19th Conference on

Computer and IT Applications in the Maritime Industries

COMPIT'20

Pontignano, 17-19 August 2020

19th International Conference on

Computer and IT Applications in the Maritime Industries

COMPIT'20

Pontignano, 17-19 August 2020

Edited by Volker Bertram

19th International Conference on Computer and IT Applications in the Maritime Industries, Pontignano, 17-19 August 2020, Hamburg, Technische Universität Hamburg-Harburg, 2020, ISBN 978-3-89220-717-7

Schriftenreihe Schiffbau
Schwarzenbergstraße 95c
D-21073 Hamburg
http://www.tuhh.de/vss

Technische Universität Hamburg-Harburg
Institut für Entwerfen von Schiffen und Schiffssicherheit
Schwarzenbergstraße 95 Gebäude C D-21073 Hamburg

Sponsored by

tutech.de

www.siemens.com/marine

www.prostep.com

www.foran.es

www.aveva.com

www.friendship-systems.com

www.sarc.nl

www.krs.co.kr

Index

Survey Simulator – A Virtual Reality Training Tool for Ship Surveys

Volker Bertram, DNV GL, Hamburg/Germany, Volker.Bertram@dnvgl.com
Tracy PLowman, DNV GL, Hamburg/Germany, Tracy.Plowman@dnvgl.com
Przemyslaw Feiner, DNV GL, Gdynia/Poland, Przemyslaw.Feiner@dnvgl.com

Abstract

This paper presents Survey Simulator (SuSi), a Virtual Reality (VR) based training tool developed by DNV GL. Survey Simulator offers virtual environments of various ship types and one offshore structure. Various deficiencies (technical and safety-related) can be simulated in the virtual environment. Trainees practice in VR finding and documenting these deficiencies. The paper describes key features, practical aspects of creating virtual environments, training options, and feed-back from users.

1. Introduction

Survey Simulator (in short SuSi) was developed for internal training of DNV GL surveyors, but then developed into a training solution for external customers. It presents a "concentrate" of corporate knowledge and "good practice" experience related to surveying maritime structures. The main idea behind the SuSi's approach is moving dangerous, time and cost consuming educational tasks into the safe and inexpensive environment of Virtual Reality, Fig.1.

Fig.1: From real environment (left) to virtual environment (right) - faster, better, less expensive professional training

2. Assorted aspects of training using Survey Simulator

2.1. Target groups and training modes

Designed originally for highly qualified DNV GL surveyors, the applications very quickly expanded to address training needs of diverse target groups:

a) Highly educated professionals (surveyors)
b) University students
c) Ship crews

Requirements for each group differ, leading to a coherent and complementary educational program with four training modes with increasing complexity and trainee challenge, Fig.2:

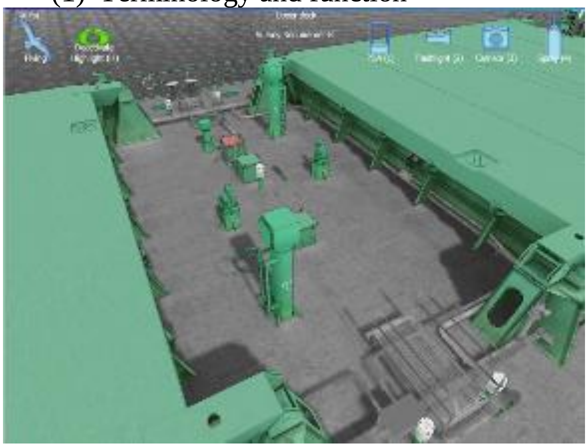
1. Introducing terminology (vocabulary) and function of ship structures

2. Identifying areas of special attention, requiring understanding of structural loads and failure modes
3. Identifying objects covered by specific survey and class rules
4. Finding (partly hidden) deficiencies, including an exam.

This approach allows adjusting the training to needs of a very wide public - from novice students to experienced surveyors.

(1) Terminology and function

(2) Areas of special attention



(3) Identifying object covered by rules

(4) Finding deficiencies

Fig.2: Training modes with increasing complexity and trainee challenge

2.2. Modelled VR scenarios

Using Unity 3D (https://unity3d.com/) as simulation environment, regions of three ship types (bulker, tanker, containership) and a semi-submersible offshore platform were modelled. Regions include tanks, cargo holds, decks, engine room, accommodation area, nautical bridge, etc., Fig.3. In addition, the outside of the bulker in a drydock and sections in a shipyard were added as regions.

In VR modelling, you must balance two opposing objectives:

1. Visualization quality – requiring high scene complexity, high level of detail and photorealistic textures.
2. Real-time performance – requiring geometry simplification, limitation of model extent and removal of small details.

The final model is always a compromise between these two objectives. In case of SuSi, the challenge was significant. Ship structures have large dimensions and the training tasks required thousands of small but important details restricting possible model simplification. For example, for the steel structure, small differences in shape and color are critical for finding deficiencies, Fig.4. In addition, scenarios contain not only the main geometry, but also (thousands of) alternatives showing different failures, deficiencies and wear conditions. The model size then exceeds many times the size of the primary geometry.

Fig.3: Selected regions modelled for Survey Simulator

Fig.4: Relatively small difference in shape and color must be modelled to indicate deficiencies

Unlike typical computer games, the 3D models in Survey Simulator were based on real ships. Generally, the available 3D models (e.g. for structural design, finite-element analyses or CFD) are ill-suited for the requirements of real-time visualization, *Lindenau and Bertram (2003)*. The conversion of an existing 3D model to a suitable VR model requires considerable effort, Fig.5:

a) Geometry preparation in CAD/CAE systems (mostly, we re-used existing data; in some cases, we produced models from the scratch) employing features for semi-automatic model editing.
b) Model simplification in DCC (Digital Content Creation) software; this is so far a highly manual process requiring skills and experience, which accounts for the largest part of development cost.
c) Integration of resulting geometry with entire scene, including connection of 3D objects with database contents describing of their non-geometrical properties.

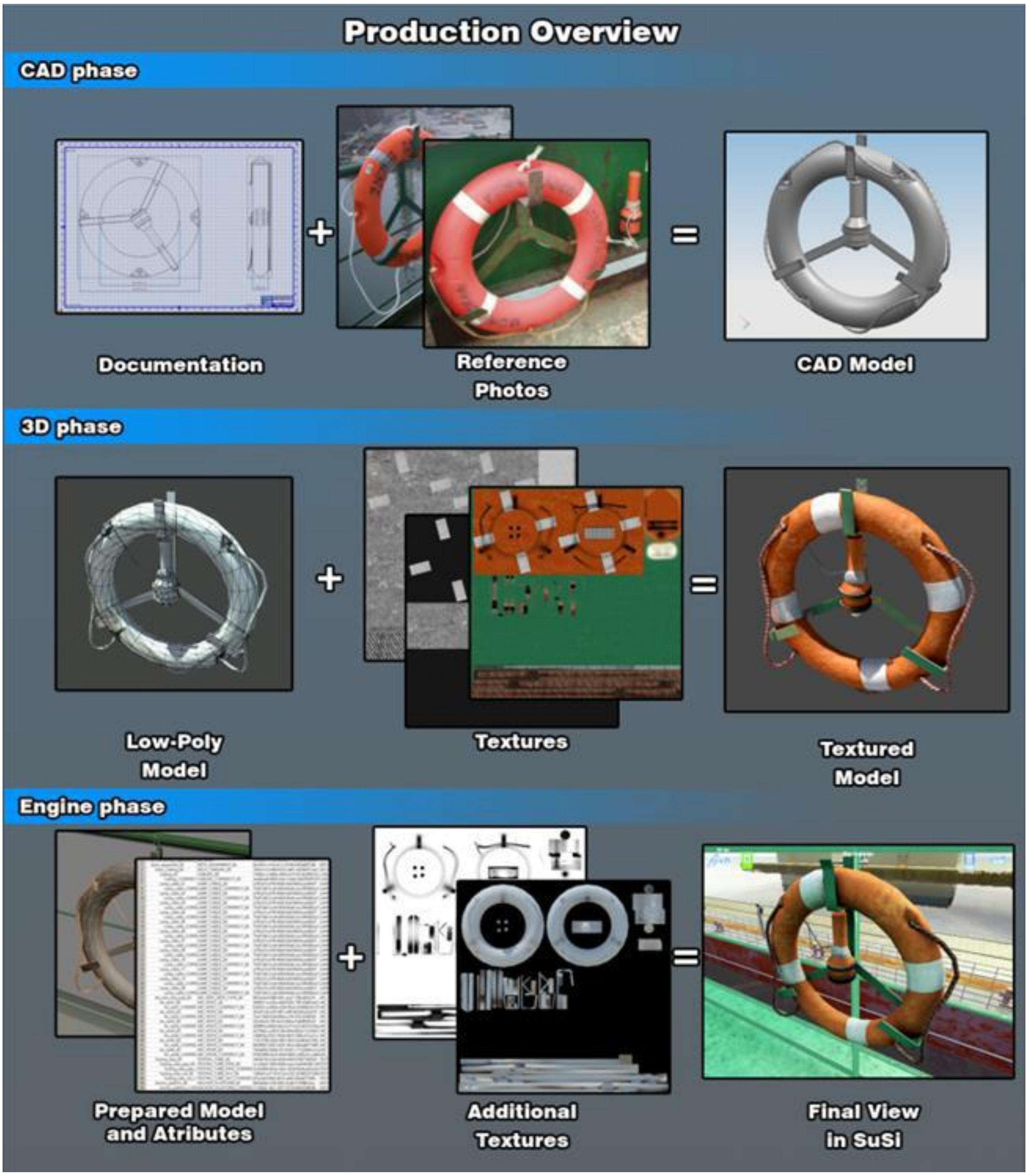


Fig.5: VR model production process

Modelling of deficiencies (cracks, buckling, etc.) is a special issue. CAD and DCC software allows efficient modelling of regular, “technical” shapes (blocks, cylinders, etc.). However, most deficiencies on ships have irregular shapes or patterns, requiring manual modelling. And a complete scenario may contain several thousands of modelled (optional) errors. This increases the size of the VR model and requires more sophisticated database management in handling the various options. The user can then choose in a central menu the condition of the ship, e.g. corrosion condition from perfect condition to heavily corroded, Fig.6. The possibility to create randomly various deficiencies in virtually unique sets prevents trainees from memorizing scenarios, forcing them to maintain high focus all the time.

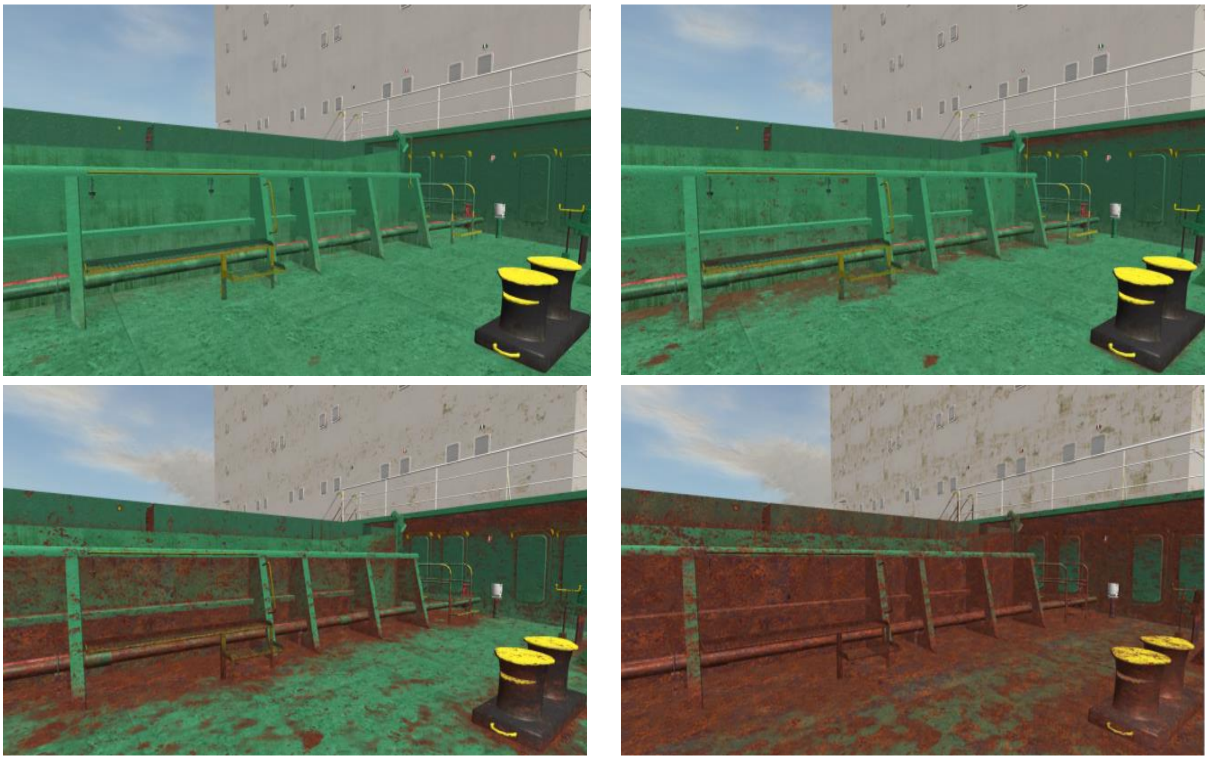


Fig.6: Progression corrosion as example for irregular deficiency modelling

2.3. Key features

The most visible asset of SuSi is its photorealistic, real-time visualization of maritime 3D models. But another important factor of good simulation is interactivity. SuSi offers virtual tools and machinery as used in real surveys, for direct interaction with the 3D model:

Smartphone – enables access to essential information about naming, survey requirements and findings during virtual inspection; gives access to attached documentation, Fig.7, photos, drawings, manuals and reporting templates.

Flashlight – necessary in dark areas

Camera – essential in preparing documentation of deficiencies during virtual inspection; allows storing of SuSi screenshots as standard jpg files and using them in the same way as pictures produces by real camera.

Spray – handy tool to mark-up deficiencies

Telescopic boom – useful when checking upper part of ship’s structure

Raft – common way of inspecting upper parts of internal structures of tankers when tanks are filled with water.

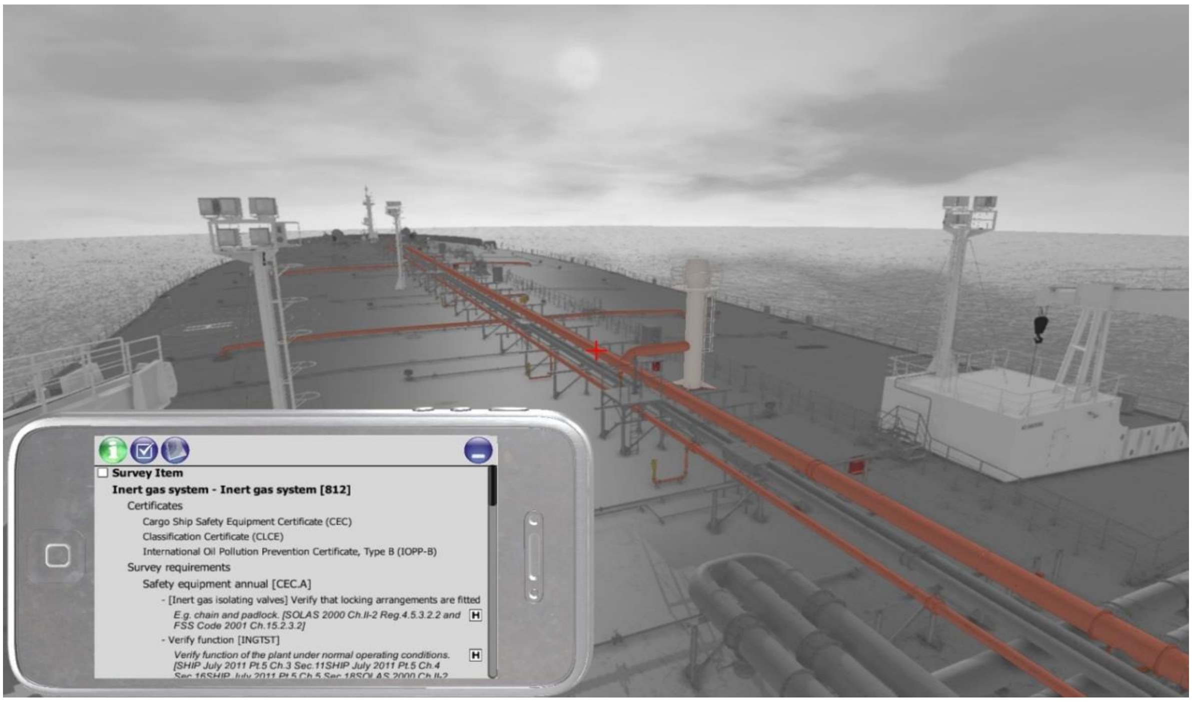


Fig.7: Example of visualization of filtered database content (only queried components are colored) with associated information displayed on virtual smartphone

Fig.8: Virtual reality finding and corresponding real-world finding

Survey Simulator is more than a video game or a visualization tool. A key value-adding part of SuSi is a knowledge base of thousands of photos of actual surveyor findings linked to the deficiencies incorporated in the VR model. Trainees can then use the virtual smartphone to retrieve photos of associated real-world defects with attached text information adding to the training effect, Fig.8.

Survey Simulator contains collision avoidance features between moving components, most notably the virtual user's body and ship parts. Users then have to navigate around the site as in a normal survey. Survey Simulator allows simple changing of user position (standing, kneeling, crawling).

The interaction with the scenario may also include the simulation of floating objects (e.g. an inflatable boat in a half-full tank moving with the water surface), refraction of light on transparent objects (water, glass) and manipulating movable parts (doors, valves, levers, etc.), Fig.9.

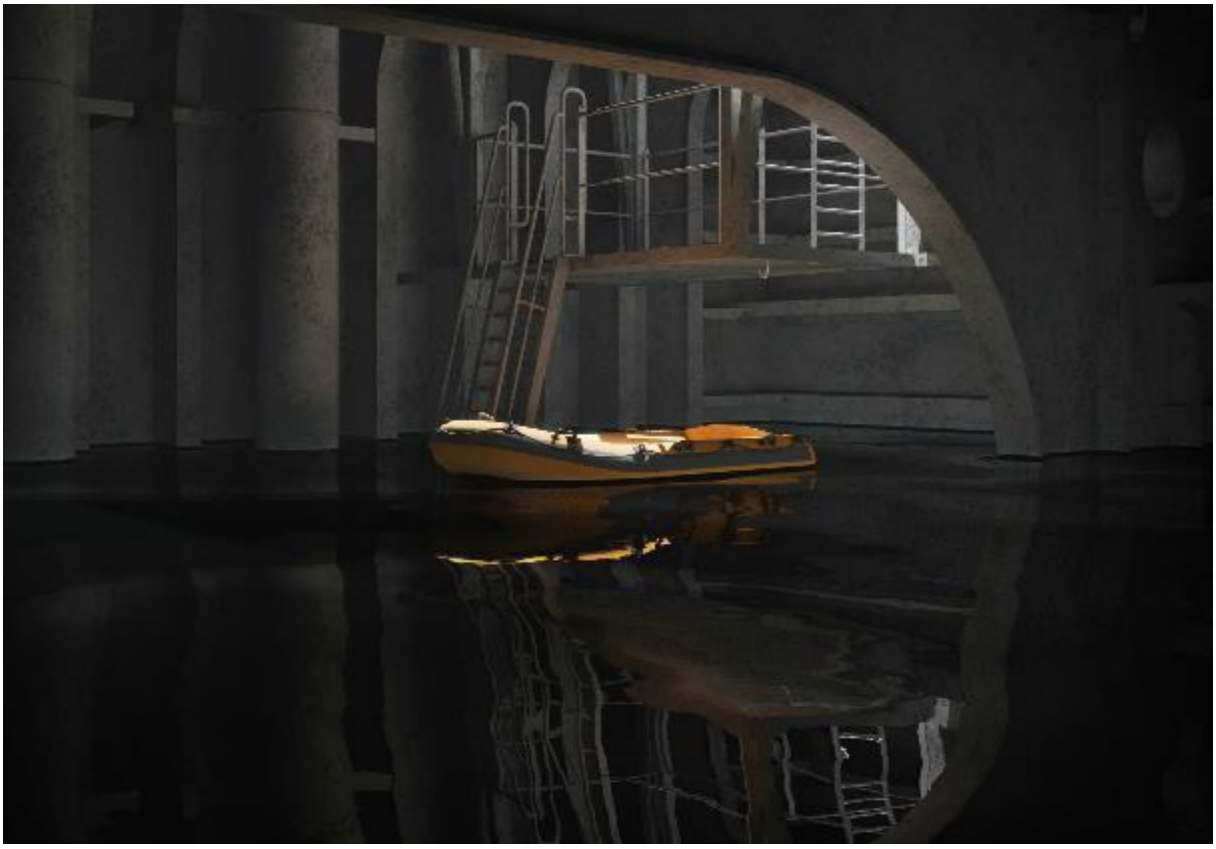


Fig.9: Moving inflatable boat on water surface; note refraction of light on water surface

2.4. Options for conducting training

Survey Simulator is a versatile training tool. The way it is used depends on the number of trainees, their computer literacy, the training objectives, available trainer and infrastructure. The options are manifold:

a) Trainer navigates through the scenario with his laptop screen projected on a large screen providing a common view for all trainees. This option works for arbitrary group size (even large audiences in university settings), but offers only limited interaction and trainee involvement. For example, the trainer can slowly navigate through a scenario, prompting trainees to interrupt when they spot a deficiency.
b) Trainer and up to six trainees work in our specialized facility in Gdynia, Fig.10. All users perform the same task, the trainer can see their progress, Fig.11, and at any time he can display a trainee's workspace on a large common screen to discuss an interesting case (error, new solution, etc.).

c) Trainer and trainees may use a CAVE (Computer automatic visualization environment) for 3D viewing. Similarly, Survey Simulator may be expanded to be used with head mounted devices such as the Oculus Rift. The 3D view impresses initially and gives stronger immersion, but some trainees feel discomfort after some time in this degree of immersion.
d) Trainees can install a demonstration version on their own laptops or smartphones (simplified scenarios for iOS and Android systems) for self-driven e-learning, e.g. for refreshing courses, Fig.12.

An embedded tutorial explains concept, navigation options, virtual tools, and training set-up using short videos on demand. Most users learn key functions such as navigation within minutes enabling them to explore the virtual environments intuitively. Less frequently used and more advanced functions are then explored as needed, and users can at any point in time toggle between the software manual/tutorial and resuming the SuSi exploration.

Fig.10: Training facility in Gdynia with central instruction and six individual "gaming" stations

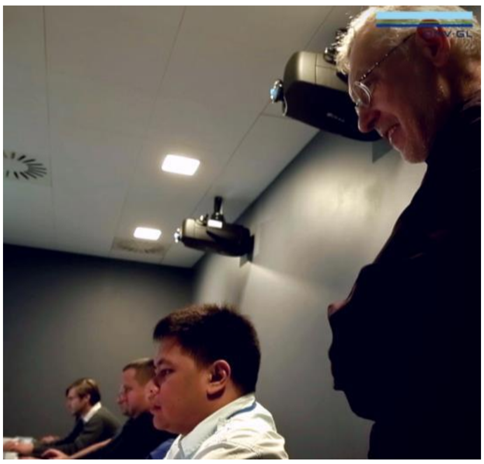


Fig.11: Supervised "gaming"

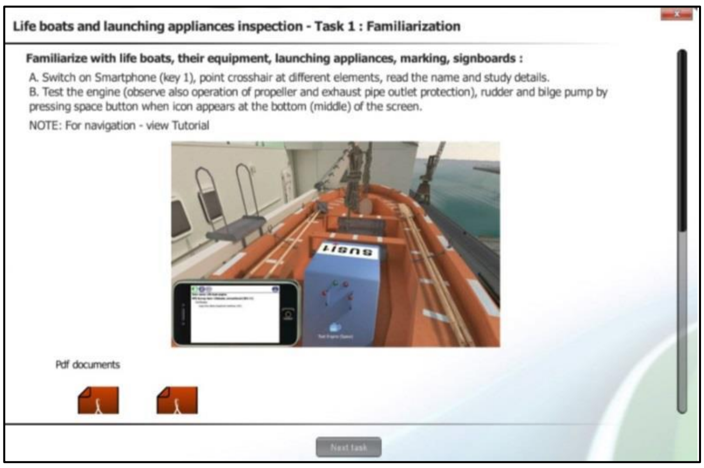


Fig.12: e-learning application

2.5. Trainee feedback

All DNV GL Academy trainings include a survey on customer satisfaction, Fig.13. In general, the Survey Simulator based training is very well perceived with an average of 4.7/5. The lowest observed marks (on average 3.9/5) are related to level of "reality impression". This is still acceptably good and the level of immersiveness and realism is chosen on purpose for reasons of pedagogy and response time. After all, Survey Simulator is a training tool and not a video game. Some important training aspects impose some tradeoffs in nice-to-have "gaming" visuals.

Feedback differs with computer literacy, gaming experience, professional background and training purpose. Generation Y (born after 1980) and younger trainees embrace the technology more rapidly. After implementing SuSi at the Faculty of Ocean Engineering and Ship Technology of TU Gdansk, students' interest in active participation in education process increased. For example, the number of new members of students' research clubs tripled. At the university, SuSi is used mainly for teaching technical English and structural design of ships and offshore structures.

Did you like the Survey Simulator (SUSI)?	0	1	2	3	4	5 X
Will applying of Survey Simulator allow to shorten the training time?	0	1	2	3	4	5 X
Would you recommend the Survey Simulator based training to others?	0	1	2	3	4	5 X
Does the virtual reality presented by Survey Simulator look realistic?	0	1	2	3 X	4	5
Does the construction presentation show details properly?	0	1	2	3	4 X	5

Fig.13: Example of feedback given on SuSi based training

3. Conclusions

Survey Simulator is a powerful training tool combining high-resolution models and graphics, knowledge base of photo images of real survey findings linked to the VR deficiencies and different training modes to address various trainee skill levels and training objectives.

References

LINDENAU, O.; BERTRAM, V. (2003), *The making of a VRML model for an SES with streamlines and pressure distribution*, 2nd Int. Conf. Computer and IT Applications in the Maritime Industries (COMPIT), Hamburg, pp.5-14, http://data.hiper-conf.info/compit2003_hamburg.pdf

Extended Reality Platform: Data-Driven Architecture

Mika Karaila, Valmet Automation, Tampere/Finland, mika.karaila@valmet.com
Vida Arabzadeh, Valmet Automation, Tampere/Finland, vida.arabzadeh@valmet.com

Abstract

For marine applications, an open Extended Reality platform provides the possibility to use any content in known formats. Shipbuilders can use 3D models with lower resolution and remove certain hull structures to hide proprietary design knowledge. End customers and ship operators can extend the content with their own 3D models, videos and scenarios. The platform merges the multi-content worlds of virtual, augmented and mixed reality into one, offering a valuable tool for accelerating training, immersing crewmembers in proper maintenance details and improving safety. Extended Reality offers users the freedom to customize scenarios according to a specific vessel. All existing digitalized materials are integrated into one open platform for easy access. Training scenarios can be fire evacuation, man overboard and lifeboat loading procedures, port arrival and departure routines, shift maintenance rounds and more. Maintenance includes remote support via the internet.

1. Introduction

Virtual reality (VR) technologies typically target the demonstration of a specific product or provide training for some product assembly or maintenance actions. This paper presents a way to integrate all necessary features into one program that is a platform with multi-functionality. The platform can be extended with new features and updated for all existing users. Furthermore, the code can handle stored configurations to be compatible with older formats. During the design phase, fewer new features were expected. Yet in real life, new features were added in each build. The development phase used agile means with multi-location teams.

This article will go through the first Extended Reality (XR) platform with selected features and discuss how the implementation process was selected. XR technology supports different 3D model formats. Real-time data is tied to the Open Platform Communications Unified Architecture (OPC UA) protocol, https://opcfoundation.org/. Since OPC UA is already widely adopted, the benefit is that most systems support it. In addition, since the solution also contains alarms and conditions, it is possible to use both immediately.

This paper will also go through the use cases that better explain how end customers are using the features and adding their own content to extend the actual virtual reality.

Finally, the conclusions and future visions will summarize the architecture solution presented.

3. XR platform

Extended reality (XR) refers to both real and virtual environments created by computer science and wearable devices. The X letter in “extended reality” has a diverse role, combining the elements from V (virtual), A (augmented) and M (mixed). To have a comprehensive understanding about mixed reality, *Milgram et al. (1994)* discussed the progression of the mixed-reality process, Fig.1.

Virtual reality (VR) is an immersive experience generated from real-world content through 360° videos, digital content or a hybrid of both, *Olmedo (2013)*. Augmented reality (AR) refers to digital objects overlaid on a real environment as well as the real-time interaction with these objects, *Rampolla and Kipper (2012)*. Mixed Reality (MR), in turn, refers to hybrid reality, which merges both real and virtual environments and produces a new world in which all real and computer-based objects exist together and even interact with each other in real time.

Real and virtual environments are the two ends of this progression, while augmented reality lies close to the real environment, *Milgram et al. (1994)*. The process continues with augmented virtuality and a virtual environment, which engages users within a world entirely replaced by virtual objects. Simply put, everything between augmented reality and virtual reality can be called mixed reality. Augmented virtuality superimposes a real image over the virtual environment. In this regard, one simple example is the weather forecast in which the real image of the reporter is overlaid on virtual maps.

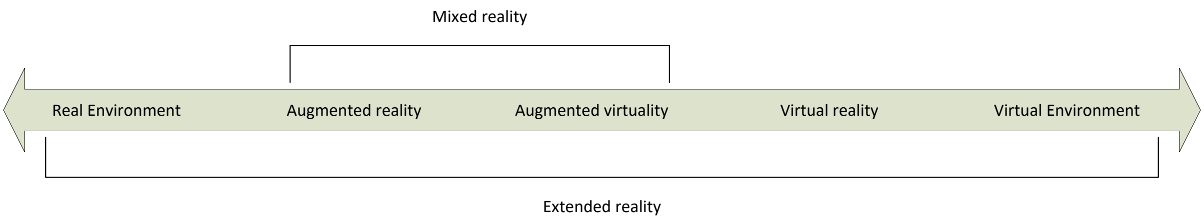


Fig.1: Continuum of extended reality, *Milgram et al. (1994)*

As this futuristic technology becomes more sophisticated, the business opportunity for marine industries and all other enterprises starts to grow far beyond expectations. These new paradigms are exciting and integrated with a lot of hardware, leading to more businesses investing in new digital resources based on human-computer interfaces, *LaValle (2018)*.

To develop extended reality solutions that could cover different phases of customer projects, all relevant features, such as design, remote support and collaboration, should be integrated into one unified extended reality platform as explained in the upcoming sections.

3.1. Scene

To develop a complex platform, it is important to choose a suitable engine. In this case, the Unity engine was selected. Models packages, such as Navisworks or Catia programs, need to be accurately reduced before exporting into the Unity game engine. This way, metadata or any other valuable detailed design information can be secured. Once the FBX files are imported into the Unity engine, a Unity Asset Bundle (UAB) package with a 1:1 model scale is ready to be applied in an XR platform.

The XR scene is normally compiled into an executable file. As different domains need to be supported, the 3D models are uploaded from the files. A file is considered to be an appropriate UAB, https://docsunity3dcom/Manual/UnityManualhtml. With this, when there is a need for a rebuild or modification of an existing environment, a completely new scene can be built from multiple 3D model files. Furthermore, through built-in settings, it is possible to create different surroundings based on specific cases, such as to render a model in a warehouse, on water or simply on the floor.

3.2. External bindings / links

The XR platform provides an opportunity for the stakeholders of shipbuilders to comprehend full-scale design vessels models in the early project phase. It is not only suitable for new building projects with available 3D or CAD models but also for existing structures. In this case, the photometric scanning technique can be used to generate a realistic overview over point clouds.

This platform facilitates the design review process by allowing a point-of-interest layout to be checked before installation, detecting any collision, checking tube routes or implementing construction changes. As a multiuser platform, it is also possible to engage in remote collaboration with other users or experts, and even via avatars, in virtual vessels. Some features created for a design review are listed in the following sections.

3.2.1. Point of interest

A point of interest (POI) connects a tag to the OPC UA server. For the POI, the panel will show either a numerical value or an animated symbol. Alternatives are the 3D tank, where the level is updated by actual measurement value; the gauge, where the needle shows the value; the motor, where the run status is indicated by color; and the valve, showing an open-closed status update. The POI has a context menu that can be extended from the settings demonstrated in Fig.2.

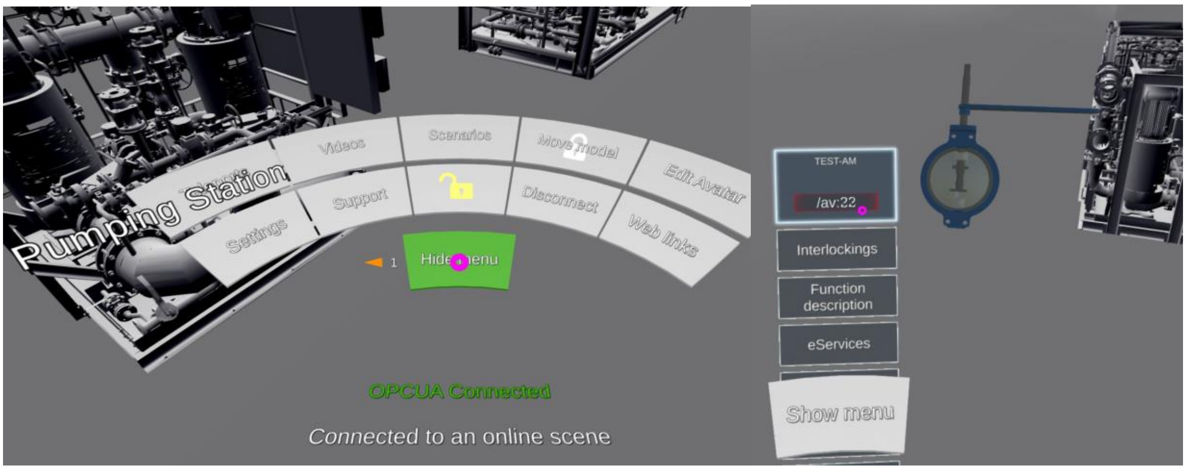


Fig.2: Alerts and context menu

A large ship might have over 100 devices, making it challenging for operators or maintenance personnel to remember all the correct locations. The connection through the OPC UA server, combined with the alerts feature, provides a list of parts requiring diagnostics. This makes it easy to teleport directly to the location of the point of interest on a vessel. Therefore, operators can check live values, which have been already tested with 5000 points. In addition, the OPC UA server enables communication between the OPC UA and the simulator, making it possible to simulate a process, which is valuable for training purposes.

3.2.2. Digital twin REST API

Based on the digital twin Application Programming Interface (API), the presented XR platform responds quickly to real-time events, taking safety to a completely new level. The Node-RED programming tool can colorfully demonstrate device status, indicating if it is in normal, warning, error or failure condition. This add-on feature facilitates the condition monitoring process and helps operators find the correct machinery part when urgent replacements are needed, since those parts are normally provided as ID tags in SIP systems.

In addition, there is the possibility to connect to a maintenance system that can provide condition predictions. With this, all parts can be highlighted with yellow, green or red colors, depending on their conditions. Fig.3 shows an example of how the Node-RED flow can be used to rotate a part along with changing its color to red to highlight the failure of this part in action.

3.2.3. Web links and tags

As the platform itself does not contain any code or information, it is possible to introduce web links that can be placed into the scene. Through the web link feature, websites can be inserted that are tied to the ship operation server to optimally use resources, documents, and links, such as eLearning.

Local experts can use head-mounted devices integrated with the extended reality platform to access the augmented information, allowing them to surf the embedded browsers in the real ship in question. Fig.4 shows how to access the web browser through the virtual environment.

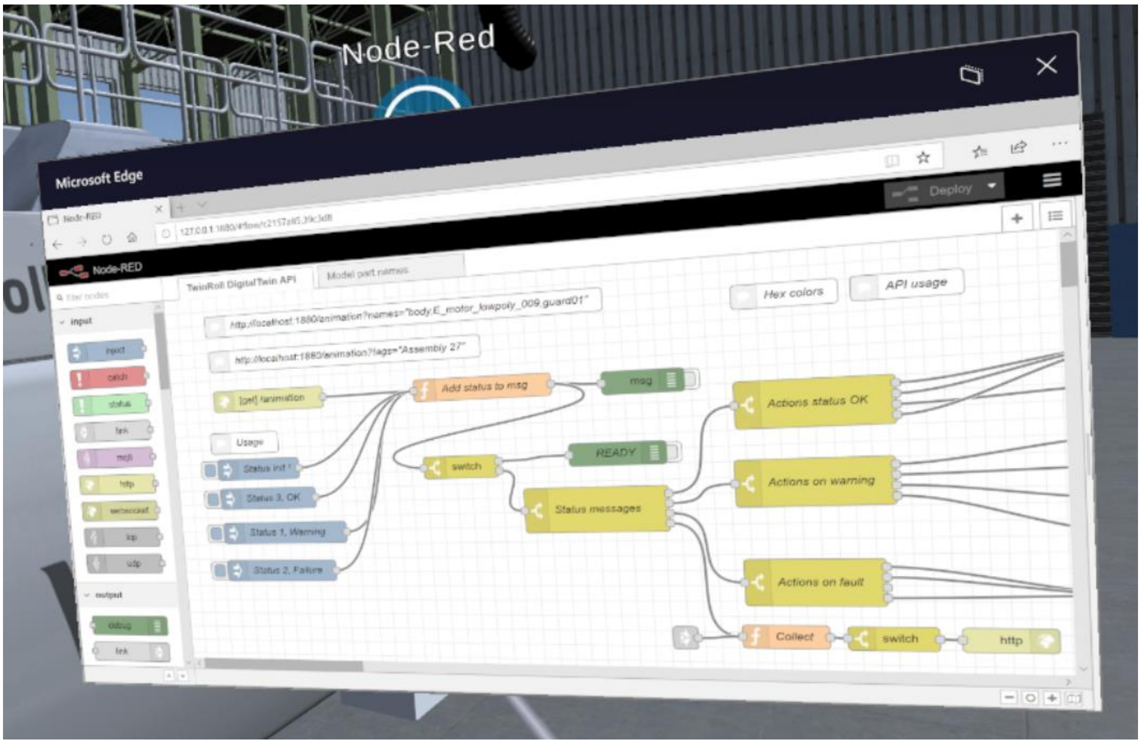


Fig.3: Digital twin REST API

Fig.4: Web links embedded into the model

Fig.5: Tag descriptions

Equipment or parts descriptions can be tagged to the virtual model. These will pop up automatically at a certain distance as demonstrated in Fig.5. The ship operator is also able to insert new tags and pin information to the equipment to access a scrubber part, for example, while exploring the vessel's surroundings in a virtual model.

3.2.4. Video links

Video files are an easy and cost-effective way to embed additional information into the scene. As illustrated in Fig.6, users can embed 2D or 360° videos into the extended reality platform to be played while walking around the virtual ship.

These videos could be provided by training teams or shot using 360° cameras from the real environment. Moreover, the same option is available using mixed-reality head-mounted devices. Service or maintenance crewmembers can have instructional videos to bring along in the real environment right next to a scrubber or cabinet.

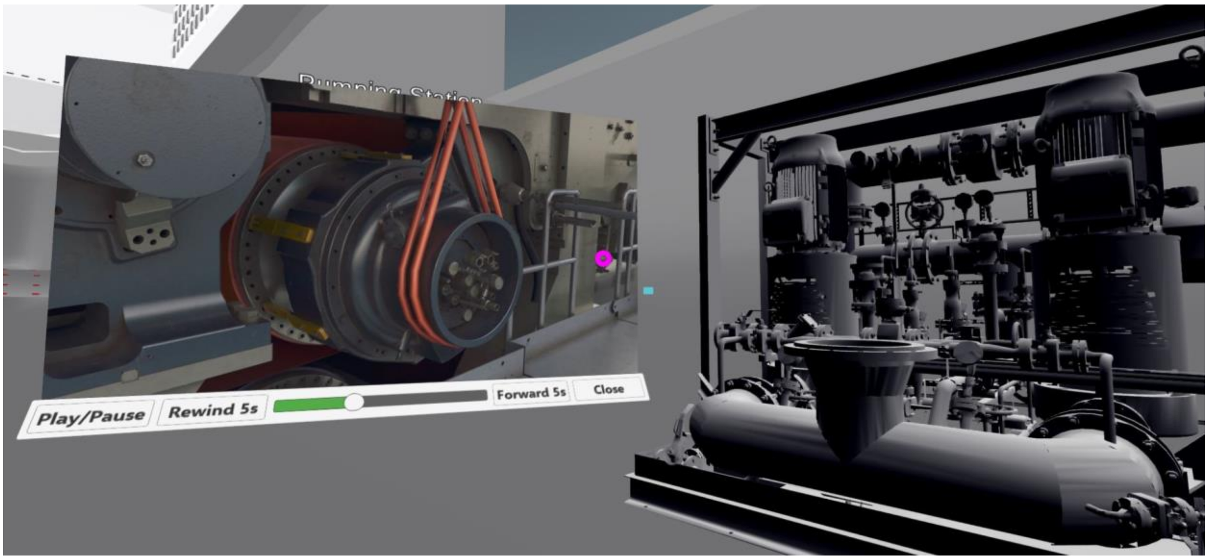


Fig.6: Video links

3.3. Settings and a built-in editor

An end user can add any web or video link, after first defining it in the settings. The next step is to place the link into the correct location inside the 3D model. The user can freely move around the environment to find the location that needs a document link or a training video.

Typically, the extended reality platform combines multiple models. Therefore, it is possible to move every object separately inside both the virtual and mixed environments in separate layers, Fig.7.

The user interface uses gamifying end-user engagement during the design review process to simplify and visualize communication. Users can use VR or MR headsets to work individually or collaboratively.

The UnlockEdit mode is another valuable tool to insert animated virtual objects, such as a tank, pump or measurement tool, or to add video links and alarms in the ships. Moreover, this feature enables shipbuilders to create their own content in both mixed and virtual environments either for training courses or preplanning maintenance services. Operators or crewmembers can implement all necessary modifications right from the platform user interface.

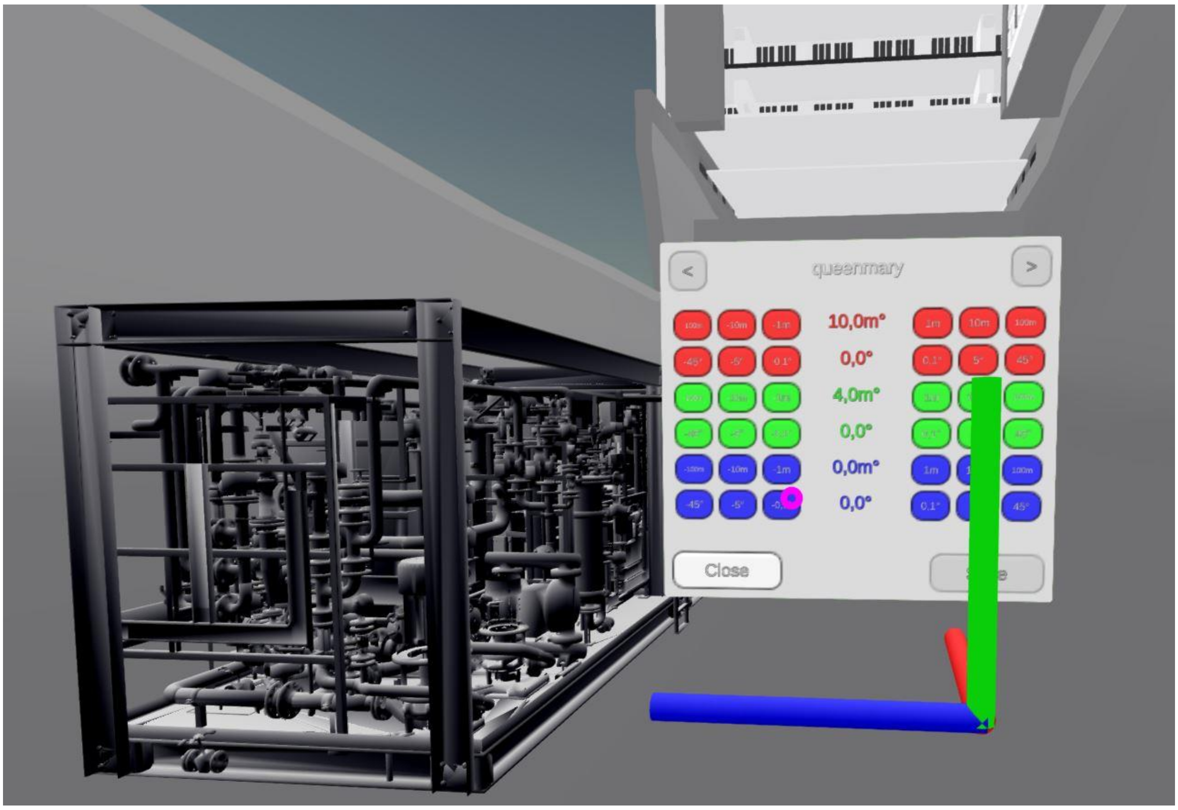


Fig.7: Model relocation

3.4. Remote support

The platform presented as an open mixed reality platform improves remote collaboration between crewmembers and ship operators by providing real-time communication through avatars or augmented mobile applications. Moreover, on-site experts can monitor live values while wearing mixed-reality head-mounted glasses and walking around in a vessel, Fig.8.


Fig.8: Remote support expert

Real-time remote collaboration with experts via avatars or augmented mobile applications help deliver information faster and more accurately, Fig.9. Virtual meetings with avatars and experts can easily cut downtime and possible errors in complicated operations, *Mourtzis et al. (2017)*.

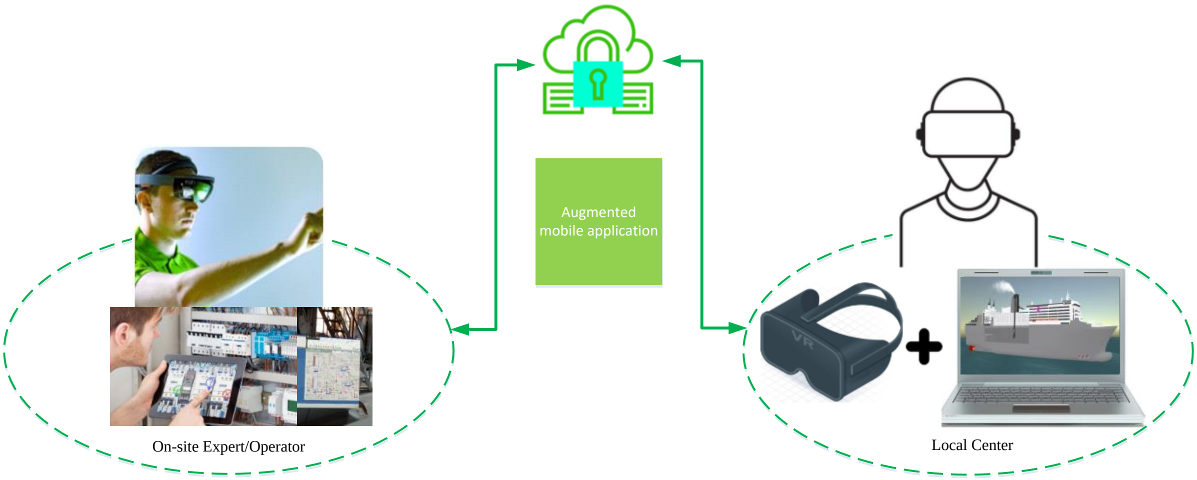


Fig.9: Embedded augmented mobile application

It is fast and easy to change an avatar's logo, name and color. It is possible to customize avatars directly in the virtual environment, as shown in Fig.10. The platform offers a unified automation system with a multiplayer mode, which can help experts and other parties meet virtually within a virtual vessel in 1:1 size and discuss the specific issues at hand.

Fig.10: Avatars

4. Use cases

Today's shipbuilders are implementing new technologies and innovative tools to create a new approach to increase value for their customers. The influence of new technologies, such as extended reality in the manufacturing industry, is manifold – from enabling a design review or developing new products to facilitating data communication and providing remote services – all of which aim to streamline business processes.

The value created by XR innovation can be discussed from the three following perspectives. First, improving collaboration between human and technology – that is the main reason for developing extended reality solutions in companies. Second, enabling companies to reach a higher level of maturity. Finally, improving business efficiency and quality. All of these can be enabled by implementing extended reality solutions. Surely, there are more criteria as well to justify the value that is possible to obtain, *NN (2017)*.

For all manufacturing industries, new technologies can be implemented to boost a customer's interaction by developing extended reality simultaneously in different areas, from the production line to services and training. To have efficient customer interaction, the extended platform must be an open platform, allowing customers to see a 3D virtual ship vessel demo, Fig.11.

Fig.11: Marine vessel model

Customers gain benefits by using a 3D visualized model via the real-time information they get from different perspectives, such as assembly, maintenance instructions or training. For example, in an assembly line for scrubbers, operators can practice the process through instructions with a head-mounted device by using the keypad, gestures and voice commands for real-time interaction. Furthermore, they are able to repeat the same tasks, such as maintenance routines, for urgent cases by immersing themselves in a virtual ship environment and simulated scenarios. This feature helps them retain knowledge during training programs and reach a higher understanding of safety at the same time. For new people who are visiting a ship for the first time to carry out maintenance, these scenarios can be pre-training practices to find a safe path to the device.

The open extended reality platform is also helping offsite support teams in local centers see the operator's on-site view and improve remote communications. Local support centers can see live data and warning spots in addition to teleporting themselves around the virtual vessel environment to different areas. When there is a connection to the OPC UA server, then the connection status in the open platform is visible. In the case of some failures in the on-site tool functionality, an alarm turns orange, warning both local and offsite experts about the problem.

The most important reason for using extended reality tools in a B2B company is that the technology should be seen as an opportunity to enhance and maintain customer relationships, rather than only a tool to bring in more revenue. In this regard, the developed extended reality platform is a multi-content application, bringing information about the real situation in several areas, such as sales and marketing, safety training or remote collaboration and support.

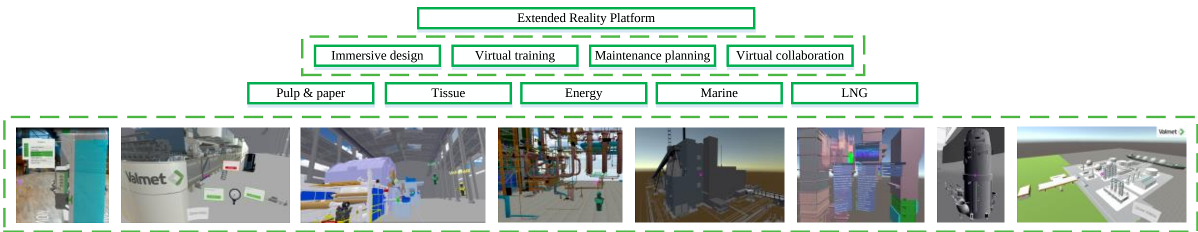


Fig.12: Extended reality platform scope

Fig.12 demonstrates the basic categorization of the extended-reality-based tools in areas such as remote collaboration and training. Several applications have been developed in each area covering different business units. For example, to assist sales and marketing, the AR kit platform in IKEA style is developed for handheld devices, such as the iPad, enabling customers to locate a product. The

Queen Mary platform has been developed for the marine industry as well as a turbine model for power plant businesses. Simultaneously, extended reality platforms have been developed over the past few years to cover all the domains in a customer's expertise, enhancing operation performance by enabling support teams via virtual remote collaboration.

5. Conclusions and future visions

We can see that mixed reality can be used in all types of industries. The next step is to support large 3D models that cannot be reduced statically to one or multiple models. New streaming-based service solutions are coming into use. The first products have been developed, but they are not yet available or widely used.

Manufacturing enterprises are interested in transferring their business models and operation to the new ecosystem known as digitalization. Digitalization is revolutionizing the economy from tangible and intangible services to customized and seamless experiences for customers. One area that manufacturing enterprises are increasingly focusing on to gain a bigger market share is extended reality, https://www.digitalmedia.fi/julkaisut/.

Extended reality technologies are reinforcing the importance of real-time remote support and virtual collaboration. Therefore, this platform can be used by any manufacturing company that is aiming to offer revolutionized services to their customers. The extended reality platform, introduced in this paper, was meaningful, since this technology provided opportunities that facilitated an interdisciplinary method to overcome common limitations in conventional company service approaches. Furthermore, these extended reality tools can support a series of services in customer projects.

The end users' willingness to pay for these new solutions is mostly tied to their expectations of the return on investment. What constitutes the return on investment for end users varies slightly, but the commonality is that they believe these extended reality solutions will reduce production downtime and increase expert availability by expanding collaboration in immersive training procedures.

Acknowledgments

The work presented in this paper is supported by the Valmet Automation research and development project "Valmet XR." The authors are greatly thankful to Soili Stadter who made it possible for us to present at this conference.

References

BURNS, E. (2020), *Transfer of tacit knowledge in organizations*, Evolving Pedagogy, https://verkkolehdet.jamk.fi/ev-peda/2018/05/23/transfer-of-tacit-knowledge-in-organizations/

DAMPNEY, K.; BUSCH, P.; RICHARDS, D. (2007), *The Meaning of Tacit Knowledge*, Australasian J. Information Systems 10/11, pp.438

LAVALLE, S.M. (2018), *Virtual Reality*, Cambridge University Press

MILGRAM, P.; TAKEMURA, H.; UTSUMI, A.; KISHINO, F. (1994), *Augmented Reality: A Class of Displays on the Reality-Virtuality Continuum*, CiteSeer, pp.282–292

MOURTZIS, D.; ZOGOPOULOS, V.; VLACHOU, E. (2017). *Augmented Reality Application to Support Remote Maintenance as a Service in the Robotics Industry*, Procedia CIRP 63, pp.46-51

NN (2017), *A Manager's Guide to Augmented Reality*, *Harvard Business Review*, https://hbr.org/2017/11/a-managers-guide-to-augmented-reality

OLMEDO, H. (2013), *Virtuality Continuum's State of the Art*, Procedia Computer Science, pp.261-270

RAMPOLLA, J.; KIPPER, G. (2012), *Augmented Reality: An Emerging Technologies Guide to AR*, Elsevier, pp.7-27

Evolution of Autonomous Surface Vehicles

Marco Bibuli, CNR-INM, Genova/Italy, marco.bibuli@cnr.it
Gabriele Bruzzone, CNR-INM, Genova/Italy, gabriele.bruzzone@cnr.it
Massimo Caccia, CNR-INM, Genova/Italy, massimo.caccia@cnr.it
Cristiano Cervellera, CNR-INM, Genova/Italy, cristiano.cervellera@cnr.it
Roberta Ferretti, CNR-INM, Genova/Italy, roberta.ferretti@inm.cnr.it
Mauro Gaggero, CNR-INM, Genova/Italy, mauro.gaggero@cnr.it
Angelo Odetti, CNR-INM, Genova/Italy, angelo.odetti@inm.cnr.it
Enrica Zereik, CNR-INM, Genova/Italy, enrica.zereik@cnr.it
Michele Viviani, UNIGE-DITEN, Genova/Italy, michele.viviani@unige.it
Marco Altosole, UNINA, Napoli/Italy, marco.altosole@unina.it

Abstract

This paper describes the technological evolution of Autonomous Surface Vehicles (ASV). The ASV 'SWAMP' has been developed and built by CNR, reflecting the general trends. The catamaran hull is equipped with Artificial Intelligence to perform its missions.

1. Introduction

The constant observation and characterization of seas, oceans, lakes, rivers and generic water reservoir is of strategic importance for the purpose of sustainability and maintenance of the environment. Maintenance and protection plans of these environments are based on a continuous observation and measurement of such areas of interest.

For these reasons in the last 30 years, there has been an increasing trend in the development and exploitation of autonomous systems capable of easing the burden of observation and monitoring of extended water areas. One of the most representative class of autonomous tools employed for a wide variety of activities in water is the one composed by the so-called Autonomous Surface Vehicles (ASVs) *Bertram (2006)*. In this time span, ASVs have evolved from research laboratory prototypes to consolidated commercial devices, providing a number of reliable functionalities.

As shown in the surveys of *Schiaretti (2017) and Moud (2018)*, a great number of Autonomous Surface Vessels (ASV) with more or less enhanced capabilities was developed in the last years arising from academic and research institutions. The growing maturity of autonomous technologies with the possibility of long-duration of the measurements and observations desired, the possibility of removing personnel from the risk, the quality of surveys with respect to traditional methods and the possibility of extending the operations to otherwise inaccessible zones make ASV the most reliable solutions for the environmental surveys. ASV and in general USV have long been used to develop technology and control systems, but are now available as COTS products with main players like Liquid Robotics (Wave Gliders), TeledyneMarine, SeaRobotics, Maribotics (Scout) and research and academic institutions that play an important role in the widespread. Thanks to these and other players starting from 2010, the number of commercial vessels has risen from almost zero up to a good number of vessels, all responding to the needs of environmental monitoring and protection that means that the ASV has become a reliable and affordable system.

However, a huge number of specific applications and interventions remain dependant on human operators, since the current technology still lacks adequate design, reliability, robustness or cost effectiveness. This aspect leads to the need of innovating ASVs starting from the design phase, as well as redefining the concepts and methodologies for effective guidance, control and interaction with the user.

With this aim in mind, this work has the goal of presenting the innovative ASV 'SWAMP', designed and realized by CNR-INM. This vehicle was designed expressly focusing on the wetlands monitoring.

These are often remote areas that require robotic solutions characterised by low logistics, low weight, small dimensions and modularity.

The concept of Wetlands, which importance is widely recognised, embraces a wide variety of mixed terrain areas: swamps, marsh, bog, floodplains, coastal areas, rivers, lake, polar regions and many others. It is easy to understand that, if there is a profound difference between these areas, on the other hand there are several peculiarities that are common to most of these zones that represent operational (and design) constraints.

Wetlands are, for definitions, areas where the earth is covered by shallow water so, in the wetlands surveys are carried on in narrow and shallow waters that are often difficult to access with classic designs. This makes the analyses of difficult implementation. The practical peculiarities of the surveys in wetlands and shallow waters lead to a development rationale that lies within a set of design constraints such that a valuable step forward can be set with respect to the current ASV technological readiness.

1. Highly reduced size and weight
   The easiness of logistic and rapid deployment of autonomous tools is nowadays an almost mandatory features; for this reason, suitable materials and efficient shaping is taken into account in the design phase. Reduced size and weight are also strictly related to reduced costs, since the ASVs become simple tools equipped with the essential payload. In many cases they can also become expendable.

2. Very shallow water navigation capability
   In the last decade, a particular attention has been posed towards the monitoring of the so-called 'wetlands', indicating such geographic areas present where water meets the land and include a variety of environments: swamps, bogs and marshes, rivers, lakes, shallow coastal areas etc. Their essential ecosystems, *Kelly (2010)*, are considered among the world's most productive environment and various international conventions, directives and projects work on their protection and monitoring. Nowadays observation and monitoring operations in such areas are usually carried out by human operators, executing tedious bathymetric campaigns in order to characterize the river-bottom profile. Depending on the specific river of interest, such manual operation could not be carried out because of deep water, rocky areas, rapids and unreachability by means of boats; such occurrences let the areas of interest completely unknown. Therefore, specific autonomous agents must be properly designed and developed.

3. Extreme modularity
   Application flexibility is one of the important keyword that has to be dealt with during the design: ASVs have to provide the capability of being adapted towards the different operational needs, including the reshaping of the structures (in order to modify the displacement, payload capacity, manoeuvring characteristics), as well as the chance to add/substitute sensors and devices on the fly.

4. Implementation of innovative control schemes
   An advanced actuation scheme is a need in order to face the most variable operating conditions, ranging from high manoeuvrability to precise dynamic positioning. Effective and redundant thrust configurations have to be studied and implemented to guarantee the motion capabilities in very different environments. Such thrust configurations also require proper advanced control scheme in order to execute the required commands, at the same time compensating for external disturbance and possible faults. Alternative control methods are based on the application of machine-learning approaches, which are recently arising interest in the community and providing encouraging results.

5. Extension to cooperative frameworks
   Extending operational area monitoring or collective a greater number of measures does not mean to enlarge the single ASV system. It has been proven by many applications that incrementing the number of employed ASV units allows to extend the monitoring coverage, collect more data at the same time, enhance the robustness of the overall system and reduce the costs. Thus, the design of a single ASV agent has to be conceived keeping in mind that a cooperative multi-vehicle system with shared information flow will be sooner or later deployed. The emerging trend of the *'Extended Ship'* also foresees the employment of autonomous agents as cooperative systems exploited as advanced observation and maintenance tools.

2. Design

On the basis of the requirements described in *Odetti et al. (2018)* an Autonomous Surface Vehicle which characteristics can be modified by specifications of the different missions and task was designed. The vehicle, Fig.1, whose acronym is SWAMP, is called Shallow Water Autonomous Multipurpose Platform, is an unmanned surface vehicle expressly conceived and designed to work in shallow and confined waters. Dimensions and hull layout of the robotic platform is shown in Fig.2. The vehicle is a sum of innovative concepts that are summarised in Fig.3.

Fig.1: The SWAMP ASV

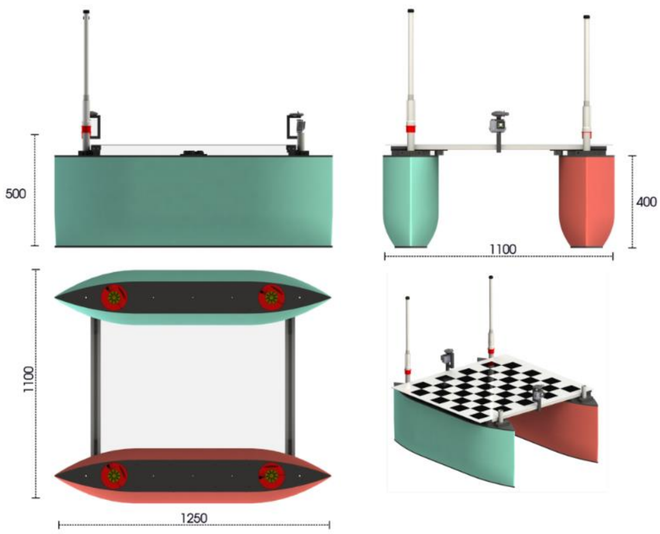


Fig.2: Vehicle general layout with measures

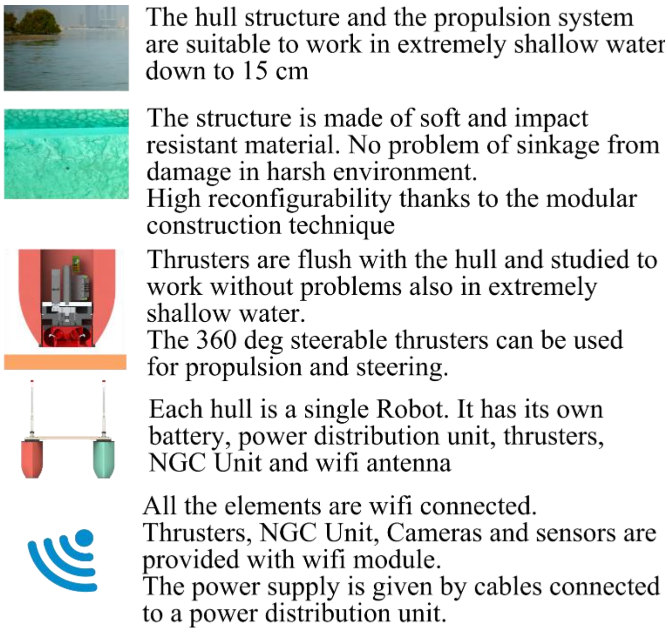


Fig.3: Vehicle main specifications layout with measures

SWAMP is a full-electric Catamaran 1.23 m long with a design breadth of 1.1 m by adopting a sliding structure the breadth is variable between 0.7 m and 1.25 m. The hull height is 0.4 m and the vehicle with the structure and the antennas is 1.1 m high. SWAMP lightweight is 38 kg with a draft of 0.1 m, the standard maximum payload is 20 kg with a consequent maximum design draft of 0.14 m but the reserve of buoyancy of SWAMP allows to embark up to 60 kg with a draft of 0.22 m. The small dimensions of the vehicle comply with the idea of a reduced logistics. The hull shape is inspired by the double-ended Wigley series but with a flat bottom as shown in Fig.4. The double-ended hull form and the propulsion layout is characterised by equally efficient sailing ahead and astern with the possibility of manoeuvring in narrow spaces, Longitudinal centre of buoyancy LCB is centered mid-hull and the hull has large B/T ratios . The hull configuration shape was chosen both for hosting four Pump-Jet azimuth thrusters expressly designed and studied for this project and to create an innovative structure that also avoids the presence of sharp edges on the hull bottom. Indeed one of the main peculiar aspects of SWAMP is the use of light, soft and impact-survival flexible structure made with a sandwich of soft closed-cell HDPE foam, HDPE plates and pultruded bars. With this design SWAMP is a completely modular vehicle that can be dismounted and transported to be remounted in various possible configurations. This flexible design allows to host various types of tools, thrusters, control systems, samplers and sensors.

Also for this reason for the propulsion the choice has fallen on the design of a modular propulsion unit based on Pump-Jet that can be easily installed on the vehicle. Such a solution allows also to remove, if necessary, some of the thrusters and/or substitute them with sensors, tools or even other thrust units. Using four azimuth thrusters gives SWAMP the controllability that is required for high quality surveys. The Pump-Jet thrusters were built in the CNR-INM labs and tested at various angles and with different configurations. The concept, design, construction and tests performed are reported in *Odetti et al. (2019)*.

One of the main peculiarities of SWAMP consists in the fact that each hull is conceived to be a single vehicle with its propulsion, navigation, guidance and control (NGC) and power system from the battery. Each monohull results to be an ASV and, thanks to the azimuth thrusters, is highly controllable. Moreover the intelligent core of each vehicle controls the monohull but is able to take over the control of the entire vehicle in the event of failure of the other core. This possibility is guaranteed by the existence of a Wi-Fi-based communication architecture.

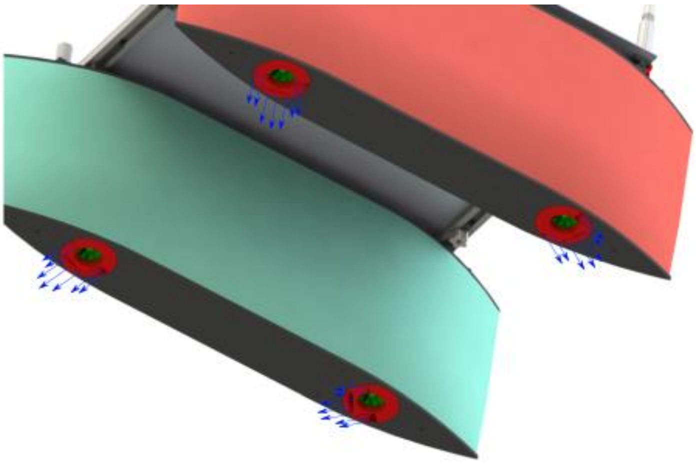


Fig.4: The propulsion layout

The SWAMP hulls were tested, also in shallow water, in the DITEN-UNIGE towing tank at various depth, payload and breadth. Maximum speed of SWAMP in infinite depth waters is 1.6 m/s, while the speed in extremely shallow waters down to 200 mm is reduced to 1 m/s due to the peculiar hydrodynamic effects occurring in shallow waters as shown in *Bertram (2011)* and *Briggs (2006)*.

The software architecture of SWAMP is based on Commercial off-the shelf components. As mentioned before one of the main hardware innovations introduced in SWAMP is the elimination of most of the possible wiring reducing the number of wires to just the power connections. The basic NGC package of each hull is composed by an IMU and a GPS. The communication is created by one communication module each hull that provides a communication framework for both its same hull and for the other hull's modules when its work is required. A resume of the main characteristic of SWAMP is reported in Table I.

Table I: SWAMP main charactersitics

Overall Length	1230	mm	
Minimum Breadth	700	mm	
Maximum Breadth	1250	mm	
Distance Between Hulls	800	mm	
Construction Height	500	mm	
Maximum Draft	150	mm	
Light Weight	35	kg	
Maximum Weight	60	kg	
Buoyancy Reserve	70	kg	
Propulsion Unit Thrust	12.5	N	
Propulsion Units	4		
Autonomy at maximum power	2	h	
Operating Speed	1	m/s	
Maximum Speed	1.6	m/s	
Max Power Cons. (single hull)	260	W	
Number of batteries nr 2	2 x 36 V - 13 Ah		
Communication Wifi - Radio	Wif-Radio		
Steering Ability DP	DP		
NGC Sensors	2 x (GPS, AHRS, Camera)		

Thanks to its characteristics, the 'SWAMP' ASV can be effectively exploited for observation, monitoring and sampling activities is harsh environment and difficultly reachable zones. A remarkable result has been obtained during a river sampling campaign, carried out in November 2019 in Liguria (Italy), with the goal of a precise and high-resolution bathymetric mapping of specific areas of interest of the Roja River. Fig.5 shows an overall bathymetric representation superimposed to the river map.

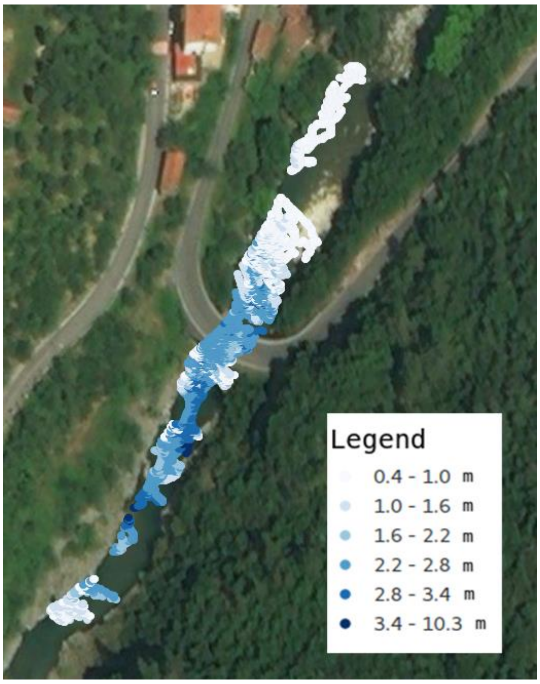


Fig.5: Range measurements gathered along the Roja river (Liguria, Italy)

3. Advanced control

This section presents the design and experimental evaluation of a model reference control for the yaw-rate regulation based on the non-linear terms compensation by an automatic adaptive scheme. The development process is inspired by the methodology described in *Ioannou and Sun (1995)* and the present work extends the approach to the non-linear yaw-rate dynamics of a catamaran-shaped USV.

The applied methodology specifically focuses on the design of a reliable yaw-rate controller capable of mitigating at best the model uncertainties, providing a quasi-linear closed-loop behavior in such a way to have the chance to subsequently design outer control loops such as heading control or path-following guidance schemes.

3.1. Modeling

The problem of developing an adaptive controller arises from the current available dynamics' model, which is corrupted by uncertainties. The current model is obtained by the combination of a CFD pre-analysis, hull motion performance tested in the naval towing tank and practical identification procedure, as suggested in *Caccia et al. (2008)*, based on proprioceptive sensor measurements obtained through proper identification maneuvers carried out during trials at sea. The combined modeling/identification procedure has led to the following dynamics form:

$$m\,\dot{u} = k_u\,u\,|u| + \,c_{u_r}\,u\,|r| + b_u\,f_u \qquad (1)$$

$$m\,\dot{v} = k_v\,v\,|v| + \,c_{v_r}\,v\,|r| + b_v\,f_v \qquad (2)$$

$$I_r\,\dot{r} = k_r\,r\,|r| + \,c_{r_u}\,r\,|u| + c_{r_v}\,r\,|v| + b_r\,\tau \qquad (3)$$

where the variables represent: u the surge speed, v the sway speed, i the yaw-rate, f_u the input surge force, f_v the input sway force, τ the input torque along the yaw axis, k_x terms (with x = {u, v, r} being the quadratic drag coefficients, c_{xx} terms being velocity coupling terms, b_x the input coefficient terms, m the mass of the vehicle and I_r the moment of inertia along the yaw axis.

Because of the reduced size and weight of the vehicle, as well as the measurement noise affecting the on-board navigation system, a significant uncertainty of the dynamics parameter set has to be considered. For this reason a number of consolidated control design methodologies are not suitable (e.g. feedback linearization) given that a high-fidelity knowledge of the dynamics model is needed.

3.2. Control design

Considering equation (3), describing the yaw motion behavior, the founding idea is to design a suitable control law for the input torque τ in such a way that the closed-loop system will behave as a linear system; a reference model is then suitably defined in order to provide a virtual behavior for the closed-loop system. To achieve the model tracking, and thus providing a reliable closed-loop yaw-rate tracking, the generated torque law will be composed by two main components: one to compensate the non-linear dynamics and the other to track the desired reference yaw-rate input signal. Clearly, the two mentioned components have a time-varying behavior in order to adapt to and compensate the current working point. The objective of the control system design will be the definition of proper time-varying components in order to guarantee the stability and tracking performance of the yaw-rate controller.

The goal is to design a suitable input control law in such a way to obtain a desired bounded and stable linear closed-loop behavior as:

$$\dot{r}_m = -a_m r + b_m r^* \quad (4)$$

where r^* is the desired yaw-rate reference, r_m is the desired yaw-rate response, $a_m > 0$ is the stable linear coefficient and b_m is the input coefficient. The system described by eq. (4) is the so-called 'reference model'.

The yaw-rate signal is then able to track the r_m state of the reference model by defining the τ control law as follows:

$$\tau = -\gamma(t)r + \lambda(t)r^* \quad (5)$$

with $\gamma(t)$ and $\lambda(t)$ being the online adapted dynamics' compensating coefficients.
The forms of the adaptive coefficients are designed following the procedure of *Ioannou and Sun (1995)*, for which the formal mathematical passages are here omitted, obtaining the following adaptation formulas:

$$\dot{\gamma} = -\eta_\gamma \, e \, r \, \mathrm{sgn}(b_r) \quad (6)$$

$$\dot{\lambda} = -\eta_\lambda \, e \, r^* \, \mathrm{sgn}(b_r) \quad (7)$$

where η_γ and γ_λ are gain factors to tune the adaptation rate, sgn(.) is the sign function and e is defined as the tracking error variable $e = r - r_m$

Fig.6 reports the evolution of the model reference state and actual vehicle's yaw-rate with respect to the desired value. Although the jerky behavior of the *r(t)* signal (caused by the conjunction of environmental disturbance, measurement noise, absence of filtering blocks for the plant signals and azimuth positioning delay), it can be noticed how its evolution tracks the reference model state $r_m(t)$.

Once the yaw-rate controller is tuned, an external loop for heading regulation can be implemented; in this particular case, a simple PD (Proportional-Derivative) scheme was implemented for the sake of rapid dual-loop heading control test. The Integral term in the heading control scheme was omitted in order to eliminate long oscillatory phases that would have extended the tuning phase. The behavior of the dual-loop heading control is then reported in Fig.7 where the satisfactory result of the heading regulation is shown. A piece-wise constant orientation is commanded to the vehicle; a small drift from the desired value can be observed and it is caused by both the environmental disturbance and the lack of a integrative term in the heading control.

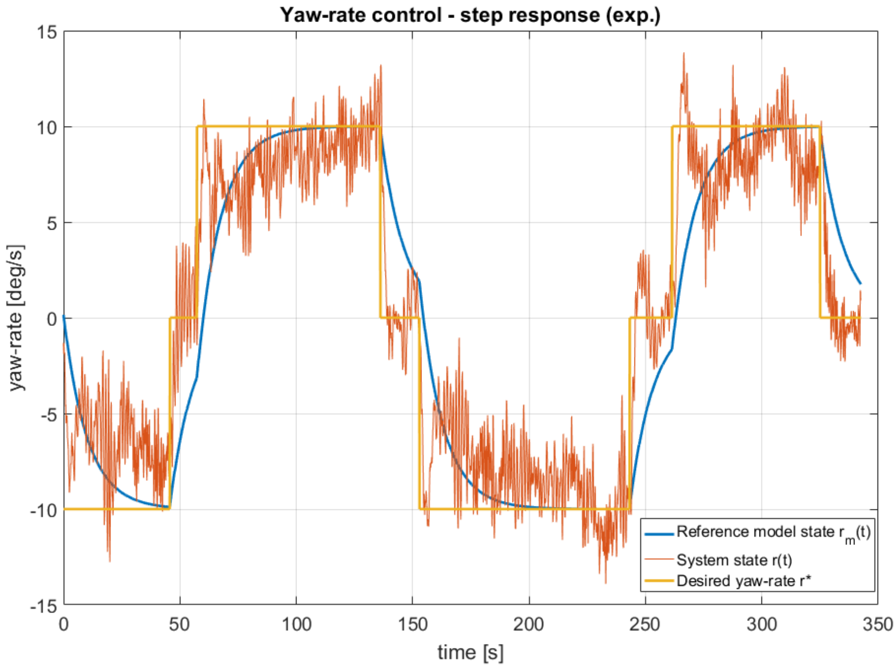


Fig.6: Step response of the adaptive yaw-rate controller during experiments

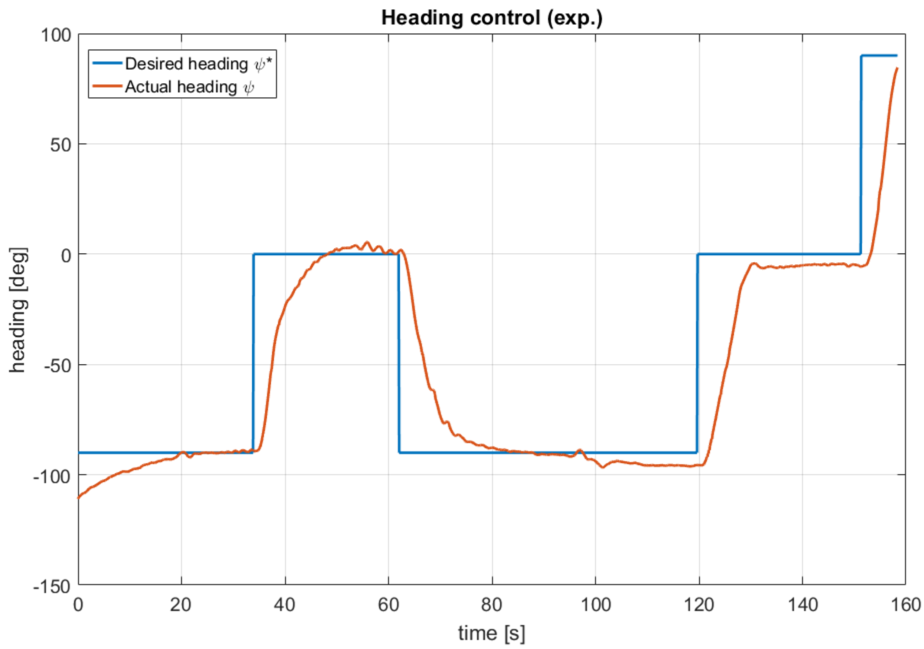


Fig.7: Heading control response experiment - desired and actual heading

3.3. Path-following

On the basis of the identification procedure and the employment of the described advanced control system, high level path-following guidance modules can be implemented. A consolidated solution, already exploited on various CNR-INM robotic platforms, is the path-following guidance system described in *Bibuli et al. (2009)* where a Lyapunov-based technique is employed to guarantee convergence and robustness of the system. The basic principle of the technique is to design a proper yaw-rate reference signal is such a way to suitably steer the vehicle towards the desired path, maintain its course along the path evolution.

Since the guidance system is decoupled from the low-level dynamics' control, the integration of the guidance module is straightforward and leads to satisfying results. A set of results where the vehicle is required to track a straight line back and forth is reported in Fig. 8. The good navigation performance can be appreciated, also keeping into account that the particular operating area was beaten by wind gusts, strongly affecting the vehicle performance.

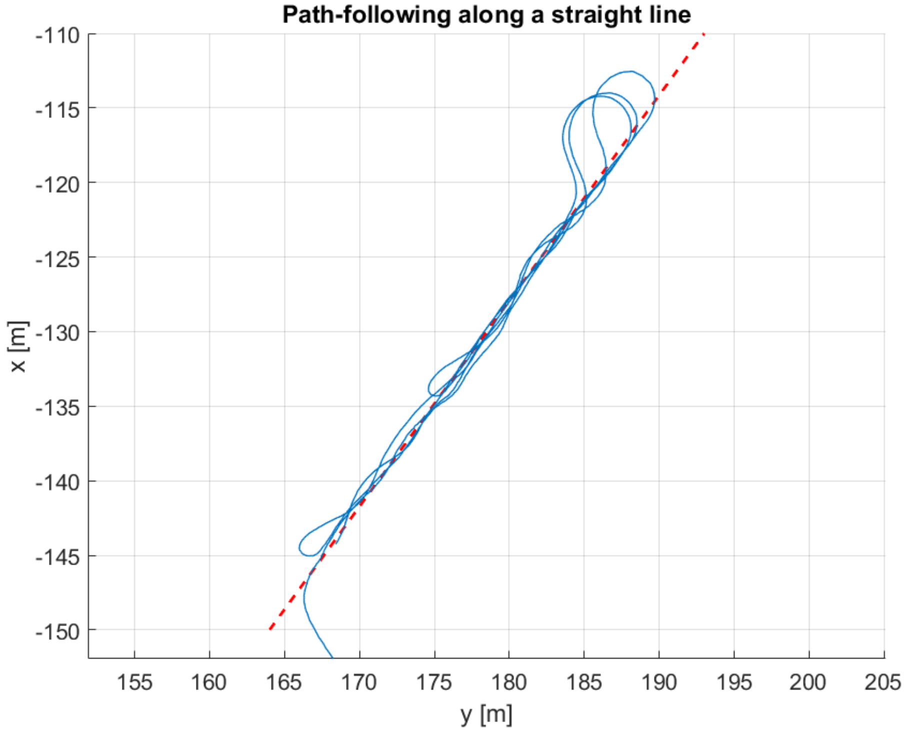


Fig.8: Path-following response experiment along a straight line

4. Application of machine-learning approaches

The development of technology in the field of autonomous robotic vehicles faces various legal and societal issues, related to public acceptance of the diffuse use of mobile robots operating in not professionally structured environments and in the presence of daily activities carried out by human beings. This has pointed out bottlenecks in robot capabilities of understanding heavily unstructured environments and situations, where complex dynamics and interactions appear.

In the case of marine and maritime systems, where strategic research is driven by the lighthouse of the autonomous ship, the possibility of applying control techniques based on artificial intelligence (AI) has received ever greater attention from the research community. In this perspective, AI is used to handle uncertain and heavily constrained dynamic systems by providing the ability to adapt to changes in the environment and to implement efficient decisions.

Reinforcement learning techniques have been applied, for instance, to ship berthing, *Amendola et al. (2018)*, navigation in restricted waters, *Figureido Nascimento et al. (2016)*, and steering of under-actuated ships, *Tuyen et al. (2017)*. Owing to its capability in capturing helmsman behavior, deep learning has been proposed to navigate autonomous vessels, and its application has been extended to situation awareness and collision avoidance.

However, operations of autonomous marine vehicles are not yet regulated by the International Maritime Organization, and several societal barriers continue to slow down the introduction of unmanned/autonomous robots in civil, commercial, and consumer domains. This is particularly evident when a closer interaction with human beings is required. Furthermore, citizens often perceive port, shipyards, and marine technology as a source of pollution or as a separator between city and sea rather than as an engine of eco-sustainable technological development. Thus, the dissemination of research activities in this field is fundamental to face the aforementioned issues. In the meantime, the European Community is supporting citizenship engagement as an effective way to connect citizens, experts, and policy makers. In this context, citizenship engagement can pave the way to social acceptance of autonomous robots interacting with human beings in their daily life.

In this section, preliminary results provided by an experiment of citizenship engagement in the research process of the 'SWAMP' ASV automatic control development, supported by AI-based approaches, are discussed. An extended presentation of the experiment is given in *Odetti et al. (2020)*.

The experiment was carried out in front of a public beach in the village of Camogli, located in Liguria Region (Italy), within the Communication Festival event, held during September 2019.

During the experiment, citizens actively contributed to data collection by driving and instructing the ASV. In more detail, volunteers piloted the ASV from a pier near the beach, with the goal of passing through two gates made up by four buoys in the same direction, thus describing an S-shaped path. The trajectories and controls performed by human beings were recorded with the aim of training an imitation system that, after collecting a sufficient number of trajectories and controls pairs, was able to drive the ASV without human intervention. A challenge was to perform training also exploiting inexperienced human pilots, such as kids or young students.

Citizens were allowed to pilot the vehicle during a couple of days. A high interest in remotely controlling the robot has been verified: this has been basically perceived as playing a kind of physical video game and a continuous access of volunteers was recorded during the whole duration of the experiment. At the end about 90 people piloted the ASV.

The paradigm under which a robot learns how to perform tasks and complex maneuvers exploiting examples from a reference controller, such as a human being, is a well-addressed research topic in the recent literature, *Hussein et al. (2017)*. This paradigm is usually referred to as imitation learning. In its basic form, often called behavioral cloning, it consists in equipping the robot with an approximating architecture that is trained in order to replicate the actions performed by the reference controller. This is obtained through a classic supervised learning approach, in which the patterns are the states visited by the robot and the corresponding targets are the actions performed by the demonstrator, i.e., the human being. In general, such an approach is not the most efficient one in terms of performance optimization. However, its intuitive principle is easily understandable by a general audience, which makes it ideal in a scenario where citizenship engagement is a goal that is more important than the optimized execution of tasks.

To learn the human behavior, recurrent neural networks (RNNs) have been employed as approximating architectures. In more detail, we have employed the echo state network (ESN) model, in which the weights of the input and recurrent layers are randomly extracted and not trained, and only the linear output layer is optimized, according to the so-called reservoir computing paradigm. This guarantees a very fast training procedure, which is an essential requirement due to the very limited available time to process data and show the results. In fact, these operations were performed in real time in front of the audience. ESNs have been already applied to robotic control in the literature, including marine vehicles and imitation learning.

Despite its overall simplicity, the combination of behavioral cloning and ESNs has allowed the trained robot to perform successfully the considered task, exhibiting a robust behavior and good generalization capabilities starting from various initial points.

In particular, the first 16 trajectories were used to train the system: as shown in Fig.5, many of them were quite far to be optimal, but considering their whole set, they covered the area between the two gates. This aspect plays a crucial role in training the ESN to execute the desired task in generic conditions. Three successful trajectories autonomously executed by the ASV on the basis of the training data are shown in Fig.5.

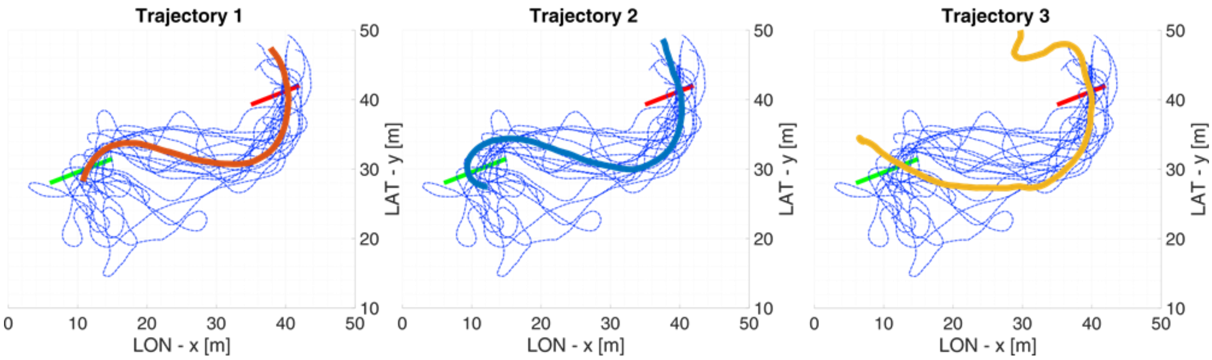


Fig.9: Three trajectories performed by SWAMP with the AI piloting the vehicle. In the background, all the 16 trajectories used for the training are reported.

A video of the experiment is available at: https://www.youtube.com/watch?v=nozHpp1TneQ

This preliminary experiment demonstrated: i) the interest of citizens with respect to technology and science, in particular where they are physically involved in an experiment directly interacting with a robot; ii) the feasibility of using imitation learning to execute (at least) basic tasks in marine robotics. The latter aspect, as well as the small number of training trajectories required by the system in this preliminary experiment, paves the way to further research focusing on how to combine learning by imitation supported by training executed using simulators and virtual environment with training executed at field with the actual ASV. In fact, one of the main bottlenecks in applying AI learning techniques to the control of marine robots is given by the cost of training with real systems.

Acknowledgments

We want to thank the technical staff composed by Giorgio Bruzzone, Edoardo Spirandelli and Mauri Giacopelli for their contribution in the design, construction and testing of the mechanical and electrical systems of the robotic platform.

References

AMENDOLA, J.; TANNURI, E.; COZMAN, F.; REALI, A. (2018), *Batch reinforcement learning of feasible trajectories in a shipmaneuvering simulator*, Anais do XV Encontro Nacionalde Inteligencia Artificial e Computacional, pp.263-274

BERTRAM, V. (2006), *Unmanned surface vehicles – A survey*, Singapore Maritime and Port Journal

BERTRAM, V. (2011), *Practical ship hydrodynamics*, Elsevier

BIBULI, M.; BRUZZONE, G.; CACCIA, M.; LAPIERRE, L. (2009), *Path following algorithms and experiments for an unmanned surface vehicle*, J. Field Robotics 26/8, pp.669-688

BRIGGS, M. J. (2006), *Ship squat predictions for ship/tow simulator*, Engineer Research and Development Center Vicksburg Ms Coastal and Hydraulics Lab, Tech. Rep.

CACCIA, M.; BRUZZONE, G.; BONO, R. (2008), *A practical approach to modeling and identification of small autonomous surface craft*, IEEE J. Oceanic Engineering 33/2, pp.133-145

FIGUEIREDO NASCIMENTO, S.; CUCCILLATO, E.; SCHADE, S.; GUIMARAES PEREIRA, A. (2016), *Citizen engagement in science and policy-making*, Technical report, Science for Policy report by the Joint Research Centre

HUSSEIN, A.; GABER, M.; ELYAN, E.; JAYNE, C. (2017), *Imitation learning: A survey of learning methods*, ACM Computing Surveys 50/2, pp.1-35

IOANNOU, P.A.; SUN, J. (1995), *Robust Adaptive Control*, Prentice-Hall

KEDDY, P.A. (2010), *Wetland ecology: principles and conservation*, Cambridge University Press

MOUD, H. I.; SHOJAEI, A.; FLOOD, I., (2018), *Current and future applications of unmanned surface, underwater, and ground vehicles in construction*, Construction Research Congress, pp. 106-115

ODETTI, A.; BIBULI, M.; BRUZZONE, G.; CERVELLERA, C.; FERRETTI, R.; GAGGERO, M.; ZEREIK, E.; CACCIA, M. (2020), *A preliminary experiment combining marine robotics and citizenship engagement using imitation learning*, 21st IFAC World Congress, Berlin

SCHIARETTI, M.; CHEN, L.; NEGENBORN, R. R. (2017), *Survey on autonomous surface vessels: Part I - a new detailed definition of autonomy levels*, Int. Conf. Computational Logistics, Springer, pp.219-233

SCHIARETTI, M.; CHEN, L.; NEGENBORN, R. R. (2017), *Survey on autonomous surface vessels: Part II - categorization of 60 prototypes and future applications*, Int. Conf. Computational Logistics, Springer, pp. 234-252

TUVEN, L.P.; LAYEK, M.A.; VIEN, N.A.; CHUNG, T. (2017), *Deep reinforcement learning algorithms for steering an underactuated ship*, IEEE Int. Conf. on Multisensor Fusion and Integration for Intelligent Systems, pp.602-607

WHITTAKER, T. (2002), *A physical study of fast ferry wash characteristics in shallow water*, MCA Research Project 457

Enhancing the Quality and Reliability of Maintenance Operations using Mixed Reality

Peter D. Petersen, MAN Energy Solutions, Copenhagen/Denmark, peterd.petersen@man-es.com

Abstract

This paper presents the MAN CEON TechGuide, a multimedia application based on mixed reality developed for mobile devices. The aim has been to make intuitive, interactive and engaging maintenance and troubleshooting instructions for two- and four-stroke engines by augmenting real life engine components and maintenance scenarios on the engine. Using mobile devices or head-mounted displays, the marine engineer has access to visual guidance in 3D, written instructions, diagrams, checklists, videos, etc. on site at the engine. The application has been tested on a vessel from Berge Bulk, where it was connected to the planned maintenance system to generate reports and update stocks.

1. Introduction

Behind the evolution of today's optimised two- and four-stroke marine engines lies a century-long tradition of engine development. A process driven by a number of changes within the maritime business, in particular emission and engine power regulations, which often govern technology development. It is our goal that engine documentation must adapt to the course and the speed of engine development but also reflect working conditions onboard and learning methods of today's engineers.

The digital transformation of technical documentation suits this goal and opens up application of different technologies in the development of engine maintenance instructions. The fast development of information technology and rapid maturation of virtual (VR) and augmented reality (AR) facilitate the development of an intuitive and interactive maintenance guidance solution aimed at a safe, reliable and fast execution of engine maintenance. AR technology, hardware and software, develops rapidly and new start-ups appear nearly every day. We have developed MAN CEON TechGuide (MCT) that combines augmented reality (AR) with a product instruction and maintenance system. The application is based on various publicly available software, which gives the agility to base the development on the latest and most suitable software.

The development philosophy behind our contemporary engine maintenance guidance is to adapt the development of manuals and instructions to the individual engineer onboard and the working environment. The challenging working environment is often characterised by reduced crew sizes, changing maintenance teams, small time slots allocated for maintenance and administrative work. This means that the crew occasionally works under stressful high-pressure conditions. They are often under time constraints and must adhere to new and more stringent guidelines. Overall, conditions that do not promote a safe and reliable execution of maintenance tasks on increasingly complex engines.

To support the engineers, our technical documentation must be available on media platforms preferred and known by the generation of marine engineers maintaining our engines onboard today. The visualisation of guidelines or instructions must enhance and make the individual information gathering experience intuitive and easy. The guidelines should promote a safe and reliable execution of specific maintenance tasks by adapting information on a task-level. Finally, it must also ensure documentation of tasks carried out according to maker's instructions.

Our research and analysis show that augmented reality can add the sought-after properties to the application, properties like intuitive, interactive and engaging, for example by augmenting real life engine components and maintenance scenarios on in-situ engines. Thus, acquiring instructions is based on real-time expert guidance rather than memorising instructions.

The digital nature of the application enables an automatic execution or user-prompted execution of the paperwork necessary to document how the maintenance task was carried out. The engineer does not have to memorise the course of events and use time to report what was done, the application prompts the user to document this as the job progresses. The application stores a report in the planned maintenance system (PMS) for future use.

2. MAN CEON TechGuide

Because of the increasing complexity of propulsion systems, in particular two- and four stroke engines, together with the increasing diversity of equipment onboard vessels, an engineer needs more knowledge, expertise and experience to operate, maintain and troubleshoot.

Fig.1: Multiplatform maintenance application

The overall intentions with the MCT solution are to minimise the time spent on finding the correct maintenance information and to improve the quality and the reliability of the maintenance by providing clear and adequate guidance. It accords with the goal to minimise human error related occurrences of engine breakdown and downtime occurring shortly after overhaul, *Malm (2018)*. The MCT allows hands-free use on mobile devices, i.e. mobile phones, tablets or head-mounted displays (HMD), Fig.1.

The application uses for example a tablet to automatically recognise products and individual components. It then loads the relevant content from the database of product and component instructions to deliver the relevant maintenance instructions quickly.

A solution where the engineer can work at the workbench or on the engine while obtaining instructions, as opposed to a computer application that requires the engineer to sit by a screen. The engineer can view instructions as augmented reality, either overlaid real-life engine components or as a model next to the component.

One of the priorities with the MCT is to ensure that only due maintenance tasks are active, that the engineer can carry out maintenance according to makers specifications, and that the ship's planned maintenance system (PMS) and spare part stock are updated after completion of the task.

Today, the MCT engine documentation platform is an integrated part of new two-stroke MAN B&W ME-LGIP engines, which run on liquefied petroleum gas.

2.1. Application development philosophy

In general, the MCT relies on 3D models created from in-house CAD data while the text is data from our content management system (CMS). We have built the MCT in Unity and used Vuforia to implement the AR functionalities: Object recognition, QR recognition and surface awareness. The data in the MCT is available data beyond PDFs, including text snippets, images, movies and HTML. Artificial voice, maintenance guides, work reports and 3D-models are added to these data.

Through the years, we have accumulated a large library of multimedia recorded on various media like VHS tapes and DVDs. In the application, the content of this library is associated with relevant instructions.

To augment content into the real world, the application requires spatial information and physical information about the surroundings as it must recognise table surfaces, workbenches, and so on. The 3D models and the AR-animations created from CAD data are not locked in a specific perspective. The user can move the tablet or head-mounted display around to view components from different angles, get explosion views of spare parts, etc.

The application sees components thanks to object recognition based on part information loaded into the application. The application recognises a component either by data matrix code or by recognising the overall design of the component. As an example, the AR-view recognises a fuel valve, uses that as a "real-life index", and shows the documentation and information related to that specific fuel valve, for example maintenance descriptions and spares. The engineer can choose to view the information as AR-animations, a movie or text.

To be of any value to the engineer, the application contains engine-specific task lists, work cards and information about the recommended wear parts. The application is usable for all products from MAN Energy Solutions with work cards.

An application-programming interface (API) supports communication with planned maintenance systems of different make. Currently, there are three ways to initiate due maintenance tasks in the application. The first is by initiating the task in the PMS, which loads it to the MCT. You can then grab your tablet or phone, go to the engine or workshop and follow the instructions in the MCT. The application has two modes termed In-situ and Table study mode for either in situ work or pre-planning work with the team.

The second way to initiate a task is by automatic on-site identification of an engine component due for maintenance. The MCT is designed with QR recognition. A QR code reader on the phone or tablet scans and reads a component code, whereby it uniquely identifies the component and supplies related maintenance information. The likelihood of retrieving information related to a different component is minimised.

The third way to initiate maintenance guidance is by recognising components by data matrix codes or by the overall physical shape of a component, the AR view recognises the gas injection valve and supplies the documentation.

To increase user friendliness and encourage the user to explore and discover content and information, the application has a search function. The application can deliver data to various platforms using iOS. One of the most important steps in the development of the application has been to test the functionalities in a real engine room. Therefore, engineers on various bulk carriers and container ships have tested it. To receive and evaluate the user experience based on the feedback has been an essential part of designing and optimising the application.

The MCT content is continuously updated with new spare parts, service letters, etc. To ensure the future development of the MCT, it has been developed with a feedback function that enables users to report errors or suggestions.

2.2. Design structure

A maintenance task that used to be simple may cover many steps today. When considering the increasing complexity and number of equipment in an engine room, it is not hard to imagine the difficulties changing crew members and maintenance teams face when trying to remember the manuals.

Instead of returning to the engine control room, browsing through a manual, leaving the engine for help or struggling to recall previous training, the engineer will have the information available on-site directly in the line of sight and at the fingertips or on, for example, a tablet.

It has been our aim to show the maintenance task in a more modern way in steps together with the needed information, safety instructions, preparation of the engine, allocation of the necessary tools, tightening values, etc., and to gather the information in one place. Fig.2 shows the design structure of the MCT with a description of the graphical user interface.

2.2.1. Multiplatform application

The engineer can opt to receive instructions according to personal preference, including instructional videos, overlaid step-by-step animations, audio instructions, or text instructions. In this way, the interactive application accommodates individual learning profiles by offering information in five different ways. The application can be voice controlled hands-free to forward, repeat or record during the overhaul or it can be controlled by hand gestures.

Fig.3 shows the MCT application in use on a tablet. In this example it provides an augmented view of a step from an overhaul sequence of a gas injection valve. It shows how AR animations guide the engineer through the different steps required (dismantle the valve, overhaul, replace wear parts, etc.) by highlighting or pinpointing specific parts or visualising part design.

If the engineer wants to read the instructions, these can be shown as text below the valve in Fig.2.

It is also possible to see the individual steps or just one aspects of the overhaul by watching an expert carry out the specific overhaul in a maintenance movie, as illustrated in Fig.4.

Fig.2: Graphical user interface of the MCT

1. Documentation index
2. Category filter for documentation index (1), work cards, tools or spare parts as shown in Fig.2
3. Details:
    a. part list for spare parts
    b. step-by-step instructions for work cards
4. Parts or timeline navigation (3)
5. Subtitle area, only visible for work cards to complement the speech for instructions
6. Extra media, such as PDF from the traditional manual and video clips when available
7. Mode switch, In Situ for when standing by the components in the workshop and Table Study for when preparing to perform a task, such as at a brief meeting in the control room or similar
8. UI switch to show or hide the sidebars 1 and 2 or 3 and 6
9. Content area, usually the camera view that AR technology relies on


Fig.3: Augmented step from a gas injection valve overhaul procedure

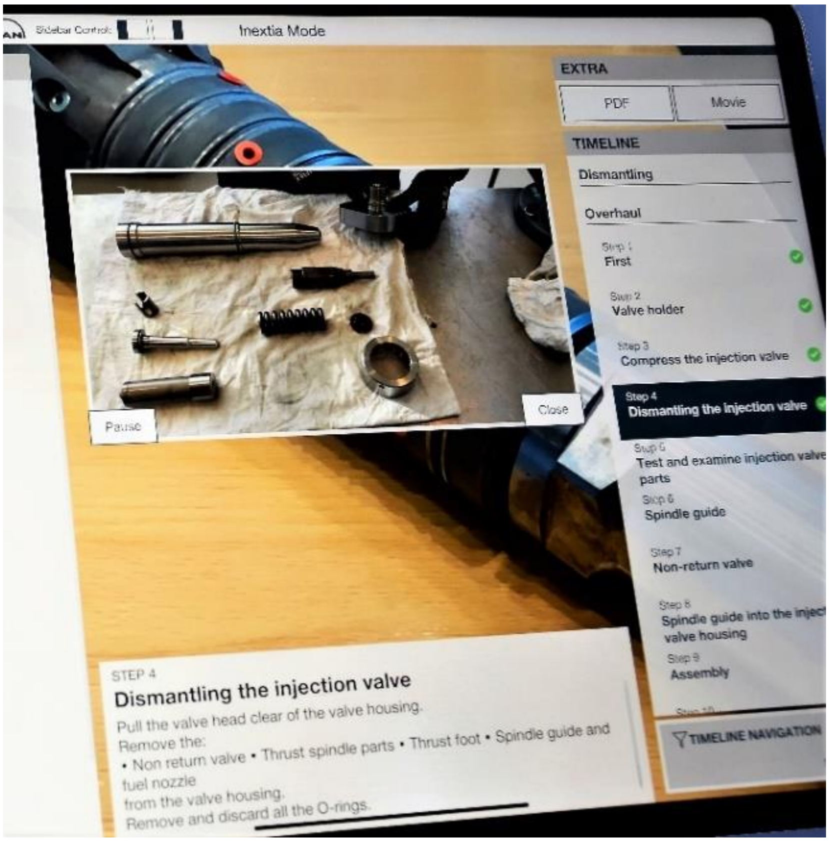


Fig.4: Step from the injection valve maintenance movie

Overhaul instructions are still available in a conventional pdf, as Fig.5 shows.

Fig. 5: Conventional PDF-based instructions

2.1.2. Maintenance and documentation

This subchapter describes maintenance of a gas injection valve for a two-stroke MAN B&W ME-LGIP engine performed by following the MCT guidance.

Fig.6 shows the maintenance screen, where the left menu or documentation index shows all information necessary to carry out the maintenance (work card), i.e. technical data, tools, safety precautions, replacement and wearing parts.

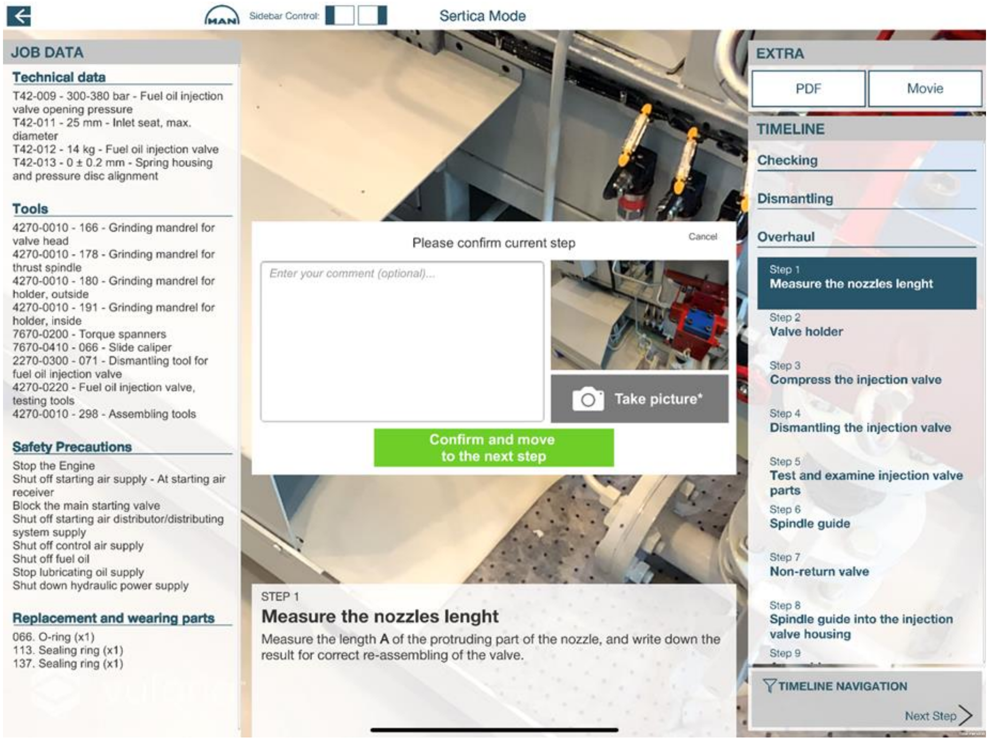


Fig. 6: Maintenance screen layout showing completion of step 1 in the overhaul procedure

The right-hand side menu contains the different maintenance steps and an overhaul timeline (Checking, Dismantling and Overhaul), which shows the progression of the task. As seen in Fig.7, the completed steps appear to the right as green check boxes. The engineer also chooses to view the extra media here, i.e. to see the maintenance movie or read the instructions in a pdf file.

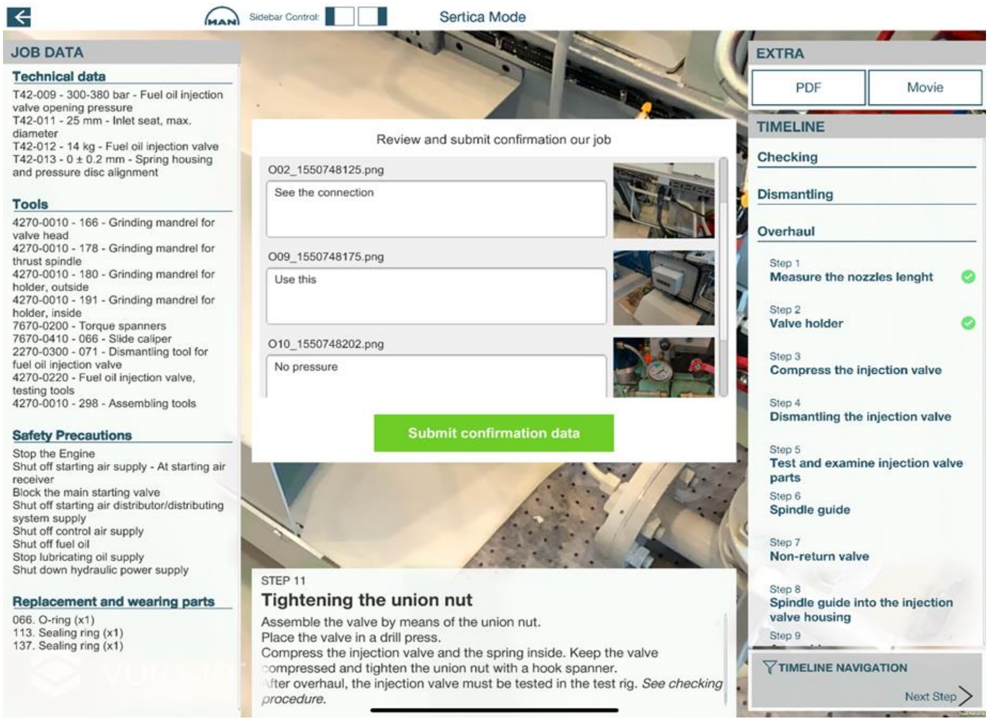


Fig. 7: Maintenance screen layout showing completion of the overhaul procedure

As part of the maintenance procedure, the engineer has the option to record a video or make comments to the individual steps of the maintenance task. In general terms, it can be observations of irregularities or defects of any kind or just observations of importance to the present and future engine crews.

The application prompts the user to take a screen dump after completion of each step and after completing the overhaul. These photos are stored together with comments from the engineer in a report in the PMS for future reference.

The engineer gets information about store status and location of the required wear parts for overhauling the injection valve when the MCT connects to the PMS. The engineer must enter the number of used wear parts into the application, which ensures an updated spare part status in the PMS. If the recommended wear parts for some reason are not used, the application prompts the engineer to enter an explanation.

The built-in reporting feature also benefits the shipowner, who can document that maintenance is according to maker's specifications.

2.2.3 Spare-part visualisation and knowledge

Once the gas injection valve is identified, the digital solution offers the engineer a range of choices. The valve can be examined in an exploded component-level view or an X-ray view of the valve can be projected over the gas injection valve in question.

The engineer can change the 3D-view of the valve thanks to the interactive 3D spare part visualisation.

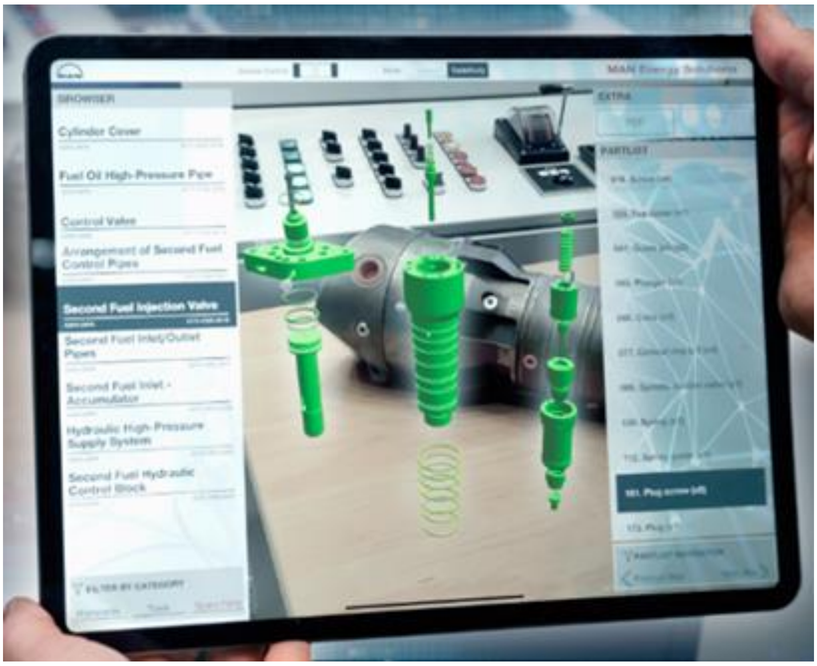


Fig. 8: 3D interactive view of the exploded gas injection valve

The structure of the interactive spare part catalogue in Fig.8 is different major spare part categories to the left (documentation index) and components (details) for the chosen category to the right. The content area in the middle (camera view) shows the augmented 3D exploded view of the injection valve.

The benefit of the 3D view is that the engineer can examine and gain knowledge about not only the gas injection valve, but also about different engine spare parts. The 3D presentation makes it easier and faster to gain knowledge about the different spare parts of the valve and of the interior valve design.

The engineer can click on a spare part and access stock availability and location data from the ship operator's inventory system. We have completed the integration with a number of different inventory management systems, for example Sertica, but can add others on demand.

3. On-site test of MAN CEON TechGuide

In October 2019, the first installation of the MCT was completed on the ore carrier 'Berge K2', Fig.9. The MCT was installed onboard the vessel and integrated with the PMS to enable communication.

- Shipowner: BergeBulk
- Ship name: *Berge K2*
- Flag: UNITED KINGDOM
- Home port: DOUGLAS (ISLE OF MAN)
  Gross tonnage: 133,500
- Deadweight: 262,583 t
- Length overall x breadth 327m × 57m
- Year built: 2015
- Main engine: G80ME-C9.2

Fig. 9: M/V 'Berge K2' and ship particulars

Fig.10 shows the on-site test of the MCT application in the workshop on the vessel. The crew carried out overhaul of fuel injection valves from the two-stroke main engine by following the MCT instructions on a tablet. During the overhaul, data was exchanged with the ship's PMS to generate a report and to update the stock system.

Fig. 10: MCT-guided overhaul of injection valve carried out in the workshop on the vessel

4. Summary and Outlook

Digitalisation and also the maturity and availability of AR technology provide us with new opportunities to create documentation platforms and applications to support engineers when maintaining equipment onboard.

The documentation development is necessary viewed in the light of today's possibilities for designing and specifying two- and four-stroke propulsion plants to specific requirements. Depending on the trade, schedule and charter, the choices are many within the topics: fuel type, emission reduction strategies, power generation, hybridisation and waste heat recovery systems.

We have addressed the potentially complex engine configurations and their maintenance by equipping these with easily located and digested maintenance guidance. MAN CEON TechGuide exploits the combination of automatic specific and updated product information with five different ways to obtain this information. The application design enables hands-free operation, the engineer can voice control the application or control it by hand gestures. Spare part availability and location on board are also available to the engineer during maintenance task.

Data exchange between the application and the PMS system ensures a seamless workflow and it reduces the administrative burden of the engineer by automatically generating a report based on input made during maintenance.

In 2020, the MAN CEON TechGuide will support more product types from MAN Energy Solutions, and workflows will be implemented to create animations and sequences as part of the daily work in the Technical Documentation department. Finally, the application will also support the first four-stroke MAN GenSet engine.

When looking into the future of engine documentation, the AR-based engine documentation will benefit from smaller, cheaper and lighter computing power and implementation of real-time augmented reality support into safety glasses. A development that makes it easier to offer an engine crew in situ online expert guidance from a shore-based location.

Currently, the solution is only available in English, but we plan to extend the choice of languages to cover the main language groups, such as Ukrainian, Chinese and others. As a next step, we are developing the functionality to extend the range of actions from maintenance to cover troubleshooting of failures or anomalies. In the future, MAN CEON TechGuide may also offer wider Internet of Things efficiency benefits.

References

MALM, L.A.; ENSTRÖM, J.; MARUSIC, M.; STÅLBERG, P. (2018), *Auxiliary Engine Damage*, The Swedish Club, pp.1-12, https://www.swedishclub.com/

Holistic Maintenance Support with a Digital Assistance System

Nikolaj Meluzov, Hamburg University of Technology, nikolaj.meluzov@tuhh.de
Robert Rost, Hamburg University of Technology, robert.rost@tuhh.de
Axel Friedewald, Hamburg University of Technology, friedewald@tuhh.de

Abstract

To ensure the technical availability of machines and equipment on board, maintenance and repair processes must be fast and error-free. In order to improve productivity for both on-board personnel and service technicians of the plant manufacturers, an Augmented Reality ecosystem has been developed that supports the preparation, execution and post-processing of maintenance tasks with targeted software modules. These include the use of the maintenance history, which is made available via an online backend, support for diagnostic findings, the display of maintenance instructions, feedback of adjustment values and documentation of the work, including storage in a digital twin. The potential estimation of industrial users showed time savings of up to 60% per maintenance action

1. Introduction

The smooth operation of ships has a direct impact on the economic success of shipowners and shipping companies. A basic prerequisite for this is the technical availability of on-board machinery and equipment. To ensure this, regular maintenance and repair work is carried out both at sea and in port. For efficient execution, the responsible maintenance personnel needs quick access to all relevant information.

Currently, employees usually have to search for information from different types of documentation. This reduces productivity and slows down the work, which is carried out either by the on-board personnel or by service technicians of the component manufacturer, depending on the specific boundary conditions. Today, the documentation of maintenance actions carried out is still characterized by handwritten /-typed notes, which are then elaborately transferred to report templates. An unsystematic evaluation considerably limits the benefit for later maintenance procedures or similar problem cases.

Mobile end devices such as tablets or smartphones, which are generally easy to use and require no specific knowledge or competence, can provide a remedy here and thus represent a suitable visualization medium for maintenance information. The connection of mobile devices to various data sources supports fast access to the relevant data. In addition, process-relevant data could be automatically recorded, documented and reported back. Current providers of digital maintenance assistance systems, however, mainly address solutions for the simple display of pdf documents and for the visualization of 3D information with Augmented Reality (AR), partly also for video-supported remote maintenance. However, a solution that addresses the entire maintenance process, can be adapted to the needs of different user qualifications and can also be used offline respectively with limited radio connections.

This article illustrates the information needs of on-board personnel and service technicians and shows how a digital AR-based assistance system platform can be designed to generate and organize maintenance information and is used on board. The support of the order-specific preparation activities and the use on site are presented and evaluated on the basis of several usecases.

2. Processes and Data of maritime Maintenance

Maintenance is divided into preventive maintenance, inspection, repair and improvement. According to *DIN35051 (2012)*, the four actions serve to maintain or restore the functional condition of technical systems.

The activities to be performed in each individual case are determined by the maintenance order and the information available to the operator or the service. A target-oriented structuring of information facilitates the access.

2.1. Information Creation

In addition to his own logbook entries, the customer essentially has access to so-called external documentation, i.e. data sheets, spare parts catalogues, (illustrated) operating and maintenance instructions and maintenance plans. The manufacturer's service technicians can also access other internal documents that describe the product, such as technical drawings, part lists and work plans. These are usually created and managed by the engineering. Based on this internal documentation, the technical editors create the external documentation. If necessary, the external documentation is extended by the service department for its own purposes. Fig.1 outlines the information release and information access of the different departments. Although the engineering is digitized to a high degree by CAD and PDM/PLM systems and 3D CAD models of the products are available, these can only be used in the technical editing department to a limited extent due to complex interface processes.

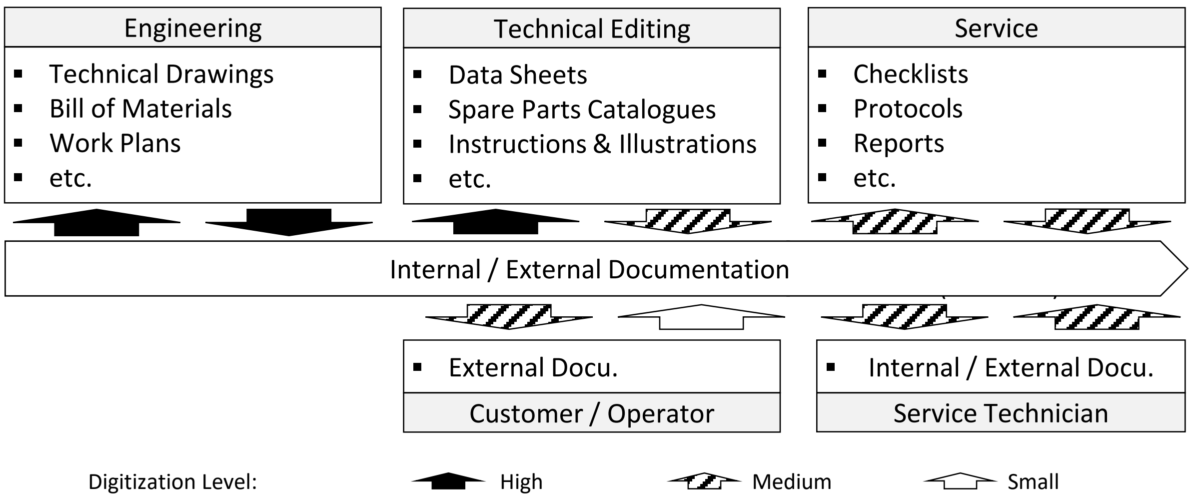


Fig.1: Information flow

The technical editors use the interactive and dynamic 3D models as templates and generate static illustrations and representations of them at great expense. Today, the descriptive texts of the instructions are created manually. Although the technical editors use pre-stored text modules, there is still a high level of post-processing effort, e.g. for the editing of texts that are not fully standardized.

2.2. Maintenance and Repair Processes

A maintenance or repair process can usually be triggered by the associated order. The process can be controlled in terms of time and content and then be invoiced. Fig.2 shows the maintenance process, which can be divided into five sub-processes, in a generalized form (*Meluzov et al. (2019) b):*

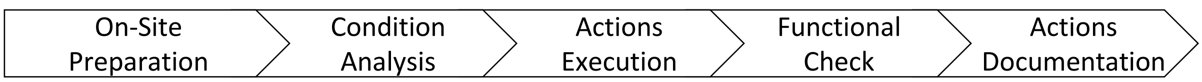


Fig.2: Generalized process representation of maintenance according to *Meluzov et al. (2019) b*

On-site preparation includes the compilation of tools and materials required for the job, but also the compilation of relevant lists and documents. The condition of the machine or plant is analyzed and documented on site. The result is a technical report and an exact estimate of the work to be carried out. After consultation with the customer and the operations management, the work is implemented and the operational readiness is checked. The work carried out as well as the corresponding modifications and adjustment values are recorded and serve as the basis for the final report.

In today's maritime maintenance, PDF files as well as paper documents are predominantly used. Due to this classic document-oriented way of working, however, the information relevant for a work step, such as information on operating materials, often has to be tediously selected from extensive instructions without hyperlinks. Even today, the maintenance reports are manually generated based on handmade notes. Although electronic report templates are available to mechanics and technicians, the transfer of notes is still mainly manual and requires a lot of time. In order to locate the data required for the maintenance sub-processes in the future, a suitable process-oriented data model is required.

2.3. Data Model

The maintenance data model must take into account both technical and organizational aspects, Fig.3. Each product has a maintenance plan that defines the tasks to be performed. Unplanned malfunctions or technical deviations are detected and corrected either by the customer or by external service providers. In both cases, a condition analysis is required to create a symptom image or technical findings as the basis for performing the actions.

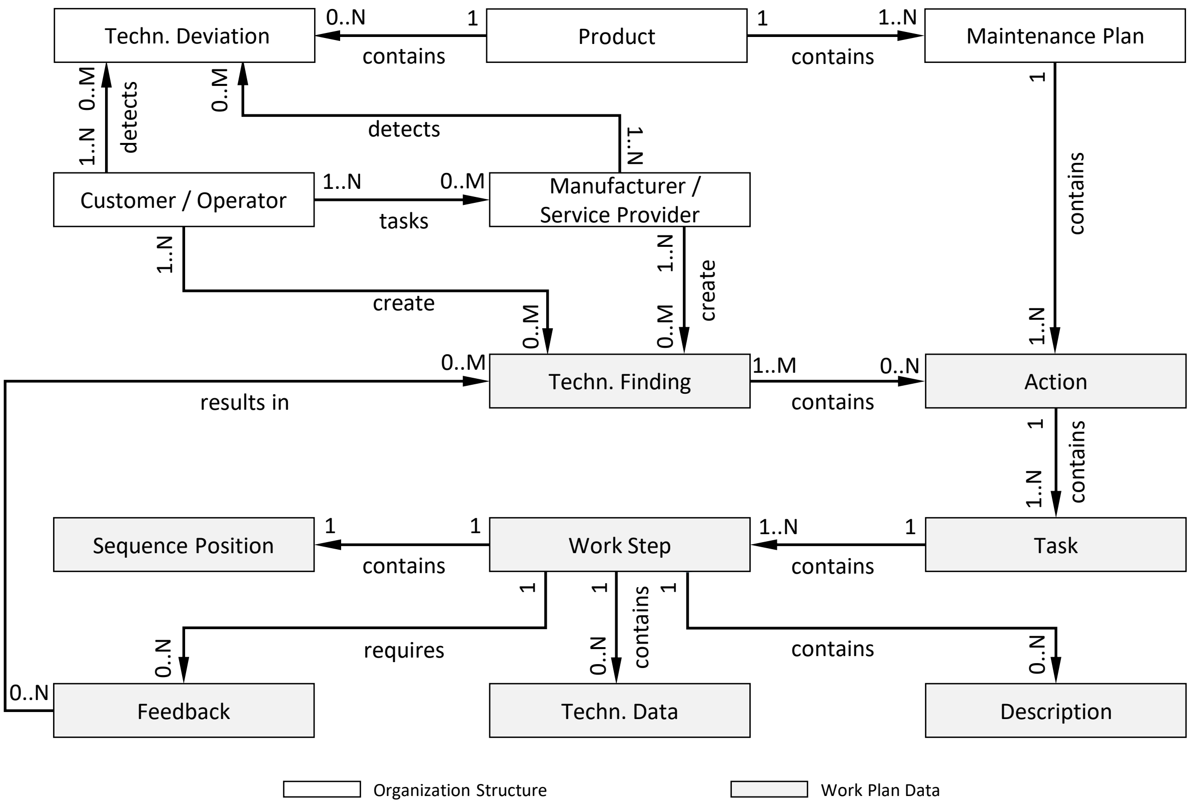


Fig.3: Schematic representation of the data model (without documentation of actions)

The maintenance actions to be carried out are derived from the technical findings, which in turn describe various maintenance or repair tasks. Each task is a sequence of individual work steps with a defined sequence position. The preceding condition analysis follows the same description system and therefore consists of different inspection tasks. Depending on the type of work step, such as dismantling or measuring, it contains varying information or requires value entries. In addition, the individual work steps include a description of the activities to be performed and the consideration of safety-relevant hints and warnings. Technical specifications, such as the correct tool or material, are stored for correct execution. Especially for measuring activities and classical checklists, the recording and feedback of defined technical values and states are required. On the one hand, these are necessary for the compilation of the technical findings, and on the other hand for the subsequent documentation. The documentation summarizes the instances of this data model, but is not shown for clarity in Fig.3.

With the knowledge about the product and process oriented data model, a holistic support of both the preparatory and the performing maintenance processes can be developed, which avoids the mentioned deficits.

3. Holistic digital Assistance System for maritime Maintenance

The support of service and maintenance is the subject of numerous applications available on the market. However, these predominantly solve only minor partial problems and do not offer a holistic approach. In the following, this problem is discussed and a suitable system structure for the maintenance platform is derived.

3.1. Mobile Applications for Service and Maintenance available on the Market

Digital assistance systems and Augmented Reality have a high potential to improve information provision and create a new user experience, according to the authors *Orsolits and Lackner (2020)* and *Langer (2020)*. These technologies are increasingly attracting attention of industrial companies and cover various application areas such as production, sales or education and training. Augmented Reality is also understood as a component of the superordinate topics: Industry 4.0 and IoT. According to *Kind et al. (2019)*, a particularly large and promising area of applications is service and maintenance.

Various mobile hardware solutions such as tablets, smartphones and data glasses are available for retrieving digital information and displaying it with AR. The use of data glasses has the advantage that the system does not have to be stored while working. The user's hands are always free and he can concentrate on the execution of the work without having to look at the distant device. However, the providers prefer to present remote support and tele-maintenance applications such as *Scope AR Remote Assistance (2020)* and *oculavis Share (2020)*. Tablets and smartphones are preferably used for the visualization of manuals and instructions, e.g. *PTC Vuforia Studio (2020)* or *RE'FLEKT ONE (2020)*. Editorial systems required for content creation are also being developed, e.g. *Scope AR Work Instructions (2020)*.

Studies by *Dini* and *Mura (2015)* on the industrial application fields of mobile assistance systems and Augmented Reality show that the maintenance of aircrafts, vehicles and facilities in particular is receiving a great deal of attention in research. In the maritime environment, there are tele-maintenance approaches by the engine manufacturer *Wärtsilä (2019)* or there is MAN Energy Solution, a supplier of large diesel engines, who has announced the first AR-based solution to visualize maintenance steps, *TechGuide (2019).*

However, both the maritime and the cross-domain solutions lack important aspects of a productivity-enhancing overall solution.

In the following, the requirements for back-office process support and for assistance systems aimed at on-board personnel and service technicians are explained. Furthermore, the development of a digital assistance system platform will be derived.

3.2 Requirements for a holistic Solution

In order to be able to meet the above-mentioned challenges and to benefit from comprehensive productivity improvements, the concepts for the use of digital assistance systems must be adapted and expanded to the specific maintenance and repair processes and the corresponding information needs:

- Easily understandable visualizations of the maintenance steps must be made available to the on-board personnel, thus supporting the execution of actions in particular
- For the service technicians, above all, the preparation of the assignment and the documentation of actions should be made less time-consuming.
- The devices and software must be accepted by the users for permanent use

- A data-saving online and offline connection must be possible, also for tele-support
- The creation of the maintenance information must avoid a tedious transfer or re-creation of the digital contents with a suitable converter solution
- The entire system must be able to be integrated into the company IT and the standard processes of order processing

To meet these requirements, a platform, which offers process- and user-adapted hardware and devices, was derived. This platform provides data or information, offline and online, via a uniform exchange format, while different user roles are taken into account.

3.3 System Structure

On the basis of the data model presented in Section 2.3, a comprehensive, component-based system for maintenance support has been developed, which supports the consistent processing of all the indirect and direct (organizing and operating) processes described. The structure of the system is divided into backend and frontend modules, Fig.4.

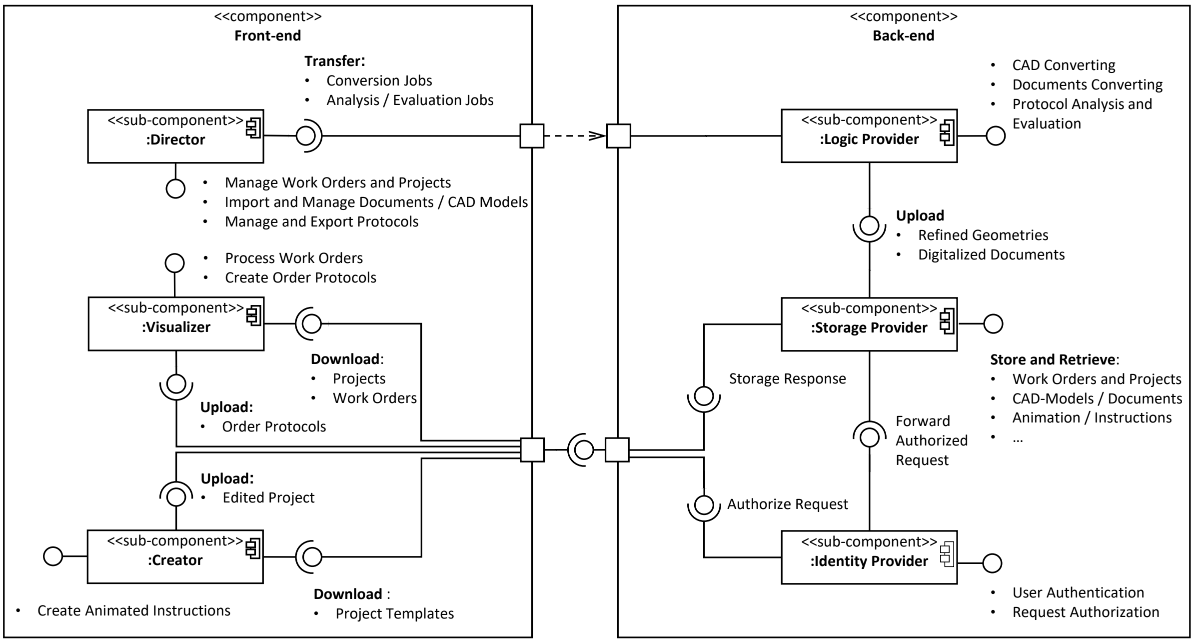


Fig.4: Overview of the system components

The base of the system is the backend, which is mainly responsible for the management and storage of digital content. The backend is built up from the three provider components: Storage, Logic and Identity. The content is supplied by the storage provider. This is a dynamic storage medium that grants controlled access to the project data and delivers it to the devices. The Logic Provider is used to import and convert the CAD models and documents. It is controlled via a front-end component, the Director. In addition, the Logic Provider forms the basis for the connection of further services, such as the evaluation of findings data or the connection of other systems (e.g. ERP, MES, PDM, PLM, CMS). The administration of user roles and the control of the authentication of the requested access rights for orders, projects, instructions and CAD models is the task of the Identity Provider.

The user interface for the different end users from the organizing and operating areas is the front-end with its three components Creator, Visualizer and Director. The creation or transfer of requests - the configuration of the import from the ERP system - is carried out using the Director. To generate the order-related instructions dynamically and interactively, the Director is used to load the corresponding CAD models or create corresponding references to the PDM / PLM system. The projects organize the instructions and take into account specific products or product variants. Existing instructions (e.g.

available as .xml / .html or .docx) can be loaded into a project via the Director and automatically be converted into the internal data format by individual processing settings. Step-by-step maintenance instructions are created in the Creator, if necessary from created templates, similar to a classic editorial system. After uploading the digital manuals, they can be displayed on the Visualizer together with the orders. The automatically generated process protocols with condition and value feedback as well as the findings data are synchronized with the back-end according to the corresponding project and are available for analysis in the Director. All components act autonomously and can be used for a longer period of time without connection. If a connection exists, the project contents are synchronized and are available in the other components accordingly.

In the following, the use of the system for the necessary back-office processes, in particular the creation of documents for maintenance execution, is described.

4. Back-Office Processes

Before the operative maintenance tasks can be carried out with the help of a digital assistance system, the necessary information must be generated in the technical editing department. It must also be possible to define user roles for these preparatory activities and to transfer CAD and metadata from the company's IT systems.

4.1 User Administration and Access Rights

Users, roles and access rights are basically controlled in the Identity Provider component. This is a generic component that follows the widely used OpenID Connect workflow. This makes it possible to replace the component with an already existing, in-house solution and results in users being able to log on to all components of the platform with their familiar companywide user accounts.

The security issues and existing processes relevant to the terminal equipment used to carry out maintenance work, e.g. for handling guest accounts for on-board mechanics, are therefore not affected. If a user logs on to one of the systems with an existing server connection, he is first registered in the respective device, the currently valid authorizations are stored and general access is protected by a PIN. The user now sees all projects with their respective data, is informed about necessary synchronization processes and can carry them out. All project data is then downloaded or updated on the device. If the same user logs on to the device with the PIN at a later time and without an Internet connection, he can work with the data available offline on the device according to the stored access rights. Offline use is only possible for a defined period of time (ideally based on the order time including pre and post-processing) and prevents access after the period of time has expired.

In addition to controlled data access, all feedback values can be provided with the respective user ID by the user login. This makes it much easier to trace the work later on for documentation purposes and can considerably improve coordination when several mechanics and technicians work simultaneously.

Furthermore, the role-based model allows the system to react to the respective user and enables functions necessary for the content creation by the technical editors. The system can also show or hide certain data for the operative maintenance, depending on the user's qualification (on-board mechanic or service technician). This considerably facilitates the separation of internal and external documentation, *Friedewald et al. (2019).*

Since all those involved in the process have to access existing data when carrying out their tasks, another essential aspect for productive system use is the connection to the company's IT, which is explained in section 4.4.

4.3 Project and Order Structuring

The interaction of necessary information is organized on a higher level by the Director module. It serves as a link between the Creator and the Visualizer. Fig.5 shows the Director interface. All data is combined in projects (1). The sidebar menu is used to prepare the project-relevant data (2).

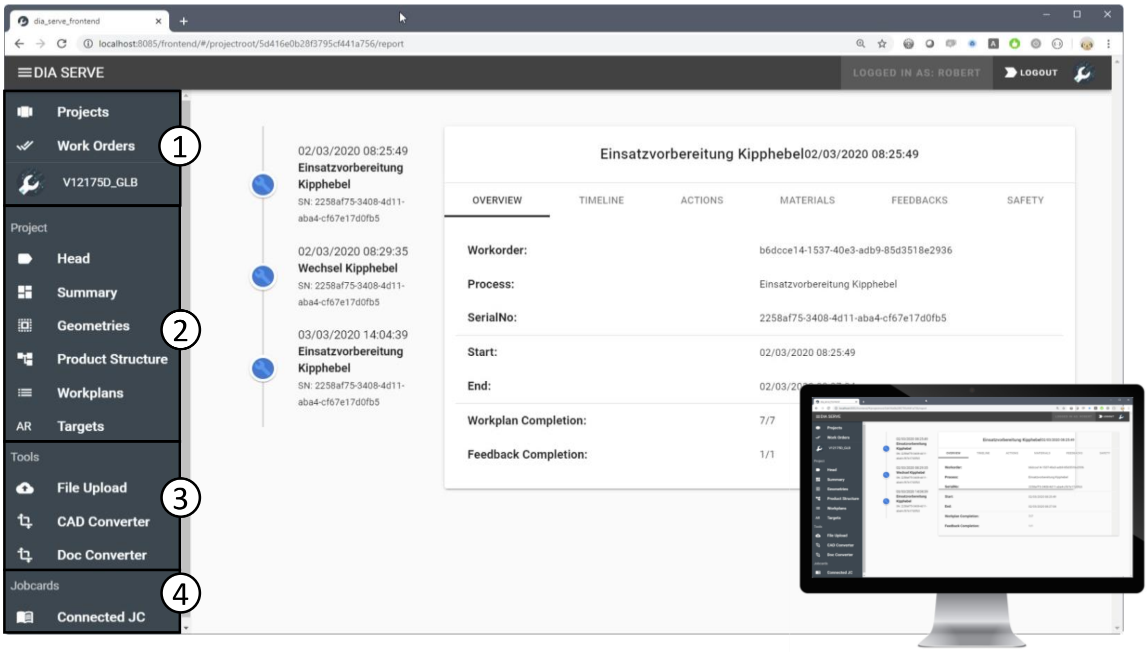


Fig.5: Director interface

This includes the 3D geometries, the parts list structure, the instructions and the tracking targets required for the AR visualization. With the help of the import functionality (3), the CAD models are converted. Existing instructions can also be converted into the internal data format or the created instructions using the import functionality. If the product or product variant has already been assigned to an order, this is also immediately visible (4). When all information is available, the creation of digital maintenance channel lines can be started.

4.2 Connection to the Company IT

Basically, the platform is designed in such a way that it can be used independently. This means that orders, projects, CAD models and instructions can be created, managed and deleted with the help of the Director. Since at least parts of the data (e.g. orders) are usually stored in other systems (ERP) and created, managed and deleted according to internal regulations, this approach would lead to redundancies.

To prevent this, it is possible to use the logic components of the back-end to set up routines for synchronizing the respective data and to deactivate the corresponding components in the Director (e.g. order management). As a result, this task is performed solely by the connected subsystem and redundancy is avoided. The situation is similar for CAD models, although an additional step is unavoidable: The models are traditionally available in parameterized and usually proprietary form in the PDM or PLM system. Since the Visualizer component is designed for use on mobile devices with comparatively low performance capacity, these models cannot be loaded directly by the application. In a pre-processing step, the models are therefore analyzed and their quality is flexibly reduced to prepare them for mobile use. However, the basic model can be obtained from the PDM / PLM system and only the converted derivative of it can be stored in the storage component. In this case, the application remembers the origin of the data set and provides an interface for PDM / PLM systems to report changes to this origin, after which the conversion is triggered again. Alternatively, query routines can be set according to which the logic component proactively queries data statuses in the subsystems at certain intervals and triggers any updates.

4.4 Preparation of digital Instructions

Digital maintenance (AR-)instructions can contain static elements such as textual information, adjustment data and references to geometric elements on the one hand and dynamic elements, i.e. animations, on the other hand. For an uncomplicated creation that takes into account the process requirements of technical editing, an authoring system (Creator) was implemented that supports sub-processes and the creation of static and dynamic information in the best possible way.

Since the basic tasks in the creation of instructions are primarily of editing and assigning character, a desktop application was initially developed for the technical implementation of the Creator. Fig.6 shows the user interface of the Desktop Creator, consisting of three areas: Product Structure (left), 3D Model Viewer (middle) and Work Plan Menu (right).

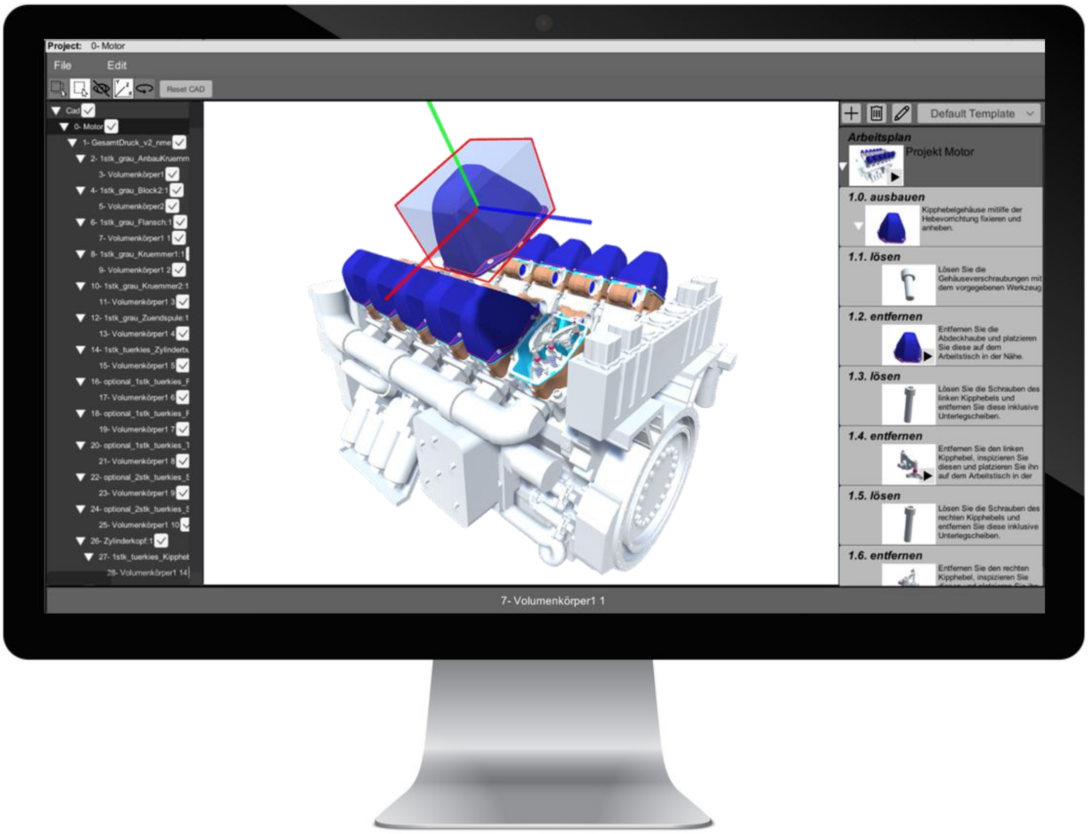


Fig.6: Desktop Creator according to *Meluzov et al. (2019a)*

The Product Structure shows technical editors the name of the components and their hierarchic arrangement. In the 3D Model Viewer, the editor sees the geometric model and can rotate, move and zoom it. Parts can be easily selected in the viewer with a mouse click and assigned to a work step in the Work Plan Menu. Using the selection menu, the work steps can be enriched with organizational and safety-relevant hints, tools, materials and technical specifications. Animations of the 3D geometries can be created and saved to illustrate the replacement and dismantling tasks.

Another front-end module was developed for the creation of complex animations. Studies on Virtual Reality (VR) show that these immersive systems offer many advantages over classic 3D systems: a simple and intuitive interaction with the 3D model as well as a clear representation of the 3D environment. Therefore, the Creator was also implemented in a Virtual Reality environment. In order to be able to easily create complex and elaborate animations, a controller-based approach was chosen.

The Virtual Reality Creator is an extension of the desktop application and enables the user to edit the projects in an immersive editorial environment, *Meluzov et al. (2019a)*. With the help of the controllers, the user operates the selection menu and interacts with the 3D model. The selection menu provides the user with various tools, such as freehand animation, to illustrate particularly complex disassembly and assembly processes. A component movement is recorded with the controller and assigned to the corresponding work step as an animation and saved as an extension of the geometry Meta data. The freehand movement is then automatically smoothed.

In comparison to conventional CAD programs, this eliminates the time-consuming moving of component axes and pivot points as well as the creation of keyframe-based animation sequences. Once the maintenance instructions have been stored, they are available for operational support as a data set and can be used with various front-ends / hardware alternatives, as illustrated below.

5. Front-Ends for Maintenance Processes

The aim of the digital assistance system is to support on-board personnel and service technicians in carrying out the maintenance processes described in 2.2 by showing relevant information directly on the display and eliminating the need for tedious searches.

5.1 On-Site Preparation and Condition Analysis

The on-site preparations of the manufacturer's service technician consists of preparatory activities at headquarters such as the compilation of documents, preventive spare parts planning and travel organization, as well as the analysis of the machine condition to complete customer's statements on site. The on-board personnel will familiarize themselves with the upcoming work steps. Both processes are supported by the digital assistance system. Fig.7 shows the user interface of the mobile tablet application (Visualizer) preferred by users.

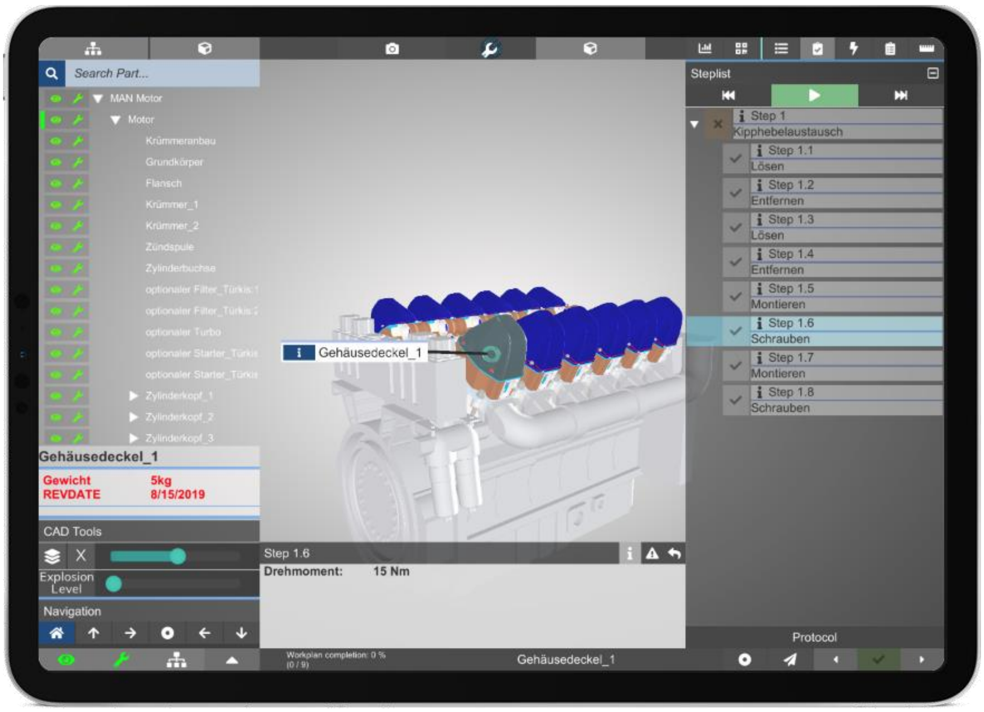


Fig.7: Overview of the user interface of the Visualizer

It is divided into three areas: The Work Plan Menu (right) shows step-by-step instructions created with the Creator, if required. The step-by-step instructions can also provide checklist functionality and include value and condition feedback for testing and inspection tasks. For simple operation, the work steps and animations can be chosen via the controller menu or be played as whole sequence. This clarifies the entire task at the beginning and gives an overview of the work to be done.

The work plan is linked to the 3D Menu (middle) and shows the corresponding CAD models and animations for a selected work step. If required, the display of the models can be switched from the CAD viewer to an augmented reality display.

The Product Structure (left) shows the 3D model structure and enables the display of component and assembly information via a list selection.

The aim of the on-site preparation is to determine the measures or work to be carried out. Although the customer order usually contains an initial symptom picture, this must be expanded by a refined diagnosis of the machines and systems on site. The Visualizer provides the appropriate check and inspection lists for this purpose. The Work Plan Menu guides the user through the inspection points of the task and queries the values and condition in order to report them. Fig.8 shows the user interface of the Visualizer during the technical diagnosis.

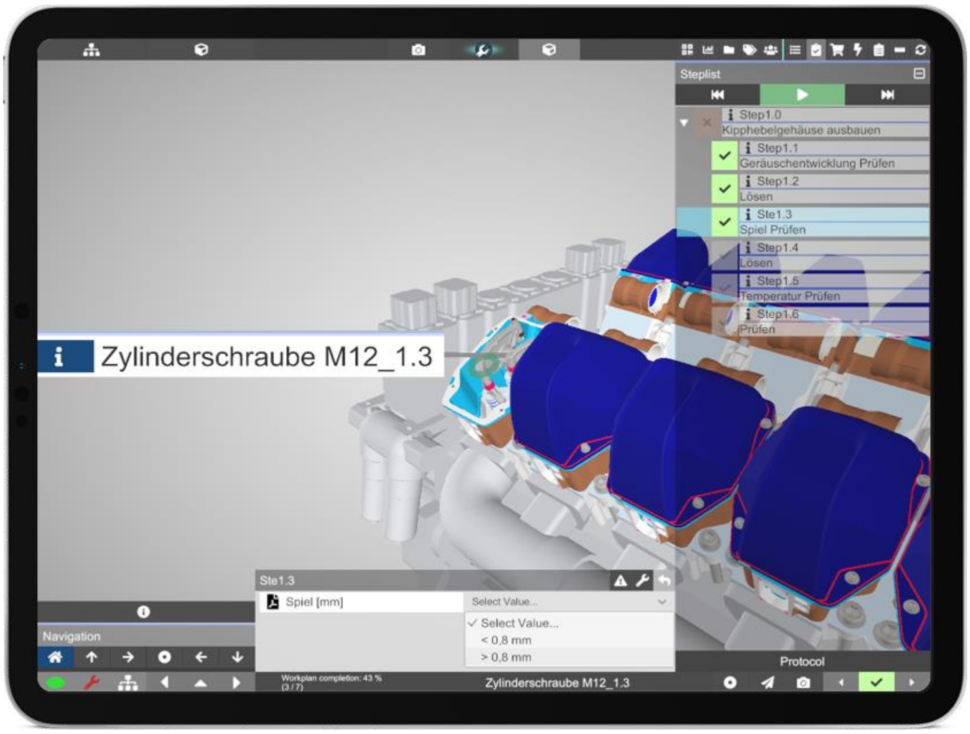


Fig.8: Value and condition report with the Visualizer

To not disregard the order of the inspection points, after a complete inspection task the work step is marked and reported as completed. The attributes are recorded in the process log. The hereby recorded state can be called up later on to complete the work either by the same or by another mechanic or technician.

For a better illustration, the affected parts or components are highlighted and displayed for each work step. In the case of testing tasks on electrical components, for example, these can be the connection points of the electrical cabinet. If no CAD models are available, they can be replaced by sections from technical drawings or illustrations. Depending on the task, the response is either a numerical value, a condition selection or a textual comment. Photos can also be created for extended documentation. The result is a digital meeting protocol with all recorded data.

5.2 Action Execution

In order to be able to carry out the necessary work efficiently, the assistance system also provides animated instructions from the technical documentation via the Work Plan Menu. The animated instructions are mainly used by the on-board personnel and are not displayed for experienced service technicians. The contents of the work step descriptions no longer have to be searched out in a tedious manner (Section 2.2). Instead, they can be called up directly via the menu. Fig.9 shows the selection options of the work step-related information.

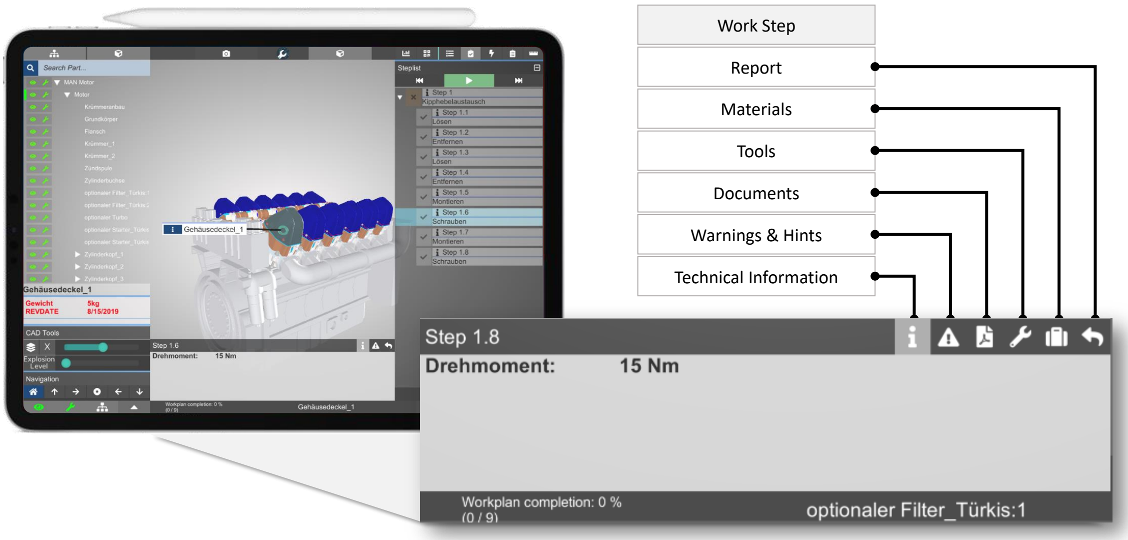


Fig.9: Digital work plans on the Visualizer

The information includes technical data and classic table values, such as tightening torques or other adjustment values. An integrated PDF viewer can also be used to display manufacturer and supplier documents. Relevant tools and materials to be used are also called up via the menu, as well as original spare parts from the manufacturers. The purchase of these spare parts can be triggered by a digital order form.

As a rule, a work task is preceded by safety-relevant or organizational hints. These are displayed as pop-ups on the assistance system and require a reading confirmation before further information is displayed. This ensures that the operator's maintenance personnel is aware of the hints and warnings.

5.3 Documentation

The users, during the technical findings, the condition analysis as well as during the execution of the actions, reports data directly and indirectly. The direct data feedback includes the value and condition report of according the machines during the plant inspection. The indirect data feedback describes the status feedback of the executed work steps from the instructions, as well as the confirmation of the read notes or warnings within the scope of the work. With the help of the Visualizer, a digital protocol is created which, with an available Internet connection, can be made available directly for evaluation in work preparation via the AR platform.

5.4 Alternative Devices

In order to meet different user requirements, the assistance system was also published for a smartphone application and AR-Glasses, Fig.10. In particular, numerous shipbuilding users in the equipment assembly department, for example, have requested the use of a smartphone instead of a tablet. In addition to scaling the user interface, a modified portrait user interface was developed, *Jahn et al. (2020)*, which corresponds to the usual smartphone use and which is currently undergoing usability tests by various users. In contrast, the developed HoloLens application, which has the advantage of hands-free use and can be used for tele-maintenance like the other end devices, was predominantly evaluated negatively in usability studies, as the duration of use and interaction possibilities are not satisfactory from the user's point of view and the use with a helmet is not possible in a comfortable way. Therefore, the tablet use in practice will be considered in the following.

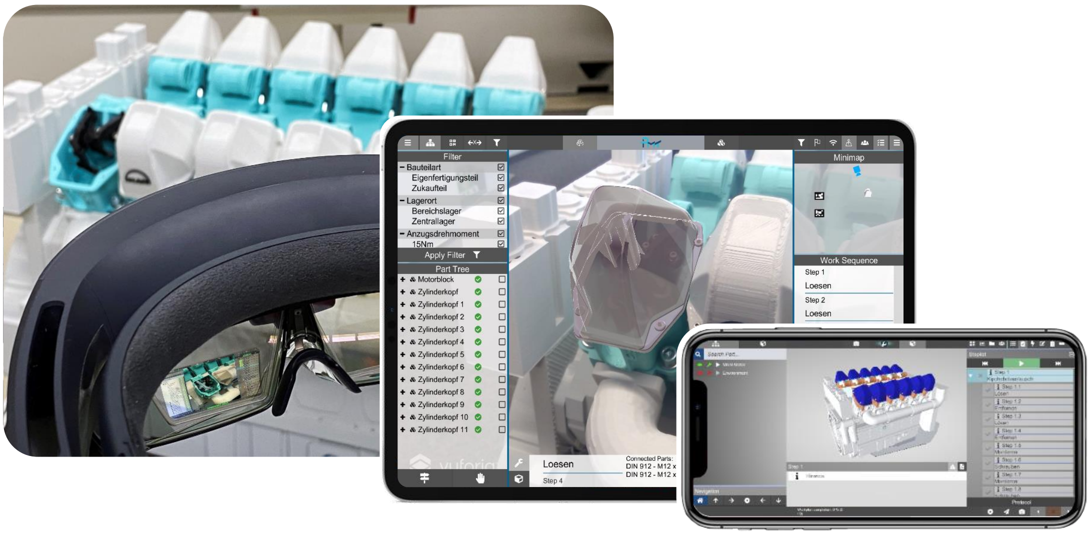


Fig.10: Available devices

6. Practical Usage

In order to be able to estimate the potential of digital information supply in practice, the assistance system was used in two companies on a trial basis. For this purpose, two service technicians and two mechanics each worked with the Visualizer during a two-day maintenance and repair assignment in addition to the conventional paper-based documents. Fig.11 shows the software in practice using the example of a control cabinet maintenance (left) and an engine repair (right).

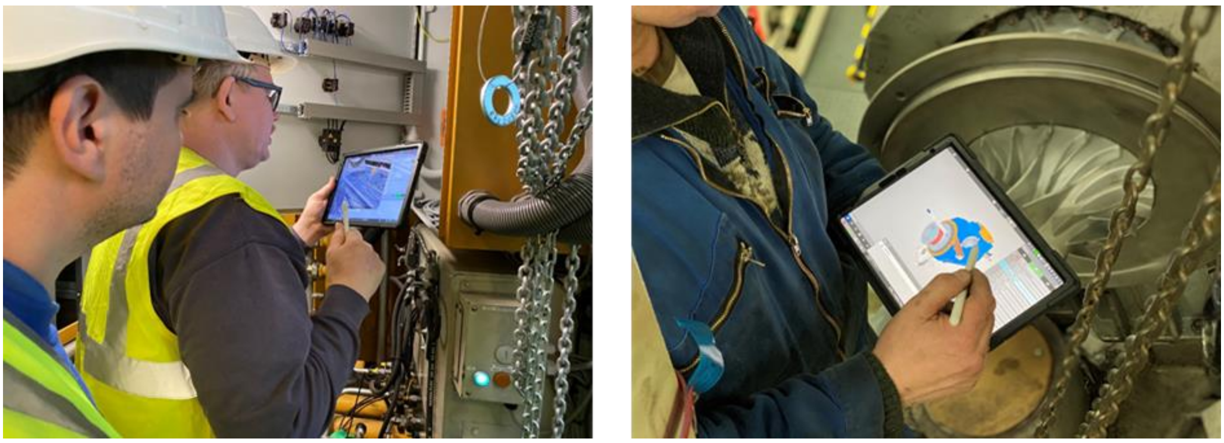


Fig.11: Visualizer in use

6.1 Productivity Studies

In order to examine productivity, the working process was first of all recorded with the time allocation. Fig.12 shows the distribution of the total time (2 days) on site to the individual phases of maintenance, based on an 8h working day.

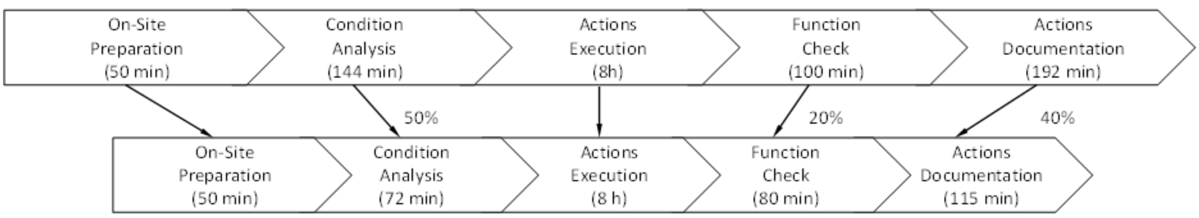


Fig.12: Time savings in the work process

The highest proportion of working time is spent on the execution of actions. In the example of the accompanied switch cabinet maintenance of the cooperating partner company, this involves the direct execution of small preventive activities, such as the retightening of loose clamp connections or the insulation of free-standing cable cores. The procedure for executing the activities was known to the employees and it was not necessary to look up instructions, rules or regulations, similar to the deployment preparation, in which all the aids and tools required for the measures were brought to the site.

The time savings using the example of switch cabinet maintenance were determined during the status analysis. The assistance system supports the service technicians in documenting test values with the help of a digital checklist and direct input fields for the actual values. Along the test points, remarks and technical findings can be stored directly in writing or noted as virtual annotations, *Rost et. al (2018)*. Fig.13 shows the corresponding user interface for both application cases.

Fig.13: Visualizer user interface for inspection actions (left) and virtual annotations (right)

Similar to the checklists from the condition analysis, the function check supports the users in checking the documentation of the plant functionality. For this purpose, the corresponding measuring points for electrical signals and values as well as the appropriate input fields are displayed. The recorded values and statuses are documented directly in the log and no longer need to be written down by hand. A final report or extract for the customer is then automatically generated from the log, so that the documentation of measures can also be significantly reduced compared to conventional processes.

Initial estimates indicate that in particular the condition analysis (reduction to 50% of working time) and the documentation of measures (reduction to 60% of working time) will be significantly improved. Due to the increased complexity of engine maintenance, much higher savings are expected, especially in preparatory and follow-up activities.

Following the practical applications, the user satisfaction was recorded in workshops.

6.2 User Satisfaction

For a successful industrial application of digital assistance systems, user acceptance is of great importance in addition to the productivity effect. Fig.13 shows the evaluation of usability following the practical application using the example of control cabinet maintenance.

The users evaluated the assistance system in a questionnaire with a scale of -2 to +2 with regard to user-friendliness in handling the software, completeness of the functional scope and applicability, taking into account industrial boundary conditions and environmental circumstances.

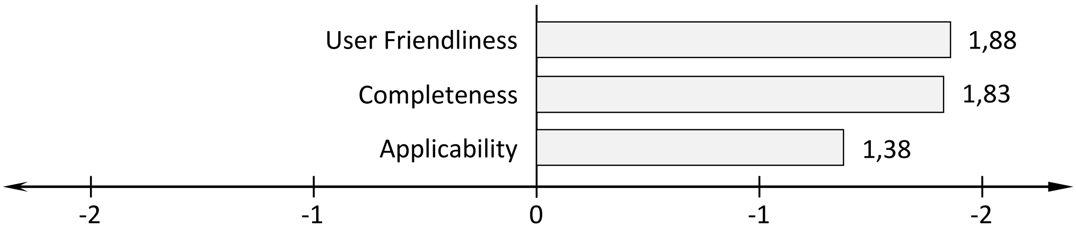


Fig.13: User satisfaction with the Visualizer

In the evaluation, the application scored very high. The intuitive operation of the software on the tablet and the user interface were particularly positively noted. The scope of the provided software functionality also opens up the information needs of the users and can be transferred to further maintenance scenarios beyond the tested use case. There was a slight point deduction for usability in cramped environments or plant rooms that can only be entered by crawling.

7. Summary

To support maritime maintenance, a comprehensive digital assistance system platform with Augmented Reality support was developed, which supports both the upstream work of information generation and the execution of maintenance. The integration into the company IT and the user-specific facilitation of the individual subtasks of maintenance contribute to an increase in productivity. Both the ongoing productivity studies and the usability tests show considerable improvements compared to the current way of working.

Acknowledgements

The presented work was done in cooperation with maritime suppliers within the joint research project WASSER, funded by the German Federal Ministry of Economics and Energy due to a decision of the German Bundestag.

References

DIN 31051 (2012), *Grundlagen der Instandhaltung*, Beuth-Verlag

DINI, G.; DALLE MURA, D. (2015), *Application of Augmented Reality Techniques in Through-life Engineering Services*, Procedia CIRP 38, pp.14-23

FRIEDEWALD, A.; ROST, R.; MELUZOV, N.; SCHRÖDER, H.; LÖDDING, H. (2019), *Weiterbildung mit digitalen Wartungsassistenzsystemen*, Wissenschaftliche Gesellschaft für Arbeits- und Betriebsorganisation (WGAB) e.V., GITO-Verlag, pp.127-142

KIND, S.; FERDINAND, J. P.; JETZKE, T.; RICHTER, S.; WEIDE, S. (2019), *Virtual und Augmented Reality*, Büro für Technikfolgen-Abschätzung beim Deutschen Bundestag

LANGER, E. (2020), *Medieninnovationen AR und VR*, Springer Vieweg

MELUZOV, N.; FRIEDEWALD, A.; MUNDT, C.; LÖDDING, H. (2019), *Produktivitätssteigerung in der Instandhaltung durch digitale Assistenzsysteme*, Digitalisierte Instandhaltung, Stand und Perspektiven, TÜV-Verlag, pp.181-197

MELUZOV, N.; FRIEDEWALD, A.; ELZALABANY, A.; LÖDDING, H. (2019), *Aufwandsarme Erstellung von Augmented-Reality-Anleitungen für die maritime Instandhaltung*, Go-3D, pp.31-44

OCULAVIS SHARE (2020), *Remote Maintenance für die produzierende Industrie*, www.oculavis.de

ORSOLITS, H.; LACKNER, M. (2020), *Virtual Reality und Augmented Reality in der Digitalen Produktion*, Springer

PTC VUFORIA STUDIO (2020), *Innovationen mit Industrial Augmented Reality vorantreiben*, www.ptc.com/de/products/augmented-reality

RE`FLECT ONE (2020), *Augmented Reality Plattform für Industrie*, www.re-flekt.com/de/rf-one

ROST, R.; JAHN, N.; FRIEDEWALD, A. (2018), *Smart Inspection: Documenting Issues in 3D with Augmented Reality*, 18th International Conference on Construction Applications of Virtual Reality (CONVR 18) Proceedings, Auckland, S. 311-320

SCOPE AR REMOTE ASSISTANCE (2020), *Augmented Reality Remote Assistance and Support Solutions*, www.scopear.com/solutions/ar-remote-assistance

SCOPE AR WORK INSTRUCTIONS (2020), *Take the work out of work instructions*, www.scopear.com/solutions/work-instructions

WÄRTSILÄ (2019), *Wärtsilä successfully tests remote guidance service capabilities*, https://www.wartsila.com/media/news/15-03-2019-wartsila-successfully-tests-remote-guidance-service-capabilities-2401858

TECHGUIDE (2019), *MAN CEON TechGuide Introduces First Augmented-Reality Maintenance Platform*, https://www.man-es.com/company/press-releases/press-details/2019/12/10/man-ceon-techguide-introduces-first-augmented-reality-maintenance-platform

Modelling of Flexibility of Shipbuilding Work by Production Simulation

Kohei Matsuo, National Maritime Research Institute, Tokyo/Japan, kohei@m.mpat.go.jp
Fumiaki Tanigawa, ASAKAWA Shipbuilding Co. Ltd., Imabari/Japan, f-tanigawa@asazo.com

Abstract

The value-added work ratio in shipbuilding is less than about 40%. The remaining 60% is incidental work and non-value-added activities. It is difficult to express incidental work and non-value-added activities correctly, as these depend on flexible behaviours by workers. The purpose of this research is to determine production plans or work procedures at the operation level with high accuracy. The production order of sub-assembly or the operation procedure (the order of attachment of parts in one sub-assembly) is determined by the production simulation. More specifically, a production simulation is performed for all patterns of work procedure, an optimum solution is output as a production plan or a work procedure.

1. Introduction

Generally, work can be classified into value-added work, incidental work, and non-value added activity in production management. Value-added work is the net work that takes a product closer to its completion. In the case of a welding operation, it is when the parts are just joined by welding. The value-added work adds value to the product. The incidental work is a work necessary for performing the value-added work, but does not directly add value to the product. In the case of a welding operation, incidental work corresponds to a work such as preparing a welding device or cleaning a welding portion in advance. Although it is necessary work, this should be reduced as much as possible by reviewing work procedures or preparing special jigs. Non-value-added activity is unnecessary waiting or unplanned breaks. The value-added work ratio in shipbuilding is less than about 40%. The remaining 60% of the work is incidental work and non-value-added activities. The following are problems on a work in shipbuilding:

- Low value-added work ratio and low productivity
- Difficulty setting reasonable standard work and standard time
- Difficulty establishing reasonable cost management and production management
- Shortage of workers and difficulty developing skilled worker

Although these problems have not yet been solved fundamentally, it can be recognized that the causes of these problems are hidden in incidental work and non-value added activity that account for 60%. Most of the incidental work and non-value-added activity depend on the human factor, and it is difficult to express them correctly. Therefore, they are left to humans. They are ambiguous because they rely on the agile behaviour of the workers. Incidental work and non-value-added activity are the kind of work can be longer or shorter, and there exists “smart way” depending on the worker’s movement. For this reason, work does not proceed as planned in the shipbuilding industry. In addition, the flexible behaviour is inherent in the individual worker, and is hard to be specified outside. This creates a skill gap. On the other hand, the value-added work is based on physical phenomena, such as cutting, bending, welding, painting, etc., and can be handled analytically. (Actually, most simulation software such as for welding deformation analysis targets the value-added work in welding). It is said that the value-added work ratio of the automobile industry is about 90%. Thorough Kaizen activities reduce incidental work and non-value-added activity and leave no room for workers to be flexible. Alternatively, they analyse incidental work and non-value-added activity to minimize worker’s flexibility.

In order to solve following fundamental issues that shipbuilding has for many years: increasing the value-added work ratio (improving productivity), setting standard work and standard time (establishing a rational production management system), eliminating the variation in skills of worker, this study focus on human factors especially included in incidental work and non-value-added activity. By scientifically elucidating human factors in shipbuilding work, the authors aim to eliminate the ambiguity of shipbuilding work and express the shipbuilder's “smart way”. This study finally aims at shipbuilding production such that “all are numerically represented, all are numerically planned, and all are completed as planned”. The authors define this as a “digital shipyard”. In recent years, innovative ICT technologies, such as IoT and AI technologies, have come to a practical stage. By effectively utilizing these, we aim to quantify the shipbuilding work composed of human factor. This is the digitalization of shipbuilding.

2. Flexibility in shipbuilding work

A “digital shipyard” in this study is defined as a shipyard that “all are numerically represented, all are numerically planned, and all are completed as planned”. The work in shipbuilding is not carried out according to the numerical plan at present, because the work by a worker cannot be definitively expressed. This is mainly due to the fact that the behaviour of the worker is “flexible”. This study tries to clarify the mechanism of flexibility of human work in shipbuilding to eliminate ambiguity. This study will elucidate the mechanism of human work in shipbuilding and digitize it numerically, rather than expressing work as a macro or expressing work like a probability distribution.

In this study, the flexibility of human work is classified into two types according to the granularity of the behaviour of the worker, and elucidation of the mechanisms of these two types of flexibility is conducted by different approaches. Fig.1 shows the structure of shipbuilding work by granularity. In principle, works at the work level (e.g. the production order of sub-assembly) or the operation level (e.g. the order of attachment of parts in one sub-assembly) can be handled in a predetermined manner in principle by analysing them carefully once the product is determined. In other words, it is an issue to output production planning or work procedure manuals, and work optimization at the work level or the operation level. Some shipyards have already dealt with them. However, in general, this predetermined production plan is not sufficiently accurate. The reason is that the parameters to be considered are enormous, the degree of influence and correlation of the parameters are unknown, and exceed the human ability for consideration. As a result, the current situation leaves this to a shop floor, and veterans replace this by using advanced experience and consider the efficient work procedures based on their own experience.

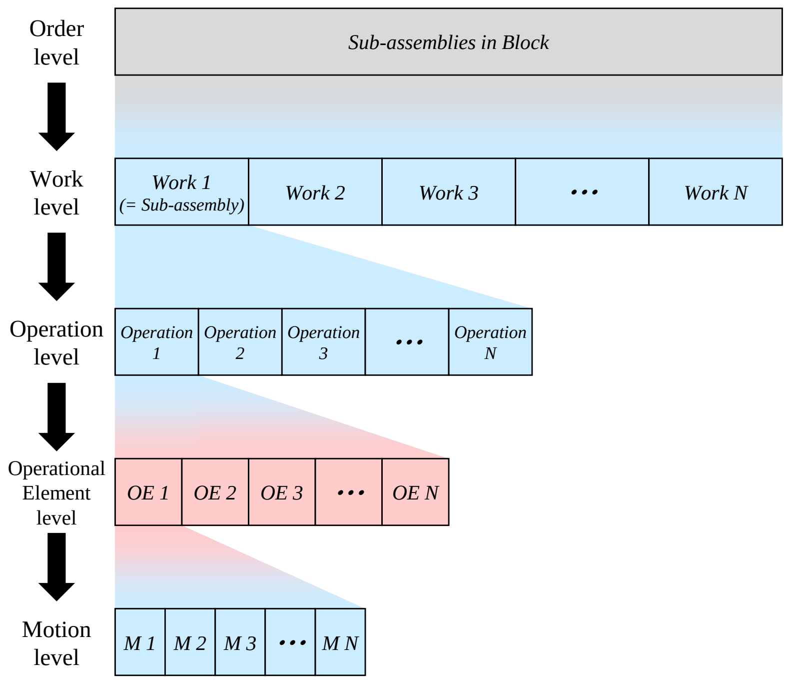


Fig.1: Structure of shipbuilding work by granularity

However, this can be handled in a predetermined manner in nature. In this paper, those that can be handled in a predetermined manner, but are left to the workers as a result and remain the flexibility of workers are defined as “flexibility as a result”. Elucidating the “flexibility as a result” means accurately obtaining a production plan and a work procedure at a work level and an operation level. The veteran's smart way should be considered in advance at the design stage, not at the shop floor.

On the other hand, even if the work is performed in accordance with a production plan or work procedure that is determined in advance, all work cannot be planned in advance, and it may be necessary to be flexible. In other words, it is an ad hoc event that can only know at that time. This paper defines this flexibility as “flexibility as an individual”. The “flexibility as an individual” is explained with an example of a sub-assembly process, since the quality variation of the parts to be attached differ from case by case, the worker responds this variation at the operational element level. As described above, what cannot be handled in advance according to the individual variation and is left to the worker each time is “flexibility as an individual”.

It should be noted that the “flexibility as a result” and the “flexibility as an individual” are not independent, but are interrelated. If the area of the predetermined production plan is increased, the area of the flexibility due to the individual variation is reduced accordingly. If all shipbuilding operations can be planned in advance even to the motion level, all individual variation can be handled in a predetermined manner (in other words, individual variation will be eliminated at this stage).

3. Production simulation for predetermined plan

Responding to the “flexibility as a result” is to accurately determine the production plan and work procedure at the work level or the operation level. For this, a production simulation on a computer deductively outputs a production plan (production order of sub-assemblies) at the work level and optimizes an operation procedure (attaching order of parts in one sub-assembly) at the operation level. Specifically, a production simulation is performed for all patterns relating to the order pf operation, and an optimum operation is searched for.

This paper introduces a case study of production simulation for a production planning problem in a sub-assembly process at the work level. In this simulation, the efficient preparation of tools is focused as an expression of flexibility of the worker, and confirm the elucidation of the mechanism of occurrence of the “flexibility as a result” caused by this.

3.1. Overview of production simulation

The production simulation targets the sub-assembly process. The workflow of the target sub-assembly process includes the distribution of the fitting parts by the pre-fitting operator, the pre-fitting of the parts by the pre-fitting operator, and the welding operation for the pre-fitting parts by the welding operator as shown in Fig.2.

Production simulation is conducted for the production of 10 sub-assemblies sets according to the above work flow. The working parameters in the simulation are set for each sub-assembly. The following parameters were set for each sub-assembly and part as working parameters:

- Number of parts to be fitted
- Distribution method (by hand or by crane) for each part
- Time required for distribution
- Tools required for pre-fitting
- Time required for pre-fitting
- Tools required for welding
- Time required for welding

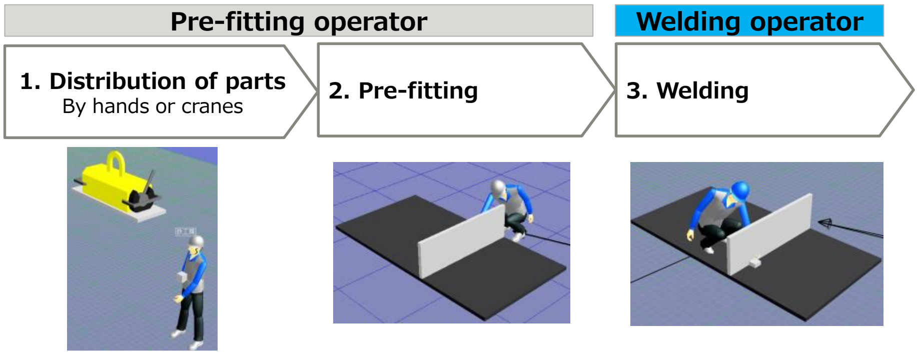


Fig.2: Workflow of sub-assembly process

Production simulation is performed by discrete event simulation. Discrete event simulation models a production process, and outputs a production time, an operation rate, and so on by inputting working parameter. Production simulations were performed using OPTEMILIS and GP4 by Fujitsu in this paper. OPTEMILIS is a production simulator that reproduces factory production, and has a built-in optimization engine that optimizes the work procedure. GP4 visualizes the calculation result by OPTEMILIS as a production flow on the factory layout. Fig.3 shows an image of the GP4 screen.

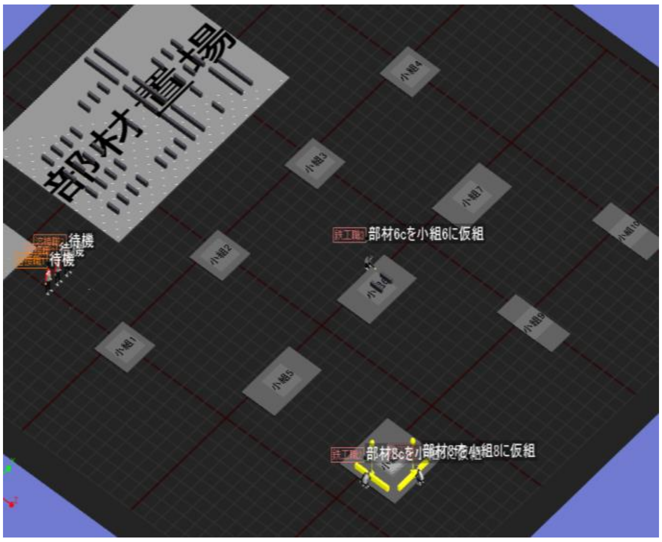


Fig.3: Image of the GP4 screen

The conditions of the production simulation are as follows:

- Two types of workers are set as pre-fitting operators and welding operators. In this simulation, three pre-fitting operators and three welding operators are set. The welding operator cannot perform the welding work until the pre-fitting work to a certain sub-assembly is completed.
- Prepare necessary tools before starting each work. Tools required for each work are set as working parameters. It takes 3 minutes to get the tools and carry them to each sub-assembly. However, one person can carry up to two tools at a time. In other words, if operators need three tools, the required time is 6 minutes (3 minutes x 2 round trips).
- The production order of the sub-assembly by each operator varies randomly. In this case study, we focus on how the preparation time for tools varies with the order of production. If the tools needed for the next work and the tools used in the current work do not overlap, those tools will be left in the sub-assembly currently being processed. Thereby, the operation required for tool preparation changes depending on the processing order, and the total time for tool preparation in the entire work may fluctuate. The tool preparation time is defined as the incidental work in this paper.

3.2. Result of production simulation

3.2.1. Result by sub-assembly production order

The production simulations are performed by randomly changing the production order of the 10 sub-assemblies, and the total production time (throughput time) and the incidental work time (= the time for tool preparation) are calculated. Detailed simulation settings for tool preparation are as follows:

- If the operator does not have the necessary tools, get the necessary tools from the toolbox if they are in the toolbox.
- If the tool currently used by the operator is not needed for the next work, leave it at the current location of the sub-assembly.
- If the operator does not have the necessary tools, or if the necessary tool is not in the toolbox, obtain it from other sub-assembly.
- Operators can only carry up to 2 tools in one move. The time required for one round trip is 3 minutes.
- When moving the tool currently used by the operator to the next work, only 2 tools can be held in one move, and it takes 3 minutes for each round trip.
- Even if the operator does not have tools when moving, it takes 3 minutes.
- As a cleaning time, the worker holding the tool returns it to the tool box. The tools left in a sub-assembly are collected and returned to the tool box. The rules are the same as before.

Table I: Working parameter of the simulation by sub-assembly production order

Pre-fitting operation	SUB1	SUB2	SUB3	SUB4	SUB5	SUB6	SUB7	SUB8	SUB9	SUB10
NO. of parts to be fitted	4	4	4	4	5	5	5	6	3	3
NO. of parts to be distributed by hands	4	4	4	4	3	3	3	2	3	3
Time required for distribution by hands	4	4	4	4	3	3	3	2	6	6
Time required for pre-fitting by hands	12	12	12	12	9	9	9	8	12	12
NO. of parts to be distributed by crane	0	0	0	0	2	2	2	4	0	0
Time required for distribution by crane	0	0	0	0	10	10	10	20	0	0
Time required for pre-fitting by crane	0	0	0	0	20	20	20	48	0	0
Tools required for pre-fitting	A,B,C	A,B,C	A,B,C	A,B,C	B,C,D,E,F,G	B,C,D,E,F,G	B,C,D,E,F,G	A,B,C,D,E	D,E,F	D,E,F
Time required to clean up tools	2	2	2	2	2	2	2	4	2	2
Welding operation										
Tools required for welding	X,Y,Z	X,Y,Z	X,Y,Z	X,Y,Z	X,Y,Z	X,Y,Z	X,Y,Z	X,Y,Z	X,Y,Z	X,Y,Z
Time required for welding by hands	16	16	16	16	15	15	15	8	15	15
Time required for welding by crane	0	0	0	0	15	15	15	40	0	0

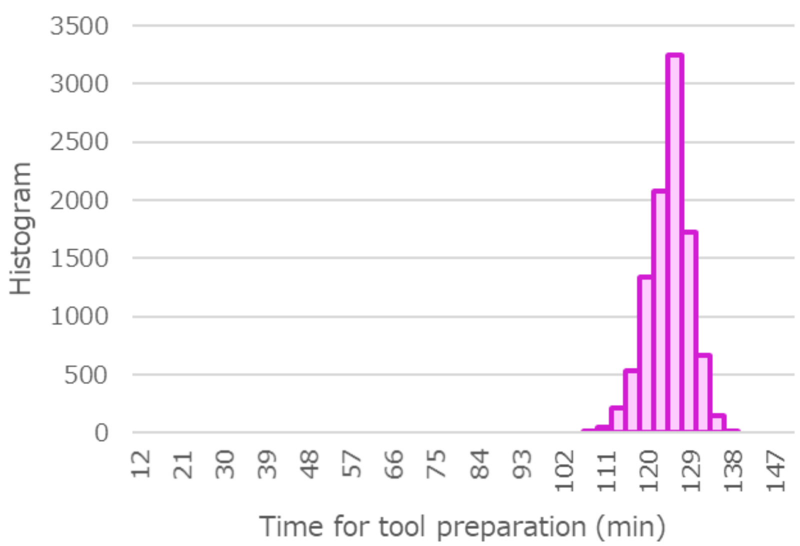


Fig.4: Histogram of the work time for tool preparation

Fig.4 shows the histogram of the work time for tool preparation as the results of the simulation for 10,000 cases. Depending on the production order of sub-assemblies, there is a variation in the time for tool preparation (incidental work). Even if the production order of sub-assembly is determined in advance, it can be confirmed that a distribution occurs in the time of the incidental work due to the flexibility of workers. In this case, setting the production order of the sub-assembly so that the same tool can be used in the next work produces efficient results. This is because, when the tool used for each work is different, it is necessary to acquire a necessary tool at the next workplace after moving there. The work time for incidental work can be reduced by minimizing the number of movements for tool preparation or by grouping the same work in this case study.

Fig.5 shows the results of simulating a veteran and a non-veteran, respectively. This simulation is set non-veteran to go to the tool box when searching for tools, as the non-veteran is unable to judge where the necessary tools were located at each time.

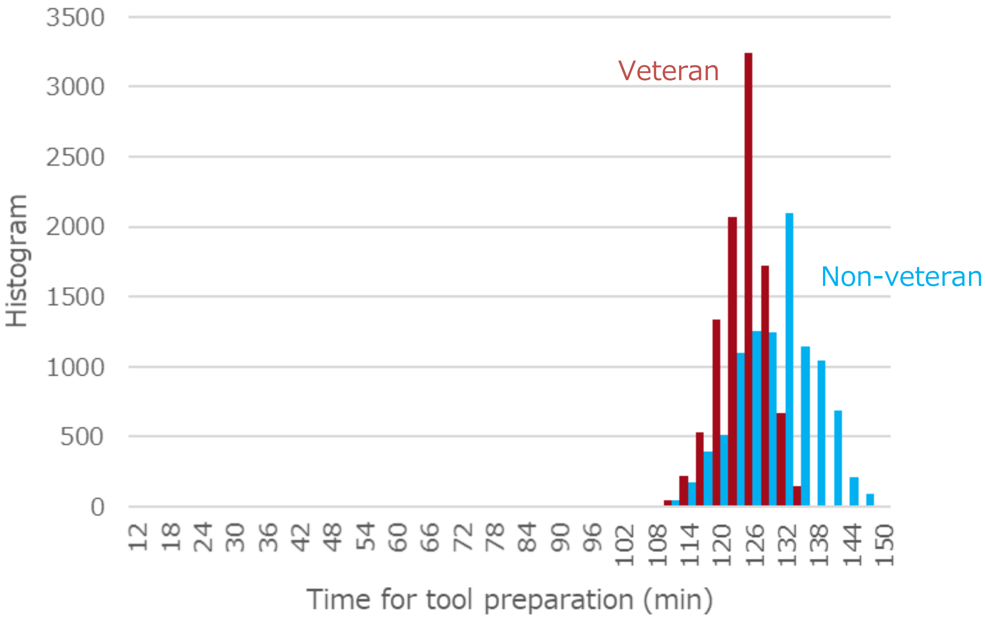


Fig.5: Histogram of the work time for tool preparation for veteran and non-veteran

In the case of non-veteran, the tail of the histogram increases to the direction of increasing the work time. However, depending on the production order, it can be confirmed that the same performance as a veteran can be exhibited by a non-veteran. For example, non-veterans unnecessarily go to the toolbox to search for tools when they do not have tools. For this reason, the production order may be designed so that the next work can be processed by the present tool.

3.2.2. Result by factory layout

Production simulation is performed by incorporating the concept of the factory layout into the simulation of 3.2.1. As a result, the work time for distributing parts and preparing tools becomes a fluctuation value according to the physical distance. It also takes into account the crane interference by including the concept of space. Table II shows the working parameters in this simulation, and Fig.6 shows the factory layout used in this simulation.

In this simulation, not only the production order of the sub-assembly but also the fitting order of the part is changed at random. By considering the order in which the parts are fitted, it is possible to more sensitively confirm fluctuations in work time due to the necessity of a crane and waiting for the crane.

Table II: Working parameter of the simulation by factory layout

Sub assembly	SUB1	SUB2	SUB3	SUB4	SUB5	SUB6	SUB7	SUB8	SUB9	SUB10
NO. of parts to be fitted	4	4	4	4	5	5	5	6	3	3
Type of parts	1a,1b,1c,1d	2a,2b,2c,2d	3a,3b,3c,3d	4a,4b,4c,4d	5a,5b,5c,5d,5e	6a,6b,6c,6d,6e	7a,7b,7c,7d,7e	8a,8b,8c,8d,8e,8f	9a,9b,9c	10a,10b,10c
Pre-fitting operation										
NO. of parts to be distributed by hands	4	4	4	4	3	3	3	2	3	3
Type of parts	1a,1b,1c,1d	2a,2b,2c,2d	3a,3b,3c,3d	4a,4b,4c,4d	5a,5b,5c	6a,6b,6c	7a,7b,7c	8a,8b	9a,9b,9c	10a,10b,10c
Time required for distribution by hands	Variable	Variable	Variable	Variable	Variable	Variable	Variable	Variable	Variable	Variable
Time required for pre-fitting by hands	12	12	12	12	9	9	9	8	12	12
NO. of parts to be distributed by crane	-	-	-	-	2	2	2	4	-	-
Type of parts	-	-	-	-	5d,5e	6d,6e	7d,7e	8c,8d,8e,8f	-	-
Time required for distribution by crane	-	-	-	-	Variable	Variable	Variable	Variable	-	-
Time required for pre-fitting by crane	-	-	-	-	20	20	20	48	-	-
Tools required for pre-fitting	A,B,C	A,B,C	A,B,C	A,B,C	B,C,D,E,F,G	B,C,D,E,F,G	B,C,D,E,F,G	A,B,C,D,E	D,E,F	D,E,F
Time required to prepare tools	Variable	Variable	Variable	Variable	Variable	Variable	Variable	Variable	Variable	Variable
Time required to clean up tools	Variable	Variable	Variable	Variable	Variable	Variable	Variable	Variable	Variable	Variable
Welding operation										
Tools required for welding	X,Y,Z	X,Y,Z	X,Y,Z	X,Y,Z	X,Y,Z	X,Y,Z	X,Y,Z	X,Y,Z	X,Y,Z	X,Y,Z
Time required to prepare tools	Variable	Variable	Variable	Variable	Variable	Variable	Variable	Variable	Variable	Variable
Time required for welding by hands	16	16	16	16	15	15	15	8	15	15
Time required for welding by crane	0	0	0	0	15	15	15	40	0	0
Time required to clean up tools	Variable	Variable	Variable	Variable	Variable	Variable	Variable	Variable	Variable	Variable

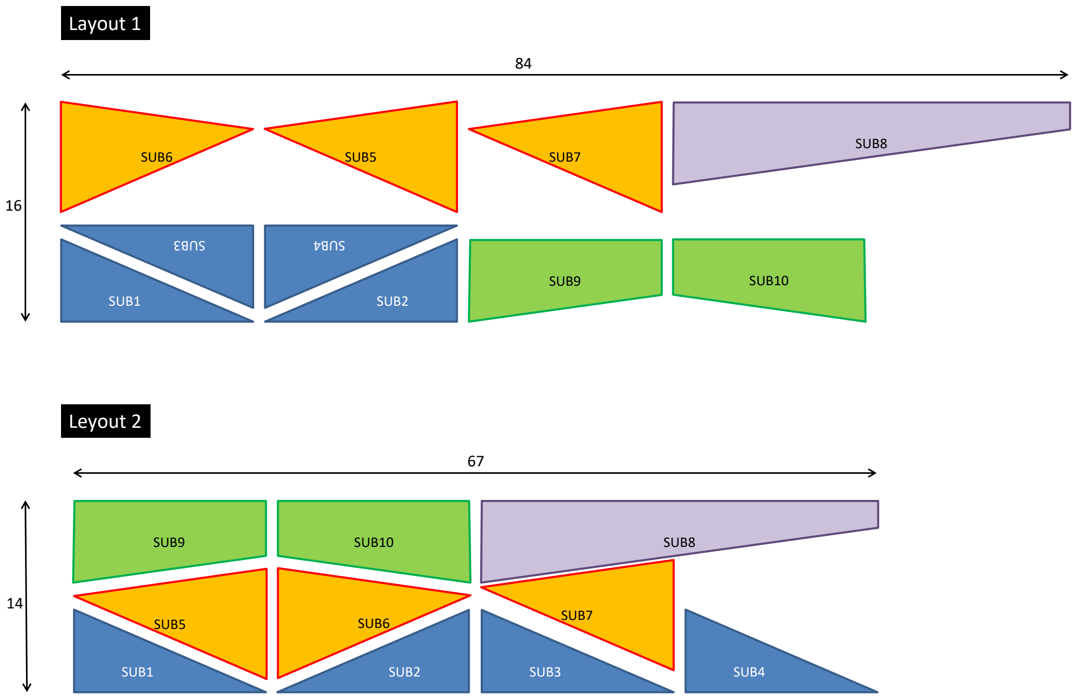


Fig.6: Factory layout used in simulation

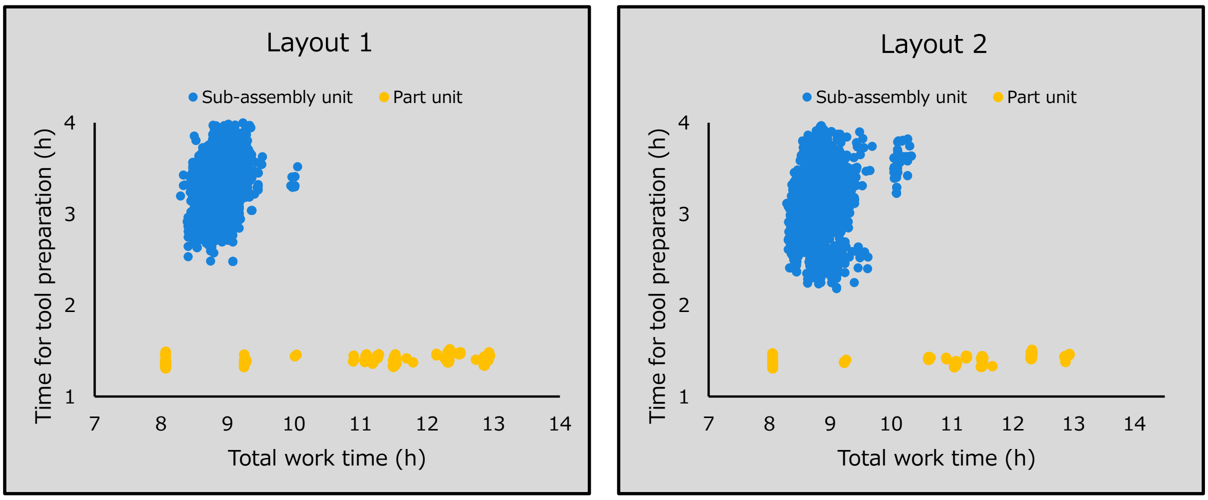


Fig.7: Distribution of total work time and tool preparation time for each layout

Fig.7 is a two-dimensional plot of the total work time and tool preparation time for each simulation. Generally, the total work time becomes shorter in the part unit than in the sub-assembly unit. This is because a large number of man-hours as a sub-assembly can be divided into parts so that they can be shared by a plurality of workers (On the other hand, tool preparation time increases since the production order changes for each part). However, when considering the crane in this simulation, it is confirmed that the total work time becomes shorter in the sub-assembly unit than in the part unit. The crane interference often occurs in case of on part unit, and the work for parts which use the crane is postponed. It can be considered that the handover to the welding work is delayed and the work completion time is extended in this simulation.

4. Conclusion

In regard to “flexibility as an individual”, authors plan to develop an AI system that substitutes worker’s judgement with artificial intelligence such as an expert system by formalizing judgement of veteran who respond to individual variation. Currently, we are studying a work monitoring system for collecting working data of workers or a three-dimensional measurement system such as a laser scanner for measuring individual variation of intermediate products. We are looking at R&D for “detailed work instructions reflecting individual variation” or “setup robots reflecting individual variation”.

The authors have considered the “Digital Shipyard” and paying particular attention to the flexibility of work in shipbuilding. Flexibility is classified into those that can be planned in nature by analysing worker's flexibility and those that need to be flexible on the case according to individual variation.

The final goal of the study is to eliminate the boundary between production planning and production execution, and create a production plan that goes as planned. Shipbuilding is not a mass production but a one-by-one production, and a ship is a complicated product. It is difficult to plan the details of the work in advance, but breaks this with new technology.

A Method to Assess the Reliability of The Machinery on Autonomous Ships

Mohammad Mahdi Abaei, Delft University of Technology, Delft/Netherlands, m.m.abaei@tudelft.nl
Robert Hekkenberg, Delft University of Technology, Delft/Netherlands, R.G.Hekkenberg@tudelft.nl

Abstract

Maritime Autonomous Surface Ships (MASS) are expected to have a considerable impact on maritime transportation. Most research focused on understanding autonomous navigation while the importance of the machinery plant has received very little attention. In this paper a systematic approach is presented to predict failures of sensitive components based on the frequency of current inspection, maintenance and repair activities by the crew. A Multinomial Process Tree is used to model categorical failures in the system by addressing weak components for an unattended Machinery Plant. Hierarchical Bayesian Inference is adopted to facilitate the prediction of disruptive events and estimate the hazard rate function of the system. The outcomes will enable strategies to prevent fatal technical failures in MASS. A real case study is adopted to demonstrate the application of the presented framework.

1. Introduction

Future Maritime Autonomous Surface Ships (MASS) are expected to have considerable impact in maritime trading. The European Waterborne Technology Platform (Waterborne TP) states that smart marine transportation is essential to safe, sustainable and efficient shipping. One of the expected benefits of MASS will be reduced maritime accidents and mitigation of human errors. DNV GL has initiated projects revolving around ship automation and autonomous control. The ReVolt project is one example that has designed as a proof of concept for autonomous ships, https://www.dnvgl.com/technology-innovation/revolt/index.html. Other projects conducted with DNV GL involvement for the Advanced Autonomous Waterborne Applications Initiative (AAWA), led by Rolls-Royce, are investigating a wide array of aspects relevant to commercial unmanned shipping from technical development to safety, legal and economic aspects as well as societal acceptance, *Rolls-Royce (2017)*.

The project Maritime Unmanned Navigation through Intelligence in Networks (MUNIN) was a collaborative research project, co-funded by the European Commissions under its 7th Framework Program, with aiming to develop and verify a concept for an autonomous ship, http://www.unmanned-ship.org/munin/.

Most research related to MASS addressed autonomous navigation and communication. However, the importance of maintenance for the unattended machinery within the engine room has received hardly any attention. Ships have multiple crew members that are employed to keep the engine room operating, and their absence may change the design of power plant significantly, *Colon (2018)*. To determine the required changes in the machinery of autonomous ships, is necessary to understand the reliability of the current engine rooms. The AAWA states that the systems of an autonomous ship “should be resilient to failure and extended maintenance intervals”, *Rolls-Royce (2017)*. The findings of MUNIN include that “Current preventative maintenance procedures need to be updated to ensure operability of components during voyage”, *Rødseth and Burmeister (2012)*.

This paper proposes a dynamic predicting approach for failure modelling of the power plant when it is unmanned, i.e. it is remotely supervised, but there is nobody physically present in the engine room. A systematic framework is presented to model how random failures propagate through the machinery and results in system disruption. The main concern is to understand the failure events that will put the system performance at risk without human intervention. To this end, a “closed-loop” reliability model is constructed to react as a framework for evaluating degradation of system under the influence of disruptive events and update itself automatically from new observation. Firstly, a Multinomial Process Tree

(MPT) is constructed to model behaviour of the entire system according to the required maintenance strategies. Secondly, a Hierarchical Bayesian Model (HBM) is employed to model the uncertainty of the variables and predict the required maintenance actions for the entire system. The event data are obtained from the expert opinions about the frequency of Repair, Maintenance and Check activities for the critical components. With the advent of obtained event data, useful information for the current features of the system is provided that can assist for modelling operational uncertainty involved in Unattended Machinery Plant (UMP). Also, it provides important input for redesign or improvement of system for identifying weak components. The developed model is intended to serve as a frame of reference to determine a performance metrics that incorporate basis of reliability with measures of system functionality through random disturbance and anticipate failures during the operation. The outcome of this research will enable strategies to increase the Level of Autonomy (LoA) and improving the safety of MASS by preventing Fatal Technical Failure (FTF) in autonomous marine operation. A real case study of a Main Engine (ME) is adopted to demonstrate the application of the presented research framework. The results highlight the importance of predicting appropriate hazard rate functions that is essential for redesign of UMP according to the maintenance activities.

2. Methodology

The focus of this study is on the machinery plant used on merchant ships. Generally, most of the ships run on four-stroke diesel engines, connected to the propeller via a gearbox. They are usually equipped with a single main engine and outfitted with at least two diesel generators and auxiliary engines. Both main and auxiliary engines are supported with a fuel oil system which cleans the fluid and prepares it for use. The engine is started with air from compressor and cooled with a triple layer seawater-based cooling system. Mostly, they are equipped with a single rudder and propeller and one or two bow thrusters. Auxiliary systems need to keep the engines running, which will be in connection with the fuel oil system, lubrication oil system, cooling water system, starting air system and exhaust gas system. An overview of machinery power plant is illustrated in Fig.1. The engine itself consists of many parts, but not all the components will be weak points if they left unattended, *Colon (2018)*.

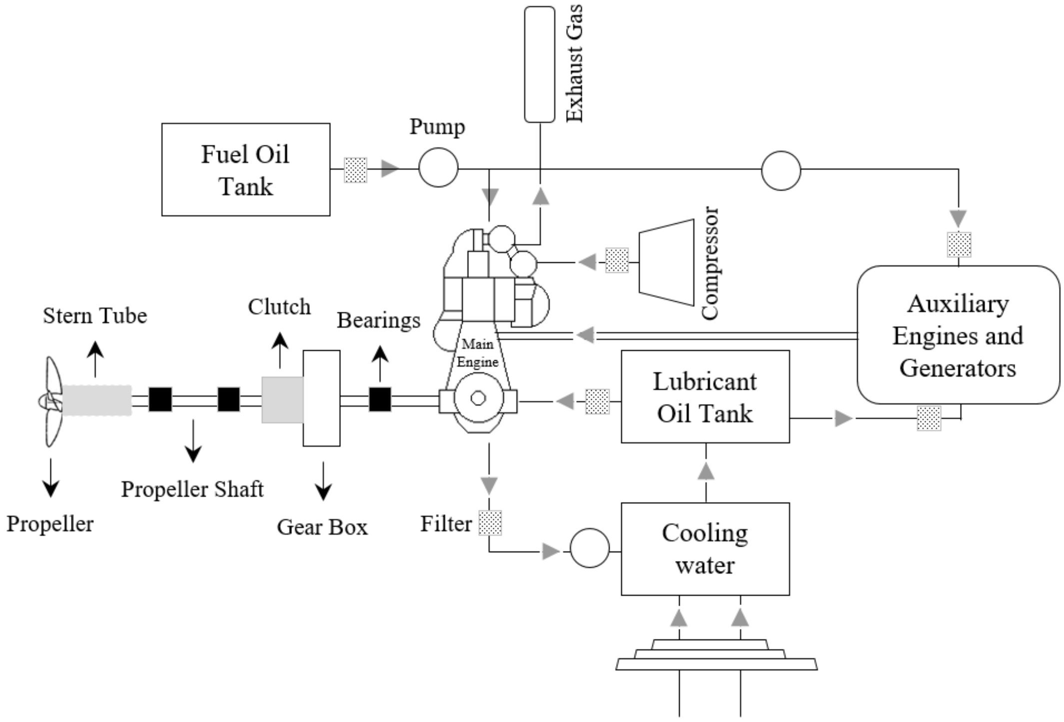


Fig.1: General overview of machinery plant on-board

To increase the profitability of operation and to extend the lifecycle of the assets without compromising safety and environmental performance, it is essential to establish a proactive maintenance process that is suitable for unattended systems. Integrating real-time maintenance planning with associated risk levels in the system will be a support tool to improve safety of operation. This will result in better asset management. By this concept, the condition of an individual component or a group of critical components at the system level is defined and will allow shipping companies to prioritize maintenance actions. As suggested by *Knudsen et al. (2014)*, “using a hierarchical framework for aggregating the health condition or reliability of components into a health status of a sub-system will give an instantaneous indication to the crew and managers about potential failures and the effects will have on the ship reliability and performance”. The performance of the system that is influenced by any disruptions can be evaluated over the operation time according to the frequency of required activities such as “Repair”, “Maintain” and “Check” of the assets, under the assumptions that all of these activities indicate that either performance was actually reduced (“repair”) , that there was an expectation that a drop in performance was likely to occur if the activity was not performed (“maintain”) or that the activity is a conservative match with the expected frequency of a failure or performance decrease of a part of the system (“check”). This will result in a prediction of the health state of the system and allows design of proper maintenance strategies to mitigate FTF.

Now, the goal is to understand how the disruption may spread through the system and lead to change the operation from its normal condition to degraded functioning of the critical components while there are no human interventions. For this purpose, the performance of the UMP will be modelled by introducing the Categorical Failure Functions (CFFs) dependent on operation time. However, defining the CFF is not an easy process, since the problem involves a great amount of uncertainties and unknown probabilities for the variables. Moreover, CFF should be defined in the context of multivariate activities that represent required “Repair”, “Maintenance” and “Check” actions for a group of components in a system. This will result in a nonlinear CFF with multiple unknown variables. As suggested by *Heck et al. (2013),* MPT is a good choice to describe CFF for the multiple failure in terms of a categorical probability distributions. MTP models are simple, substantively motivated statistical models that can be applied to categorical data and representing performance of an operation globally. MPT models address categorical data based on the assumption that the sample frequencies observed for the Event-data set follow a multinomial distribution. In the case of categorical data, a multinomial distribution is the most general and theoretical neutral statistical distribution, which is a neutral generalization of the binomial distribution to more than two categories. In the multinomial distribution, observations are independent and identically distributed (iid) over the categories and each category has a parameter representing the probability that a random observation falls into it, *Batchelder and Riefer (1999)*. Therefore, MPT will express the probability parameters as function of the system behaviour for different circumstances and re-parametrize the multinomial distribution for an objective situation. Each branch of the tree represents a different hypothesized sequence of the processing stages of the operation, resulting in a specific response category regarding obtained knowledge over the system.

On the other side, HBM is a statistical approach that represents uncertainty with probability distributions and provides a ready framework for propagation of uncertainty through risk and reliability models, *Kelly and Smith (2009)*. Due to the complexity of systems in autonomous ships, the prediction model needs to react as a function of a desired goal state regarding observed conditions of a system. Recent conducted research on evaluating reliability of marine operations and predicting availability of systems highlight key attributes of Bayesian method, namely the ability to incorporate qualitative information (i.e. evidence) into the parameters, *Bahootoroody et al. (2019a,b)*, *Khalaj et al. (2020)*, *Leoni et al. (2019)*. In this study, HBM will be employed to estimate uncertainty of random variables associated with MPT model.

2.1 Approach Description

In this study, an MPT model is constructed based on engineer tasks to preventing failure in critical, failure sensitive components in the system. The necessary activities are divided into four main actions as Repair (C1), Maintenance (C2) or Check (C3) and Do nothing/Continue operation (C4). The link probabilities for performance that does or does not deviate from normal (safe) condition are labelled as:

$$
\begin{aligned}
&P(C_1 : \mathbf{Repair}) = \theta_1 \times \theta_2 \\
&P(C_2 : \mathbf{Maintenace}) = (1-\theta_1) \times \theta_3 \times \theta_4 \\
&\qquad\qquad\qquad\qquad = \theta_3\theta_4 - \theta_1\theta_3\theta_4 \\
&P(C_3 : \mathbf{Check}) = (1-\theta_1) \times (1-\theta_3) \times \theta_5 \\
&\qquad\qquad\qquad = \theta_5 - \theta_3\theta_5 - \theta_1\theta_3 + \theta_1\theta_3\theta_5 \\
&P(C_4 : \mathbf{Safe}) = \theta_1 \times (1-\theta_2) + (1-\theta_1) \times \theta_3 \times (1-\theta_4) + (1-\theta_1) \times (1-\theta_3) \times (1-\theta_5) \\
&\qquad\qquad\quad = 1 - \theta_5 - \theta_1\theta_2 + \theta_1\theta_5 + \theta_3\theta_5 - \theta_3\theta_4 - \theta_1\theta_3\theta_5 + \theta_1\theta_3\theta_4
\end{aligned}
\tag{1}
$$

The unknown parameters θ represent the probability of occurrence of each processing branch which will ultimately lead to the probability of event category C. The categories C are a nonlinear function that stand for behavior of the system. As illustrated, in the MPT model in Fig.2, the right-hand side of the root categories represent critical failures and the rest describes non-critical or small failures. According to the description, the category C_1 needs major repair, while C_2 and C_3 need maintenance and checking of the involved components respectively. The developed MPT model for the system is plotted in Fig.2 and associated categorical functions for the feeding the HBM are presented in Fig.3.

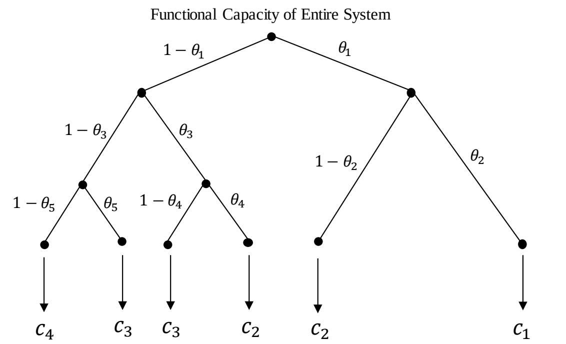


Fig.2: Constructed MPT model for evaluating performance of system influenced by random disruption

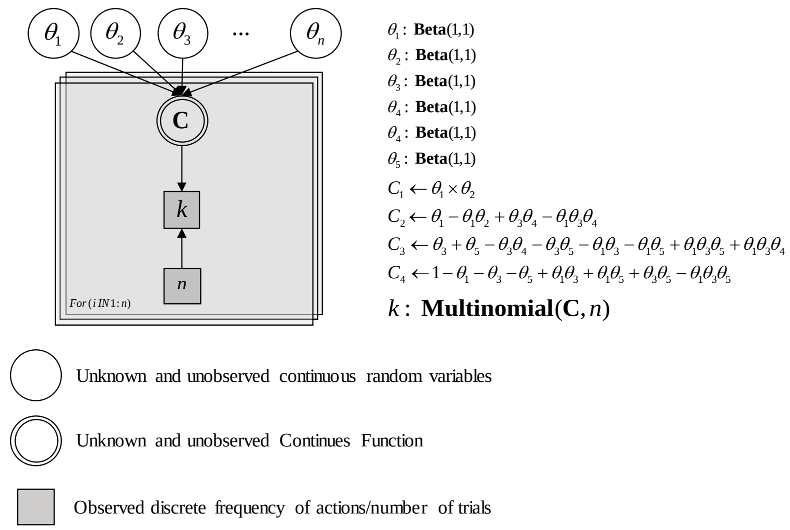


Fig.3: Constructed Bayesian Inference Network for predicting uncertain variables

The Bayesian Inference network in Fig.3 is constructed for uncertainty modelling of the MPT. In the figure, the circles represent the uncertain parameters and the square is related to the observations. The node C will be defined as the nonlinear categorical failure functions that form behavior of system in connection with branches variables. Subsequently, parameter k is represented as the observation variable that is connected to the categories C for monitoring behavior of critical components in the system. The node C will be defined as the nonlinear categorical failure functions that form behavior of system in connection with branch variables θ. In statistics, parameters within the models are regarded as random variables and having probability distributions. The first step in Bayesian Inference is to determine prior distributions of unknown parameters of process branches θ. Due to shortage of engineering data and physical information regarding details of the process non-informative prior of Beta function, **Beta** (1,1) selected. The model's likelihood is obtained by inserting CFFs (C_1, C_2, C_3, C_4) into the density function of multinomial distributions, given by k~**Multinomial**(C,n) depended on the number of n trials. Finally, the posterior distribution will be derived using Bayesian Theorem.

To perform the simulation, the open source MCMC WinBugs software package, *Spiegelhalter et al. (2003)*, is used to predict marginal posterior distributions. The presented method is able to evaluate degradation phases of the entire system according to the required maintenance activities, and identifying target hazard rate function which is suitable for examining system behavior and extending maintenance interval of UMP. The next section will describe a real case study to demonstrate the application of the presented framework.

3. Application: Case Study of a Main Engine

To demonstrate the application of the framework, critical components of the Main Engine (ME) is considered as the case study. The system breakdown for the main engine used a MAN B&W K98MC-C7-TII as reference. This is based on a four-stroke diesel engine with an output between 36 and 84 MW; suitable for most large size vessels. The structure of event data and prior observation obtained from survey analysis which is adopted from the research conducted by *Colon (2018)*. The system breakdown is filtered according to basis of retaining the most critical components that affect operations directly. This criterion is suggested by *Colon (2018)* to exclude any parts that can only be repaired or replaced during a major overhaul. Accordingly, by creating a list of critical machinery and equipment breakdowns, the analysis starts with filling in frequency indexes of crew activities regarding each item. This is done by processing data acquired from several expert engineers that is selected from *Colon (2018)*. The summary of the observed frequencies from expert activities are presented in Table I. As described in Section 2, three options were considered for obtaining frequency index of expert engineer feedbacks regarding monitoring the machineries; "Check", "Maintenance" and "Repair".

Table I: Obtained observations of FI for different actions from expert crews

	Repair (Unplanned Maintenance) C1	**Planned Maintenance C2**	**Check C3**
Cylinder cover	1	2	4
Gear box	1	2	3
Stern tube Seal Cover	3	3	5
Piston cylinder	1	2	1
Maneuvering system	1	2	4
Clutch	2	2	2
Attached pump	1	3	3

Interpreting the definition of FI

FI ranges	**Frequency (Per ship per year)**	**Definition**
FI≥6	>1000	Multiple times per day
4-5	100	Once a day to once a week
3-4	10	Once a week to once a month
2-3	1	Once every 3 months to once a year
1-2	0.1	Once every 5 to 10 years
1	> 0.01	Once in a ships' lifetime
1	0.001	Once in a fleets' lifetime

Recently, *BV (2019)* and *Colon (2018)* both created a Frequency Index (FI) to be used in risk and reliability assessment of autonomous ships. The index is defined in frequency of occurrence of an event per year per ship. The highest values of FI=5, 4 and 3 represent a frequency of 'multiple times per day', 'once a day to once a week' and 'once a week to once a month' respectively. While the lowest values of FI=2 and 1 represent for 'once every 3 months to once a year' and 'once every 5 to 10 years respectively'. The higher value of FI≥ 6 indicates for the frequency of required actions (i.e planned or unplanned maintenance and Checking) that the typical component may need for inspection or repair more than once a time in a day. Therefore, the chance of observing higher value of FI≥ 6 in the graph is only and only has the emphasize for the increasing number of required actions for inspecting health condition of the system, though the probability of its occurrence is almost near to zero. The. The details of the FI ranges and its description are described in *BV (2019)* and *Colon (2018)*. The BV used FI between 0 to 7 scale for interpreting the index while colon considered 0 to 5. In both cases they interoperated the FI as described in Table I.

To prevent the UMP from unexpected disruption, it is essential to design a system in the condition that the urgent activities for "Maintenance" and "Repair" become minimized. The joint probability distribution for the frequency index of these two activities is plotted in Fig.4. The significance of the graph is the ability to interoperate the uncertainty correlation between the frequency index of performing repair and maintenance. For instance, considering a case that the FI for maintenance of the failure-sensitive components in the Main Engine is predicted as five. Then it is possible to define a marginalized probability distribution for the FI of repairs in the condition that FI of maintenance is five. Therefore, it can be concluded from the figure that probability of occurrence of a failure that requires immediate repair of one of those critical components ranges between two and six, while the expected FI for repairs is estimated to be four according to the graph plotted in Fig.4.

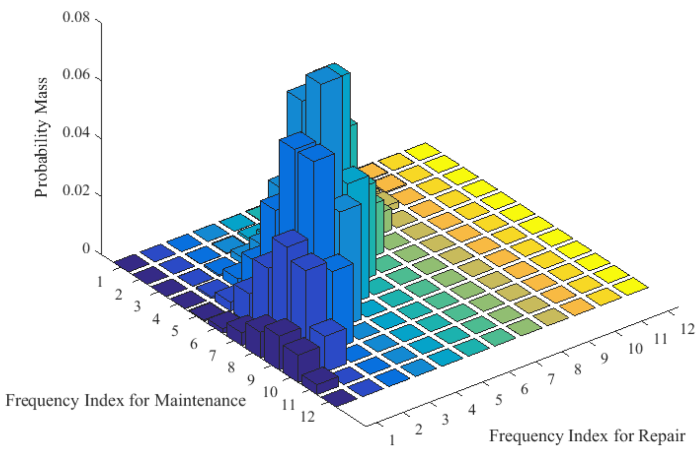


Fig.4: Joint probability of occurrence for urgent maintenance and repair activities

In order to understand the correlated dispersion of inspection activities for the selected critical components, the frequency index for each combination is obtained from the simulation and plotted in Fig.5. The bar charts are the probability of occurrence for each activity while the black dash-lines represent the number of correlated samples of each two activities (i.e. repair by maintenance; repair by check; maintenance by check and vice versa which result in six black graphs). From the correlation graph for maintenance versus repair, it can be concluded that the main engine needs a lot of maintenance to prevent failures for the failure sensitive components In most cases, if the owner does not do the required maintenance on time then the condition of the system will get worse and the critical components may have breakdowns which need immediate repair. The ideal case for an UMP is that the joint probability distribution for Repair and Maintenance shifts to the origin of the graph, i.e. FI=0; this means that the probability of required repairs will decrease and the overall condition for system will be robust to disruptions. These are a part of the important outcomes of MPT model, due to its capability for estimating degradation of the system and optimizing design of UMP under the influence of random disruption. For this purpose, the probability of the marginalized distribution of each action are predicted from the simulation. The Com mutative Distribution Function (CDF) for all three actions "Repair", "Maintenance" and "Check" presented in Fig.6. These functions will be employed to estimate the Hazard Rate Function (HRF) for multiple inspection activities over effective operation days. The HRF is used to describe the failure rate of engineering systems which is essential outcome for reliability assessment and extending maintenance interval of UMP. DNV states that "The extent of automation shall be sufficient to permit unattended engine room operation for the maximum continuous operation time."[18]. It means that UMP must be able to be resilient enough to run continuously and be able to self-perceive the expected times of disruption in process.

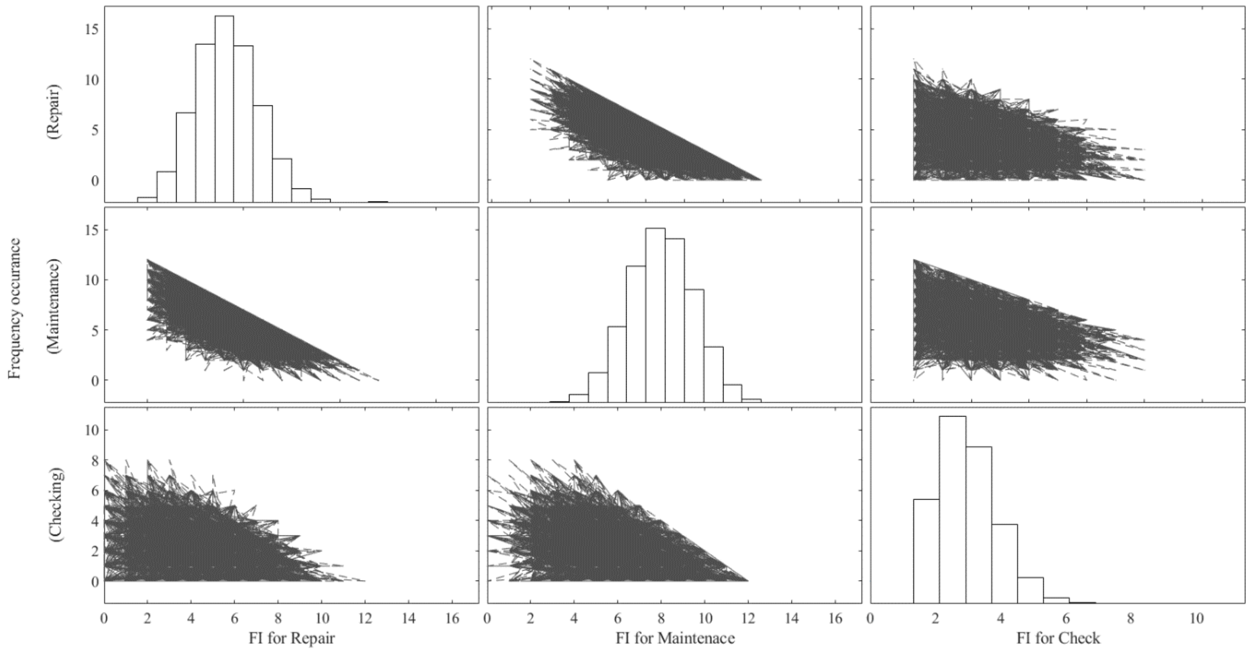


Fig.5: Dispersion of frequency for the predicted inspection activities according to the MPT model

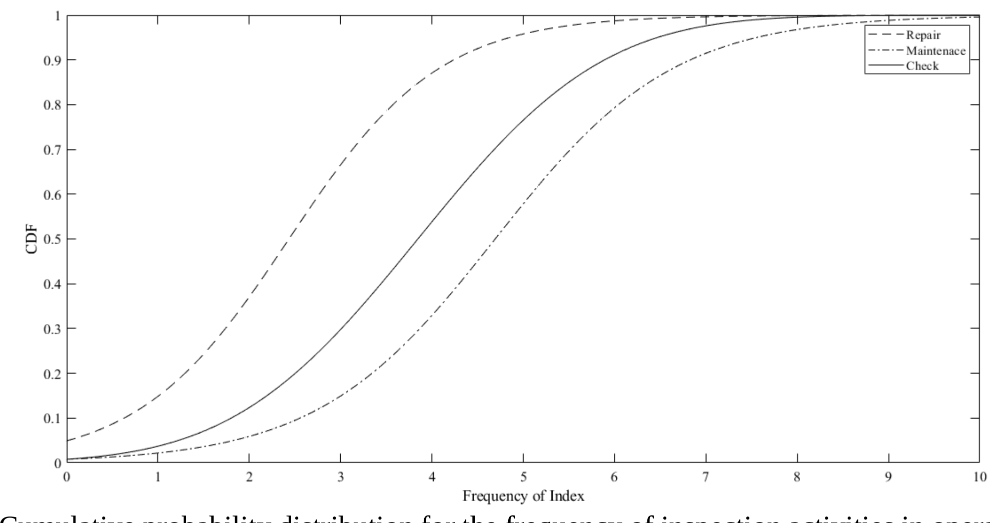


Fig.6: Cumulative probability distribution for the frequency of inspection activities in operation lead to prevent system from disruption

A comprehensive study conducted by DNV GL demonstrated that two main types of failures are expected in operation of marine vessels; Age Related Failures (ARFs) and Random Failures (RFs). Although the ARFs are directly related to the age of products, it contains almost 11% of the failures of the machineries. ARFs are much easier to be predicted than RFs due to the historical population variability of component failure frequencies. Most of the strategies done in engine room regarding ARFs are preventive maintenance based on operational hours, since it is convenient to apply planned maintenance to prevent system interruptions. However, the main concern is related to understand how the system will break down due to RFs that account for almost of 89% of system failures. RFs consist of stochastic disruption periods in a system, where failure events are low at the start of the products' lifetime, and slowly increase over time. This part of the UMP in autonomous systems suffer from limited availability of reliability records, and there is currently little no public information on the impact of human intervention on engine room reliability. The developed method in this study will be a supportive tool for understanding how the performance of UMP will be change if the human interface is removed. This is resulted in identification of the degradation rate of system related to crew activities for the critical components listed in Table I. The Hazard Rate Function (HRF), of the system is estimated to represent the degradation of the entire system regarding each maintenance activity. The total effective operation days of 1000 are considered in the simulation to predict the HRF. Moreover, the plots break to five individual time slices [0,200], [200,400], [400.600], [600,800] and [800,1000] to highlight the variation of system degradation respect to each activity, and illustrated in Fig.7. The results demonstrate the randomness of system degradation with increasing operation time. In the first time slice [0,200] the ”Repair” activity is less likely to occur while it overtakes the "Check" activity after days 158. That implies that the probability that a component fails and needs to be repaired exceeds the probability that a small error occurs that only requires a check. It should be noted that the simulation is run under the assumption of "no human action performed" from the beginning of the operation. This assumption is essential to let us evaluate performance of unattended machinery. In this period (i.e. [0,200]) the required action for maintenance of failure-sensitive components in the main engine system dominates "Check" and "Repair". This graph highlights that reliability of UMPs can be estimated on the basis of the rate of "Maintenance" activities to prevent system from risk of breakdown. Although there is still a significant probability of a need for repair, which has significant impact on operations, the consideration of performing timely maintenance will lead to minimize the cost of repair and the associated downtime. This could be done by establishing essential maintenance schedule to prevent components from drastically deviating from their nominal condition. Indeed, the crews should pay attention to the 'check' line in the graph to estimate the essential workload once the ship makes a regular port call, and use 'maintenance', to indicate the probability of having to return to the nearest port, and use 'repair' to estimate if it is acceptable to postpone maintenance and checks any longer.

The starting time of major disturbance is predicted at 124 hours which is almost the same for both "maintenance" and "Repair" activities, however, the rate of increase of "Maintenance" is notably higher than that of "Repair". The failure rate is highly depended on time of operation. As shown in the graph ,"Repair" activity overtakes "Maintenance" after 250 operation days. For the last two time slices [600,800] and [800,1000], the trends demonstrate that the rate of all activities will reach the same pattern, though "Maintenance" is slightly higher than "Repair" and "Check" activities. These line graphs can give the simultaneous failure rate of critical systems in various time slices. This will help the industry to have a better view of the degradation status of unattended machinery in an autonomous condition. For instance, in the case study of this research it is demonstrated that the unattended main engine has negligible failure rate up to 20 continuous operation days (failure rate of approximately of 0.005 per day). By continuing the operation, the rate for maintenance of critical components is drastically increased compared to other actions for Repair and Check (e.g. in 100 days of operation the failure rate for maintenance is 0.06 while the rate for repair and check are 0.01 and 0.025 respectively). Also, the result demonstrates that unless maintenance is performed before day 260, the main engine's repair-related failure rate will increase by 0.15 per day. If no action performed for almost one year (i.e. for 365 days of continuous operation of unattended main engine), then the failure rate of the system reaches to 0.3 per day. The details of the decreasing trend of reliability of unattended main engine is summarized in Table II to emphasize the degradation behavior of a system while no human is intended to be involved in the operation.

With the predicted HRF, it is possible to estimate the probability that system will survive beyond any specific time. For this purpose, the survival function of the system is predicted according to each activity and the results are plotted Fig.8. To better understand how the uncertainty will affect the reliability of operation, the graphs are plotted according to confidence bound of 5% for each action in the same plot. The colorful dash lines represent the Upper Bound Confidence (UBC) and Lower Bound Confidence (LBC) of system reliability.

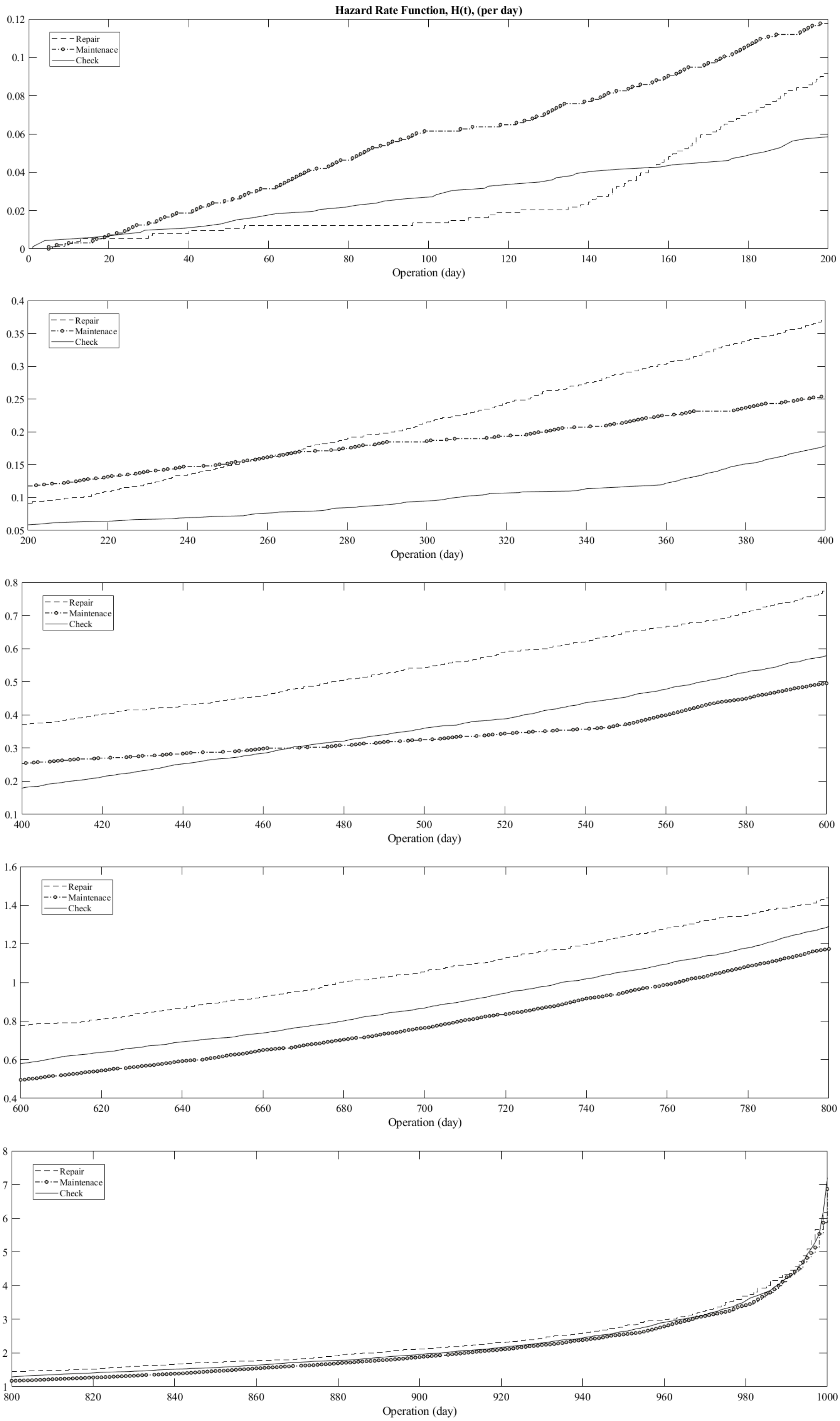


Fig.7: HRF for performing Repair, Maintenance and Check the critical components in Maine Engine

As shown in the plot, the rate of 'Maintenance' and 'Repair' always have a dominant influence on reliability-based design of machinery. According to the result, before 200 days the reliability of system decreases due to need for 'Maintenance', however after this period, the need for 'Repair' is expected to be the most dominant factor in the loss of reliability. Moreover, the UBC for preforming 'Repair' is always lower than other actions, which demonstrate the importance of RFs in stochastic process. To show in details that how the reliability of system will change according to each activity, a summary of probability results for specific time intervals are listed in Table II. The results of this study indicate the advantages of the method for understanding how the system works to maintain its essential functioning of UMP according to the required inspection activities. The approach can be adopted by continuous monitoring of the performance of system and making an autonomous system cognitive in the sense that it is able to identify and minimize influences that would disrupt its essential functioning.

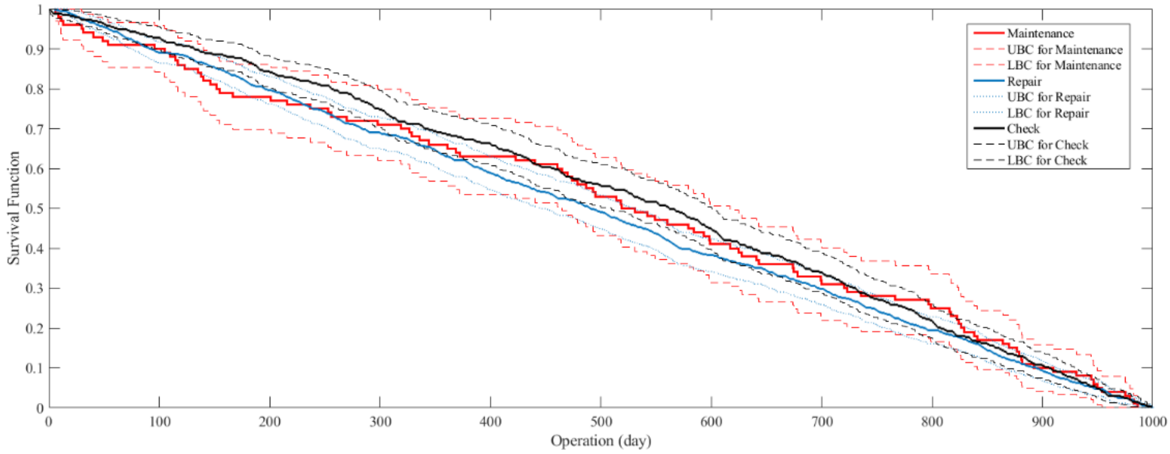


Fig.8: Survival Function of main engine with considering 5% confidence bound; Maintenance (Red Colors), Repair (Blue Colors) and Check (Black Colors)

Table II: Reliability of System respect to each activity depended on operation time (Values are in percent)

Time (in days)	Maintenance	Repair	Check
100	**0.9249**	0.9490	0.9569
200	**0.7700**	0.8400	0.8817
300	0.7687	**0.7140**	0.7814
400	0.5841	**0.5412**	0.6612
500	0.4881	**0.4652**	0.5525
600	0.4324	**0.3717**	0.0.4432
700	0.3120	**0.2929**	0.3363
800	0.2814	**0.1912**	0.2020
900	0.1050	**0.1050**	0.1084
1000*	0	0	0

* Value zero at time 1000 means the complete breakdown of the system

4. Conclusion

In this paper a predictive tree model is developed to estimate reliability of UMP. The model consists of main three eminent actions (Repair, Maintenance and Check) that may be require in operation for prevent the system failure. To this end, an MPT model constructed for failure modelling of system with respect to required inspection activities. The HBM employed to model the uncertainty of random variables and predicting unknown probabilities. The results summarized for a real case study of main engine. It is realized that for this particular main engine, the critical components of the system should be designed based on the hazardous rate function that react according to the safe limit of 2 disruptive events over 1000 effective operation. This will assure to keep machinery to allow an extend period of time with restricted physical interface in machinery. For the present study, 528 hours is estimated as the starting time of critical degradation in the system. Based on the results of the case study for degradation modelling of unattended machinery and evaluating performance of system, it can be concluded that the proposed methodology can not only be significantly helpful for mitigating risk of failure in an operation of autonomous ships but also it is applicable for extending interval time of maintenance by proposing an assistive tool for evaluating system behavior.

References

BAHOOTOROODY, A.; ABAEI, M.M.; BAHOOTOROODY, F.; DE CARLO, F.; ABBASSI, R.; KHALAJ, S. (2019a), *A condition monitoring based signal filtering approach for dynamic time dependent safety assessment of natural gas distribution process*, Process Safety and Environmental Protection 123, pp.335-343

BAHOOTOROODY, A.; ABAEI, M.M.; ARZAGHI, E.; BAHOOTOROODY, F.; DE CARLO, F.; ABBASSI, R. (2019b), *Multi-level optimization of maintenance plan for natural gas system exposed to deterioration process,* J. Hazardous Materials 362, pp.412-423

BATCHELDER, W.H.; RIEFER, D.M. (1999), *Theoretical and empirical review of multinomial process tree modeling*, Psychonomic Bulletin & Review 6/1, pp.57-86

BV (2019), *Guidelines for autonomous shipping*, Guidance Note NI, vol. 64, Bureau Veritas

COLON. J, (2018), *Identifying and eliminating weak points in ship's machinery plants: A step towards continuously unmanned engine rooms*, Master Thesis, TU Delft

HECK, D.W.; ARNOLD, N. R.; ARNOLD, D. (2018), *TreeBUGS: An R package for hierarchical multinomial-processing-tree modeling*, Behavior Research Methods 50/1, pp.264-284

KELLY, D.L.; SMITH, C.L. (2009), *Bayesian inference in probabilistic risk assessment - the current state of the art*, Reliability Engineering & System Safety 94/2, pp.628-643

KHALAJ, S.; BAHOOTOROODY, F.; ABAEI, M.M.; BAHOOTOROODY, A.; DE CARLO, F.; ABBASSI, R. (2020), *A methodology for uncertainty analysis of landslides triggered by an earthquake*, Computers and Geotechnics 117, p.103262

KNUTSEN, K.E.; MANNO, G.; VARTDAL, B.J. (2014), *Beyond Condition Monitoring in the Maritime Industry*, Position Paper, DNV GL, https://www.dnvgl.com/publications/beyond-condition-monitoring-in-the-maritime-industry-12403

LEONI, L.; BAHOOTOROODY, A.; DE CARLO, F.; PALTRINIERI, N. (2019), *Developing a risk-based maintenance model for a Natural Gas Regulating and Metering Station using Bayesian Network*, J. Loss Prevention in the Process Industries 57, pp.17-24

ROLLS-ROYCE (2017), *Autonomous ships: The next step*, Marine Ship Intelligence

RØDSETH. O.J.; BURMEISTER, H.C. (2012), *Developments toward the unmanned ship*, Int. Symp. Information on Ships (ISIS), vol. 201, pp.30-31

SPIEGELHALTER, D.; THOMAS, A.; BEST, N.; LUNN, D. (2003) "*WinBUGS user manual*," ed: version

How to Achieve Data Integrity in Ship Design through Unified Data Models

Constantin-Alexander Gresens, PROSTEP AG, Hamburg/Germany,
constantin-alexander.gresens@prostep.com
Carsten Zerbst, PROSTEP AG, Hamburg/Germany, carsten.zerbst@prostep.com

Abstract

The typical toolkit for ship design contains multiple independent applications; from tools for early, basic, and detailed design, to numerous complementary tools for particular design tasks. Transferring and harmonising information between these applications robustly and conveniently is challenging but offers long-term efficiency and data integrity benefits. This paper investigates a practical implementation of such a harmonisation effort, building upon unified data models to create lean interfaces between NAPA Steel, SSI ShipConstructor, AVEVA Marine and Siemens Teamcenter. This enables the yard to spin their digital thread from initial design to manufacturing preparation, and thus to work more efficiently without data disruptions.

1. Introduction

Sieranski and Zerbst (2019) explored two different approaches shipyards can take when deciding on the best software for their design processes. The best-of-suite approach focusses on high-level data and data model continuity and integrity by purchasing all tools necessary for the individual design phases from the same vendor. The driving force behind the best-of-breed approach, on the other hand, is not ease of integration between the tools used in different design phases, but rather to select the best-suited tool or application for each design step, independent of their vendors. This comes at the cost of having to transport and harmonise the data between the different tools, most of the time without the direct support of the application vendors. That paper concluded that, even though extra effort will have to be taken to integrate the data and data models, the best-of-breed approach offers yards the best flexibility and the ideal software setup for their ship design process.

However, this conclusion only holds true if integration costs between the different tools can be limited, both in terms of initial implementation and maintenance, but also in terms of operative speed, ease of use, and interface complexity. Hence, we examine in this paper a practical implementation of such a best-of-breed approach, giving details about a productive integration between NAPA Steel, AVEVA Marine, SSI ShipConstructor, Aras Innovator and Siemens Teamcenter. The key idea behind this implementation is to have a common point of reference for all interfaces between these tools. In doing so, we do not need to build interfaces from each individual application to the others. Instead, we connect each one to this common point of reference. Some tools will need to export data to and import data from this integration layer, while others only need one direction. This approach allows us to create maximum connectivity at a relatively low cost. Also, the integration of any two systems will not depend on the integration to any other intermediary system but can still be built upon and extended through this integration layer.

From a practical perspective, not all integration directions make sense. Why would a yard want to migrate data from their detailed or production design tools to steel or early design tools? Also, PLM and ERP systems mostly do not care about detailed geometrical information, whereas CAD tools typically care little about pricing and product life cycles. Hence, the minimisation of such an overhead while keeping the flexibility and independence of this approach intact was another main goal for this practical investigation. Keeping these considerations in mind, we will showcase data integration from NAPA Steel to SSI ShipConstructor as well as AVEVA Marine, and between SSI's Digital Hub (based on Aras Innovator) and Siemens Teamcenter. We will also see that, due to the kind of approach taken, we can also connect SSI ShipConstructor and AVEVA Marine at no additional implementation cost.

2. OpenPDM SHIP – A common point of reference

2.1. Basic Overview

This approach to data integration between multiple different production tools is not a novel one. In fact, both conceptually and technologically, we built upon a solution that has been in practical use in the automotive industry for many years, called OpenPDM. A central process engine (our common point of reference) governs and delegates tasks to a series of so-called connectors. These connectors are simply implementations of a general interface between the integration engine and a production tool, be it a PLM application, a CAD tool or an ERP system.

All these connectors offer the same interface endpoints and express the data of their respective connected systems in the form of a unified data model called BIBO. When extracting data from an application, the connectors map the system-specific data model to BIBO using the systems' official APIs. Similarly, when importing data, the connectors take BIBO objects as input and map them to the corresponding data model for their application. The process engine itself does not rely on any system-specific technologies or data models and thus offers an independent integration layer between the production applications.

With the connectors in place, data transfer between two different systems becomes as simple as telling the process engines which connectors to address and which tasks they shall perform in what order. The connectors themselves are not directly communicating with each other, which allows for the decoupling of the high-level migration process from any specific production tool. Hence, no integration essentially depends on any other interface or even the system we wish to integrate with, per se. Additionally, this approach allows for good control over data integrity between different systems at different production phases. Since data structures in any system are first mapped to BIBO before being processed further, we can simply compare the BIBO objects obtained from each tool to detect any potential issues in terms of data harmonisation early and often.

2.2. Application to Ship Design

Of course, we cannot simply take something that was developed for the automotive industry and apply it directly to ship design processes. Shipbuilding faces unique challenges and thus uses different data models and tools. As we will see in Chapter 3, some of the connectors can be used for shipbuilding pretty much without alteration. But due to the unique requirements towards data exchange between different CAD tools used for different ship design phases, we needed to extend OpenPDM, Fig.1, with functionality that allows for the transport of geometrical, parametrical and topological information, both for steel and outfitting integration.

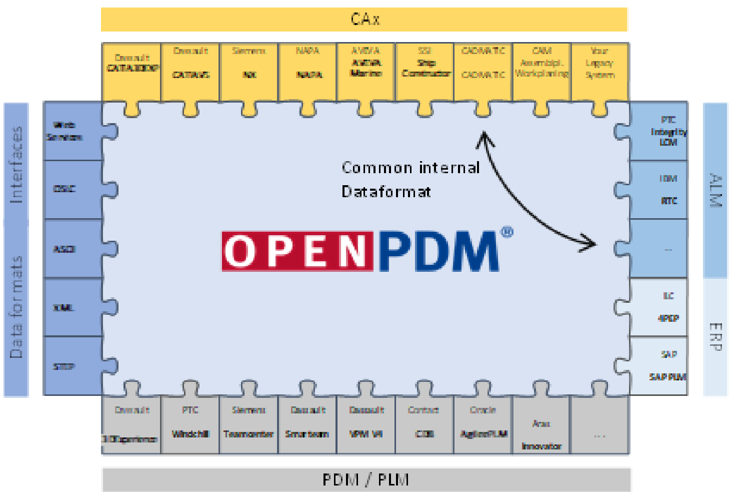


Fig.1: Open PDM

This was achieved by extending OpenPDM with a second general data model for exactly this type of data, Fig.2. Geometrical and topological information is extracted from the design tools, mapped according to this data model and stored in OpenPDM SHIP XML files, similarly to how the connectors extract PLM data from their corresponding tools and map them to BIBO.

Again in analogy to BIBO for PLM tools, this data can then be imported into any other ship design tool with an existing connector to the common process engine.

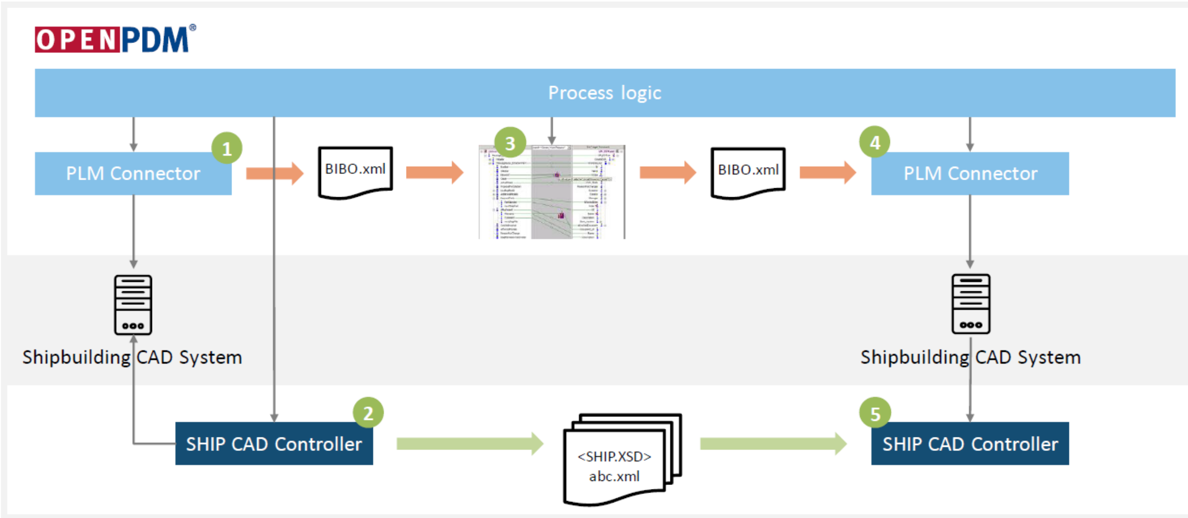


Fig.2: OpenPDM SHIP

3. Example implementation

With these basic ideas and motivations clear, and a first technological overview given, we can now investigate the details of an actual implementation of this approach. For the example showcased here, the starting point for our integration is NAPA Steel.

3.1. NAPA Steel to SSI ShipConstructor and AVEVA Marine

The NAPA Steel connector for OpenPDM SHIP comes in the form of a plugin for NAPA Designer.

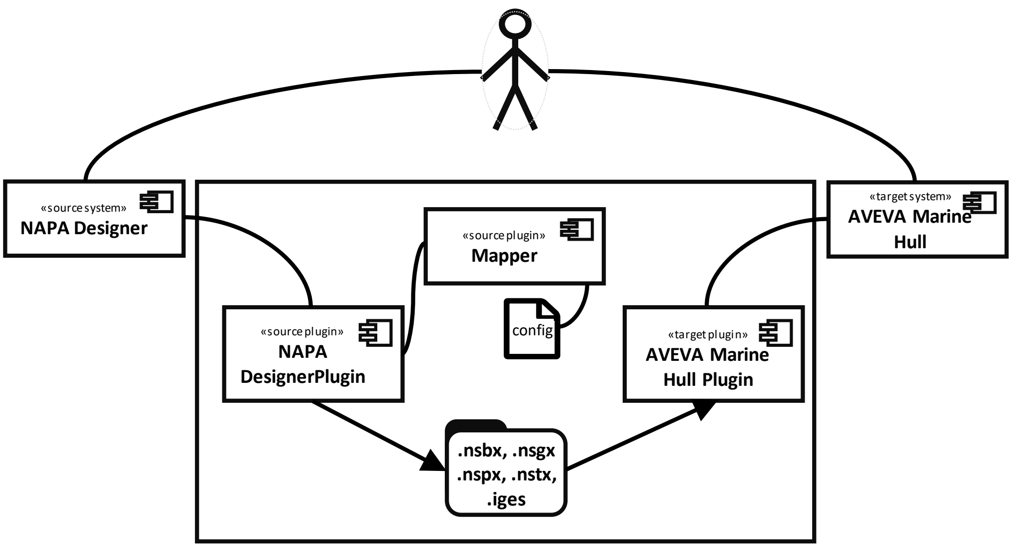


Fig.3: NAPA Designer to SSI ShipConstructor and AVEVA Marine - Architecture

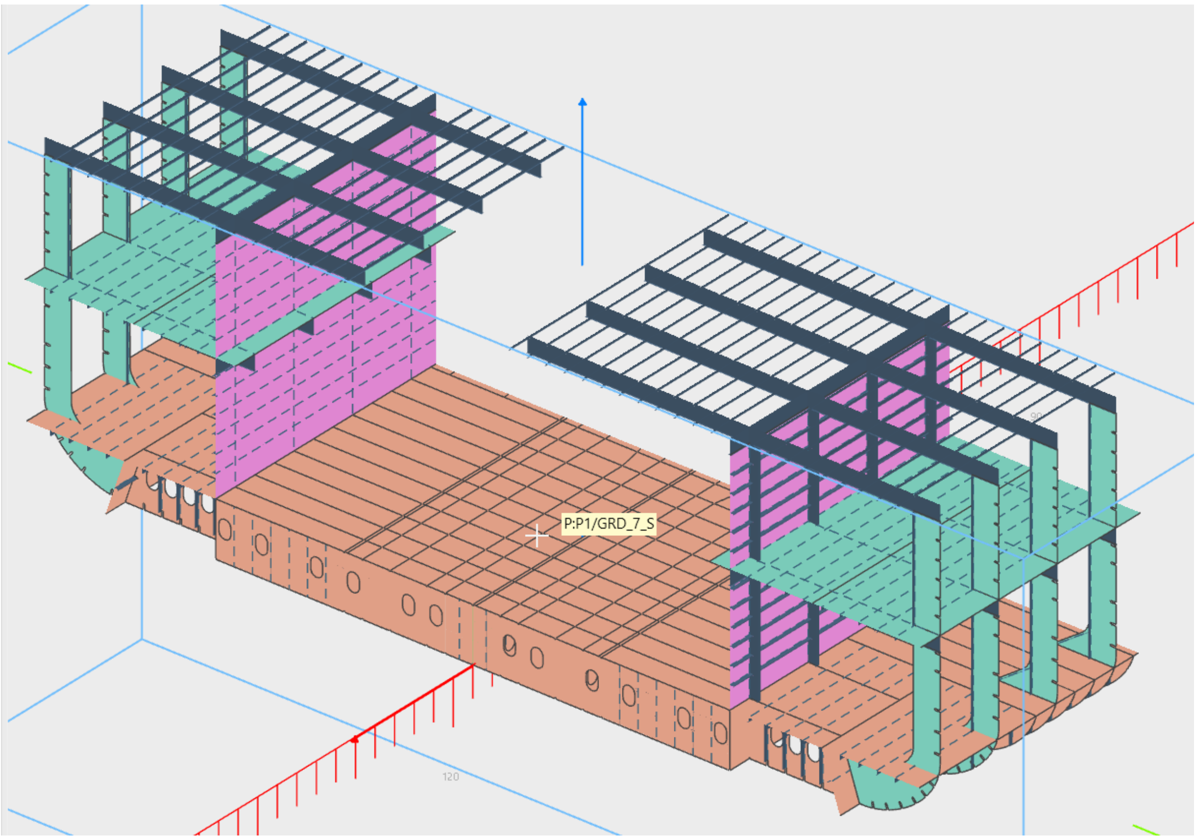


Fig.4: NAPA Designer Model

Using this plugin, the user can export geometrical and topological information on vessel, block or panel level and save it in an OpenPDM SHIP XML file. Since not all data from NAPA Designer will be useful in all systems with which we want to integrate, the user can also set different mapping configurations and export parameters to reduce overhead and data glut.

The plugin itself uses NAPA Designer's official API to extract the data and carry out auxiliary calculations to translate certain NAPA specific representations of data to a more general form so that it can be more easily used by a wider range of design tools. Also, mappers can be configured that either run during the export itself or on the already exported XML file. This allows us to both have the data in a form that is as general as possible if needed or tailored precisely to the use case of any specific shipyard.

Now that the data is exported in a level of detail to our liking, NAPA Designer does no longer play a role in the integration process going forward. As was mentioned in chapter 2, one of the great benefits of such a 'common point of reference' approach is that we consciously decouple the process from the tools to the greatest extent possible. We simply work with the exported XML which adheres to our general data model when it comes to importing the data to the other design tools.

Two such tools for which we already implemented the connectors to OpenPDM SHIP are SSI ShipConstructor (SC in the following) and AVEVA Marine (AM in the following). These connectors read the exported XML files, carry out any auxiliary calculations and mappings needed for the tool to be able to process the data and then import them into the corresponding databases. It is only at this point – the connector level – that system specific requirements are accounted for.

For example, whereas NAPA creates plates on a panel in 'real-time' while new seams are added to it, AM requires each imported seam to divide any existing panel or plate into sub-plates as it is being imported. If we import a seam that does not adhere to this, AM will not create the desired plates, even if after the entire import is completed each seam does divide the panel into plates. Hence, some additional mappings pertaining to the order of the seam import are required for the AM connector. Since the so-called Genesis API for SC does not include this requirement, this mapping does not have to be applied for the SC connector. The system-specific logic resides entirely within the respective connectors, and we can eliminate overhead and unnecessary calculations in this data exchange process, while still relying on a common basis for each integration step.

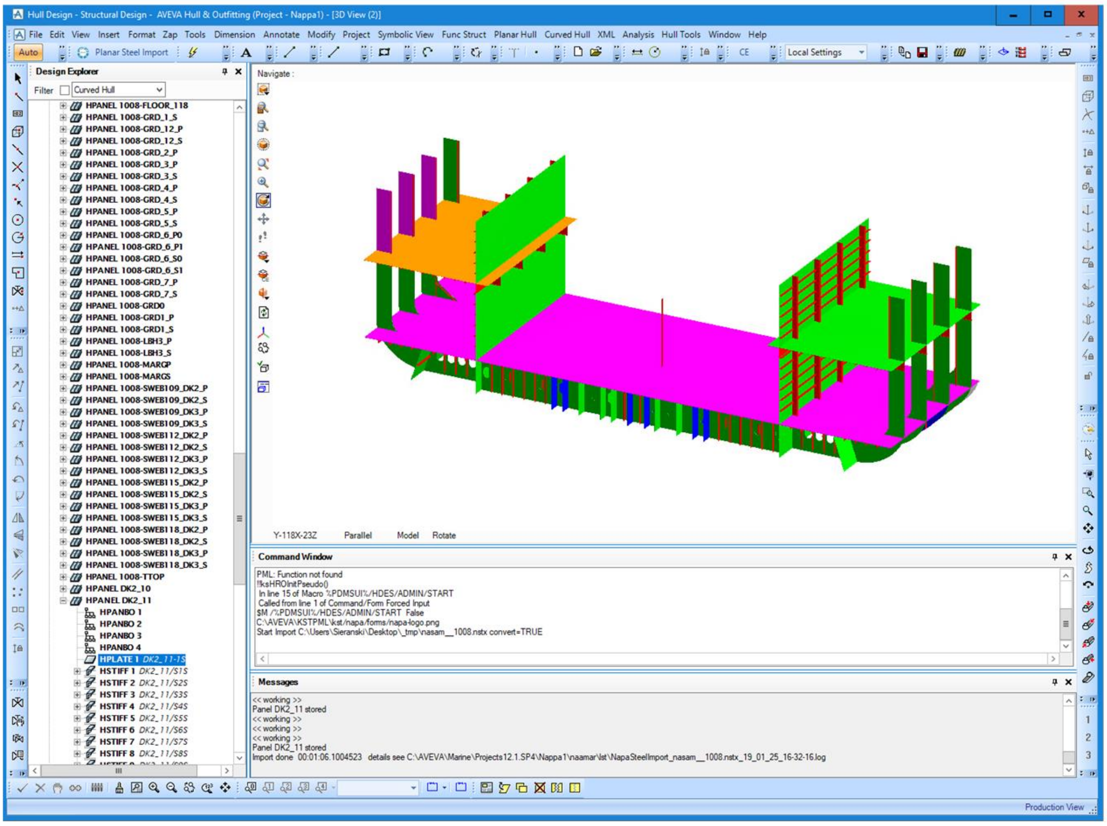


Fig.6: NAPA Designer Model imported to AVEVA Marine

We can also immediately see another benefit of our approach. Even though we started with the idea of getting data from NAPA Designer to SC and AM, we have now also implemented an interface between SC and AM, at no cost other than the cost of implementing a general integration layer, to begin with. Since both SC and AM ‘understand’ the general data model, they also understand each other.

3.2. SSI ShipConstructor and DigitalHub to Teamcenter

Now that we have had a look at the integration between different design tools, we will continue by looking at the integration between design tool and PLM system, as well as between two PLM systems. In this case, ShipConstructor is not used as a target for the design data, but as a source for other systems like ERP or production preparation. In addition to exporting, processing and mapping geometrical, parametrical and topological information, the connector to SC also implements endpoints that allow for the transfer of drawings and other files, as well as product information data.

This data is extracted independently from the OpenPDM SHIP XML file and as such can be addressed by the process engine in independent services. This way, we can, for example, export drawings from SC in pdf format and use the Teamcenter connector – again, in an independent process – to import these drawings into Teamcenter. As hinted at in chapter 2, we did not have to develop anything new for this Teamcenter connector. It was an already tried and tested component to OpenPDM and as such capable of file handling and processing of PLM-relevant data. Of course, the process engine has to be configured for each integration task, but no system-specific development work must be done once the connectors are implemented – regardless of use case or data source.

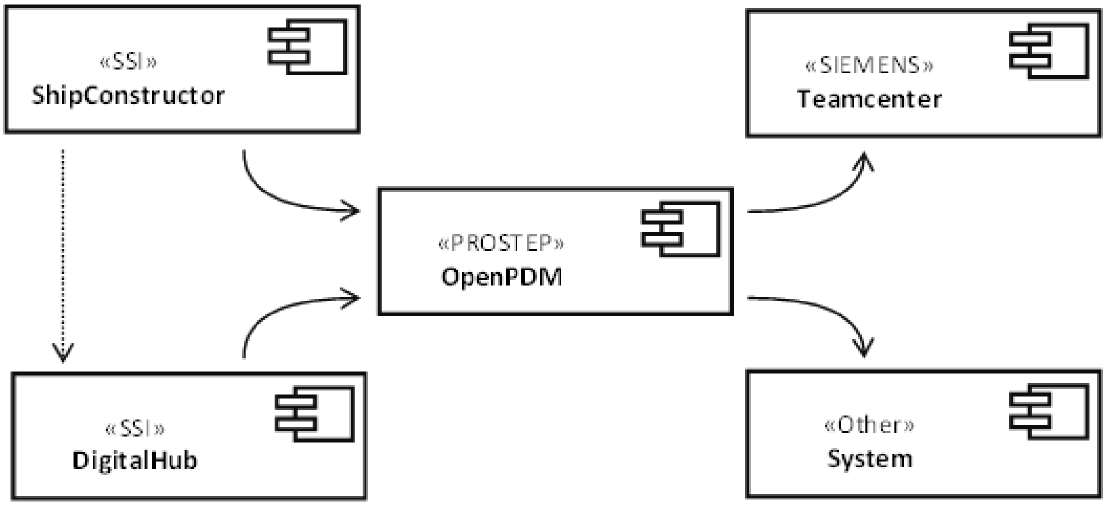


Fig.7: Several paths to connect ShipConstructor to other target systems

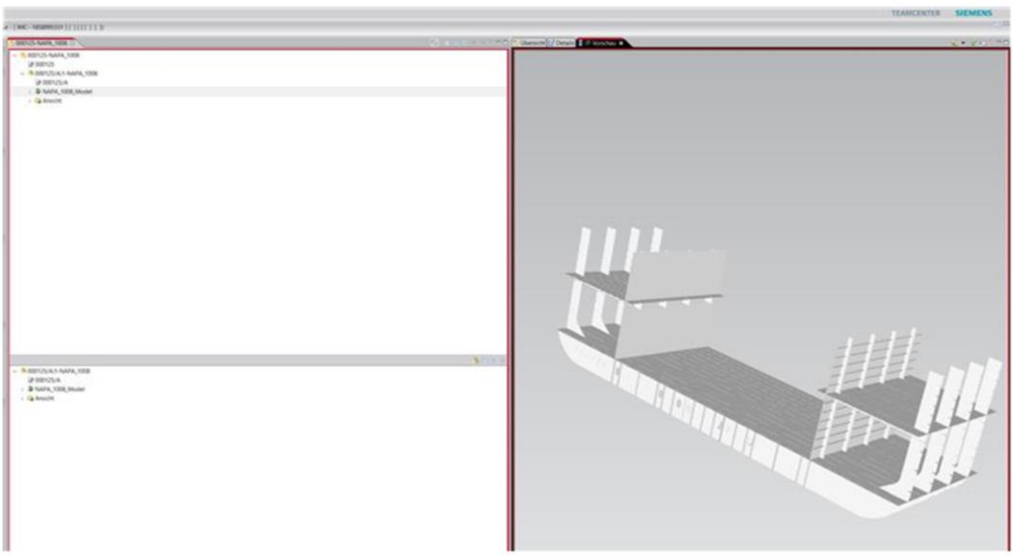


Fig.8: ShipConstructor design transferred directly to Teamcenter or via Digital Hub

In addition to its normal APIs, SSI now also offers the DigitalHub, a component to provide PLM functionalities tailored for shipbuilding processes with a built-in connection to ShipConstructor as the design tool. DigitalHub is based on Aras Innovator, so we can again use one of the existing PLM connectors to achieve an integration between DigitalHub and Teamcenter. The Aras connector uses the official IOM API provided by Aras to implement the connector endpoints and provides the same functionalities that the Teamcenter connector does – after all, they both implement the same general connector interface. Analogously to the connection of ship design tools via the common OpenPDM SHIP XML, we can connect DigitalHub and Teamcenter using our common PLM data model, BIBO (see chapter 2.1.). This means that no additional development work is required to achieve this data exchange between the two systems but to create a mapping from the DigitalHub to e.g. a Teamcenter data model.

4. Challenges

Of course, such a general approach focussing on system independence and the decoupling of the integration process from tool-specific logics and requirements does not come about for free. *Sieranski and Zerbst (2019)* explored the challenges a best-of-breed approach faces on a general level, so we will focus here on the practical challenges we faced while implementing the above integrations.

As mentioned in chapter 3.1. with the example of seams creating plates in NAPA Designer and AVEVA Marine, there were a few quirks unique to each design tool that needed to be accounted for when mapping the general data model to the tool-specific one. But these will also have to be addressed in any attempt to integrate the different applications, regardless of whether one uses a common point of reference like we did or a direct interface between two tools. The challenge that was unique to our approach was the development of the common data model for the OpenPDM SHIP XML itself. Different tools for different design phases require different types (geometrical, topological, etc) of data in different levels of detail. Enabling the model to both feed any design tool and accept input from any design tool, in turn, was thus far from a trivial task. Especially so, as we needed the model itself to be rid of any system-specific components. After all, we want it to be independent of any design tool to be in a position to fully reap the benefits of such an approach as detailed throughout this paper.

Additionally, the model had to be extensible. Future versions of applications or different tools altogether also need to be able to work with the data model, which might require amendments to it. Any such amendment mustn't break already working interfaces from other tools, without each of them needing adjustment itself (and thus undermining the idea of such a general approach as a whole).

We solved most of the issues as follows: the data model itself can store design data of any type, in virtually any practical level of detail. The different connectors for the design tools, on the other hand, know which data in the model to address and which of the information is or is not relevant to them.

Of course, exporting all pieces of information one possibly could extract from a design tool in any level of detail is quite impractical, both when it comes to performance as well as user comprehension. But there is a difference between a model being able to store all sorts of data, and a model requiring all this data to properly function. That is where the mappers included in the respective connectors play a big part. They are configurable to a degree that allows a yard to determine the exported form of the data in the OpenPDM SHIP XML to suit their toolkit. If a yard does not use a specific design tool, there is no need to include any information in a form or level of detail that is only useful for that tool. We thus add additional configuration on the connector side to keep the model free from any dependencies to any one specific design tool.

5. Conclusion

While certainly costly in the initial and continued development of the common data model, a ‘common point of reference’ approach offers a variety of benefits for integration between ship design tools and PLM systems. Independent modules being directed by a common process engine that itself relies on independent models for both PLM and design data, allows us to implement interfaces in a fast and structured way, often without any further development needed. System-specific configurations are carried out in the connector modules and no tool-unique logic is implemented outside of these connectors. Hence, all systems and tools that understand these common data models, also understand each other, enabling shipyards to implement fast, data-secure interfaces between any applications in their design and PLM toolkits.

References

SIERANSKI, J.; ZERBST, C. (2019), *Automatic Geometry and Metadata Conversion*, COMPIT Conf., Tullamore

Hull Form Design Support Tool based on Machine Learning

Tomoyuki Taniguchi, National Maritime Research Institute, Tokyo/Japan,
taniguchi-t@m.mpat.go.jp
Yasuo Ichinose, National Maritime Research Institute, Tokyo/Japan, ichinose@m.mpat.go.jp

Abstract

In this paper, a concept of the hull form design support tool based on the machine learning is described. We developed a short-estimation-time stern wake prediction method based on the convolutional neural network, and the process by which recommendations on the hull form are given is considered. The neural network model is trained using computational results by computational fluid dynamics (CFD). However, this method does not replace CFD, but extends design charts.

1. Introduction

In recent years, the propulsion performance design of ships, such as environmental requirements and EEDI, etc., requires designs that satisfy various purposes, requirements, and constraints. At present, designs utilizing computational fluid dynamics (CFD) are being performed in addition to towing tests. However, these take a long time to obtain the results, and the optimization of hull form takes several weeks in some cases. On the other hand, there are studies that uses machine learning for a simple estimation at the early stage of design. For example, *Matsumura et al (1998)* are studying a method for predicting performance values using neural networks. However, its input is contracted to some parameters such as length, width, Froude number, etc., and the degree of freedom of the hull shape is limited.

It is important for designers to understand the relationship between the hull form and the flow field, such as where to change the hull shape to obtain the desired flow field. In this paper, we focus on the supporting designers by presenting the wake results predicted by the machine learning model in a short time. The designer will be able to see the flow results almost as soon as the hull form is modified. We constructed a machine learning model that predicts the stern flow field using the hull shape expressed as an image. The model was trained using a pair of the hull form and the calculation result by CFD. The validity was confirmed by comparison with CFD results. And we also studied the optimization method using this machine learning model. The optimal hull form to fit the desired flow field was obtained in a few minutes. It is noted that our research intent is to develop cooperative design toolsets of CFD and machine learning according to the design process. The present method will extents the designer's knowledge and insight in the hull form design process, in which conventionally craftsman's knowledge and the design charts play a key role.

2. Related work

There are studies using machine learning for a prediction of the propulsion performance. *Matsumura et al. (1998), Kanai (2010), Margari et al. (2018), Cepowski et al. (2019),* have studied the multi-layer perceptron model to predict the performance values. These method predict the scalar value related to the propulsion performance such as the wave drag coefficient by using the parameters representing the curved surface of the hull and dimensionless numbers such as the Froude number. *Takagi et al. (2019)* make predictions using image data obtained by projecting the hull shape from various angles as input. These studies have good approximation results for predicting scalar value.

On the other hand, as a method of predicting the flow field, for example, *Kim et al. (2019),* developed the Deep Fluids, which is a generative machine learning model and performs time series fluid simulation such as smoke, spheres dropping on a basin and viscous dam break. *Ichinose et al. (2019),* have proposed a method of constructing a database of CFD calculation results and interpolating the stern flow field from the database.

In this study, in order to express the shape of the hull more flexibly and appropriately, we defined the input data in an image format with a different approach from that of *Takagi et al. (2019)* and *Kim et al. (2019)*. Our study is the extension of the work of *Ichinose et al. (2019)* to support designers' quick understanding and optimization using a generative machine learning model.

3. Design support tool based on machine learning

Designing the hull form satisfying many demands and a lot of restriction is not easy for not only a young designer, but also an expert. The designer need to understand the relationship between the hull form and flow field. In this paper, there are two purposes of the design support tool based on machine learning. The one is the training and supporting to understand the fluid flow for the designers, and the another is the recommendation and the automatic generation of the initial hull form design. Fig.1 shows the concept of propulsion performance design tool based on machine learning. The machine learning is conducted using the training data from the database with hull forms, CFD and experiment results. The designer can get the predicted flow instantly by changing the hull form and learn the relationship through the interaction to the machine learning model. And the recommendation of the optimal hull form is also shown. The verification is conducted by CFD and towing test. If the error of prediction is large, the re-training of the prediction model is also conducted and its precision increases. The optimal hull form will be determined in a less calculation of CFD. After this chapter, we will build a concrete machine learning model and verify the predicted results.

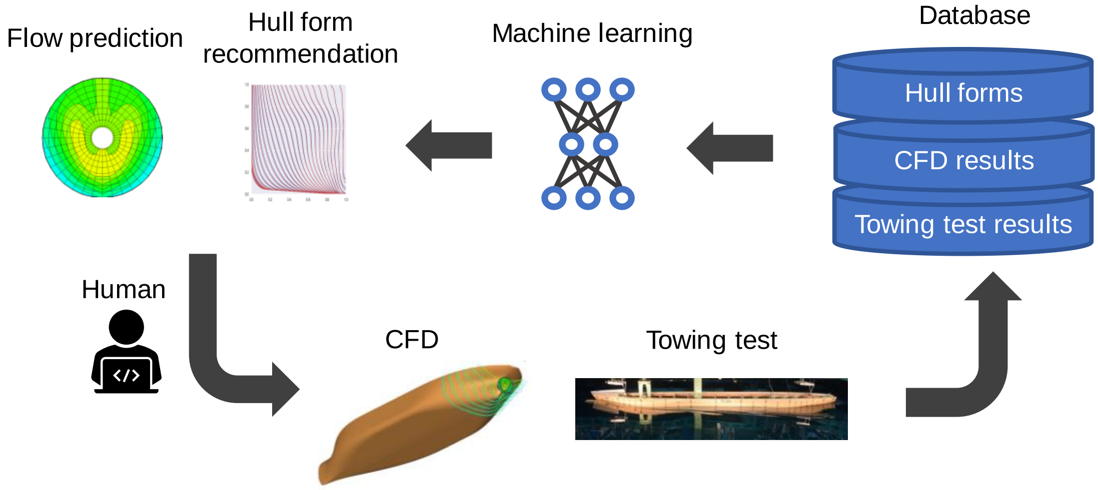


Fig.1: Flow of the design tool based on machine learning

4. A generative model for prediction of stern wake

In this section, a stern wake prediction method based on machine learning is presented. This method can shorten the time of the hydrodynamical performance prediction compared to the CFD method. The proposed method is assumed to be applied in the early design stage which usually has the limitation of time. This machine learning model is trained by the data set of hull form and CFD-estimated flow field. Conventionally, the hull form representations for machine learning had some limitation in the degree of freedom. The previous studies have attempted to represent the 3D curved hull surface with only 10~20 parameters. In our approach, the 2D structured grid expressing 3D curved surface is directly used for the input datasets of machine learning. This makes it possible to represent the hull shape with almost no restrictions. The velocity at the stern is calculated by CFD. The obtained velocity results are also expressed as image data in the same way as the hull shape, and the relationship between the hull shape and the stern flow field is used as training data. In this paper, a generative model was constructed with reference to the deconvolutional neural network used in the image generation model.

4.1. Representation of hull form and flow field

In CFD analysis, the hull surface which has three-dimensional coordinate (x, y, z) is represented with n × m two-dimensional structured grid utilizing the coordinate transformation method. This 2D structured grid data can map to the RGB color model image data with no restrictions as shown in Fig.2. The n × m two-dimensional grid which is a linear mapping of the original coordination of (x, y, z) is regarded as n × m × 3 image data which has three colors expression (Red, Green, Blue). The values of (x, y, z) are normalized, and these x, y, z correspond to RGB color model's R, G, B respectively. As a result, it can be used as an input to a convolutional neural network, and the hull shape can be expressed in a more practical form without significant restrictions. We supposed to call this conversion Image-based Hull form Representation (IHR). The grid used for training here is 64 × 128.

The stern flow field calculated by CFD is represented as discrete values on a concentric circle centred on the propeller axis. The flow velocity in the vessel longitudinal direction is also expressed as q × r × 1 image data divided into q in the circumferential direction and r in the radial direction. The partition used for training in this paper is 32 × 16. The nonlinear relationship between the hull shape as the image data and the stern flow field is predicted using a generative machine learning model.

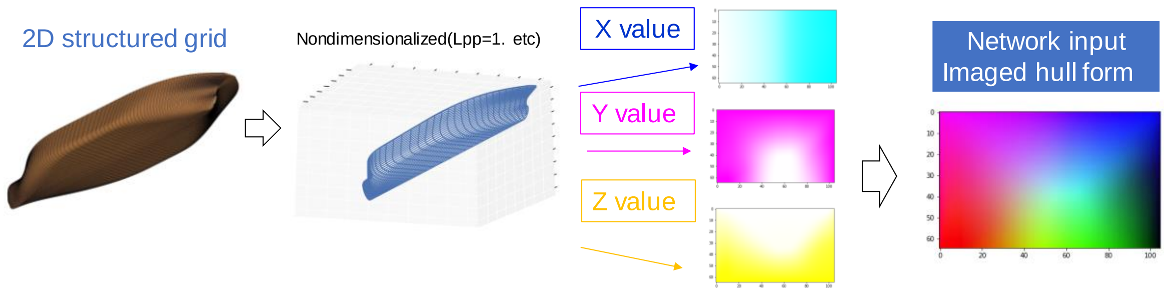


Fig.2: Image-based Hull form Representation (IHR)

4.2. Architecture of network

In this paper, we refer to the generative model such as DCGAN of *Radford et al. (2016)*. The stern flow prediction model with convolutional and transpose convolutional neural network is constructed. Fig.3 shows the architecture of the prediction model. Blue arrow shows the convolutional neural network, green arrows indicates the transpose convolutional neural network and black arrows denote the reshape operator. The left-side input is the IHR image data representing the hull shape described in Subsection 4.1, and the right-side output is the image data representing the stern wake flow. The features of input hull form are extracted by convolution, and the flow field is generated by transpose convolution.

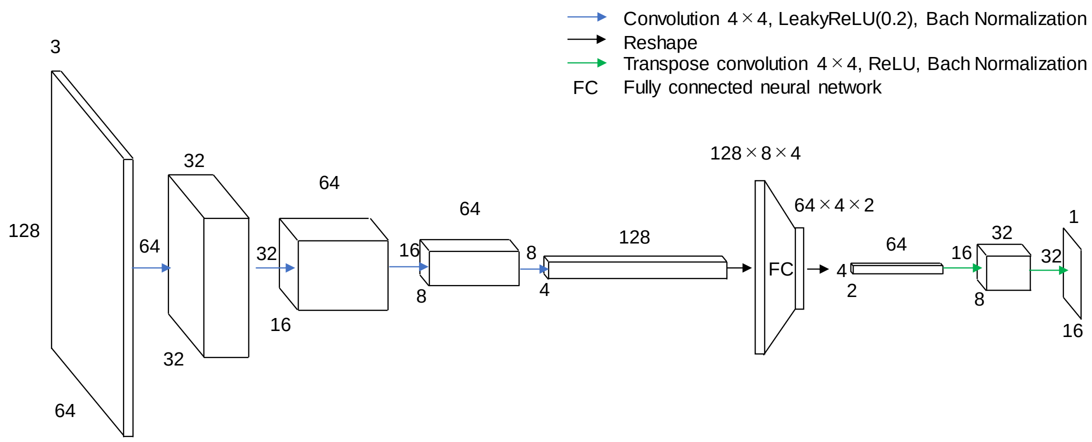


Fig.3: Architecture of the prediction model

4.3. Training data

As training datasets, we use the results of the stern flow field calculated by CFD analysis, NEPTUNE of *Hirata, etc. (1999)*. In this paper, the 16 different basic hull forms were created, and the datasets with 2200 hull shapes were constructed from these basic hull forms linear combination. Here, the sum of the coefficients of the linear combination is generated to be 1.0. Fig.4 shows examples of a line of basic hull forms. The left-side figure in Fig.4 shows the hull shape on the port side of the stern as an example. The vertical axis represents the vertical direction and the horizontal axis represents the width direction. The 8 lines shown in different colors indicate one hull shape line for each basic hull type. Similarly, the right-side figure shows the horizontal axis as the longitudinal direction of the ship. The basic hull forms have different stern shapes. Fig.5 shows an example of the stern flow velocity obtained by CFD analysis. The left figure shows the flow velocity of the stern flow field with a color map and contour lines, and the right figure shows the flow velocity at a radius = 0.7 with the horizontal axis as an angle. In this study, the left figure is predicted by machine learning.

Table I shows the datasets structure used for training and validation. 1947 data were used for learning and 253 data were used for verification. The basic hull shape data is represented from 0001 to 0016, the 1157 data is created by 8 linear combinations from 0001 to 0008, and the 790 data is generated by combinations from 0009 to 0016. A total of 1947 hull forms were used for training of the model. In addition, the 253 verification data is created by combinations from 0001 to 0004, 0009 to 00012. Since the shape of the hull is not expected to change significantly if the type of ship is determined, the variation from the average value of the basic hull types 0001 to 0008 is applied for machine learning. The stern flow field, which is the output, is also converted as the variation.

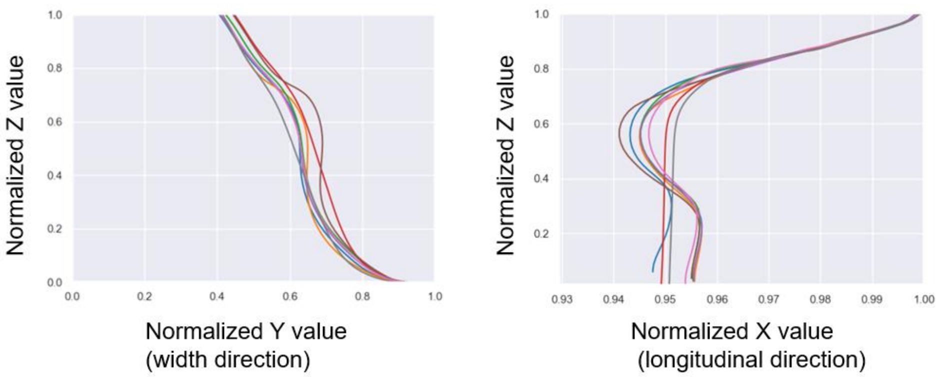


Fig.4: Examples of a line of basic hull forms

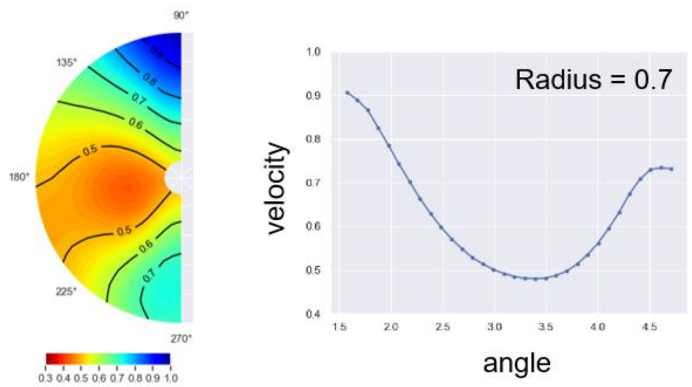


Fig.5: Examples of a stern wake results calculated by CFD

Table I: Train and validation datasets

	Datasets size		Basic hull forms number
Train dataset	1947	1157	0001 ~ 0008
		790	0009 ~ 0016
Validation dataset	253		0001 ~ 0004, 0009 ~ 0012

4.4. Loss function

In this study, we did not simply apply the error of the flow velocity to the loss function. The utilization of the gradient information is studied. The gradient of flow velocity related the fluid dynamic forces, it is important information for prediction. We added the terms of the gradient error to the loss function, which is the similar approach of Deep Fluids of *Kim et al. (2019)*. Here, the $\boldsymbol{I}_t, \boldsymbol{I}_p$ are the true and predicted stern flow image data, and the loss function L can be expressed as follows:

$$L = \mathrm{MSE}(\boldsymbol{I}_t - \boldsymbol{I}_p) + l \times \mathrm{MSE}(\nabla \boldsymbol{I}_t - \nabla \boldsymbol{I}_p) \tag{1}$$

Here, MSE indicates the function of mean square error, and l denotes the weight of the error of gradient of flow. It is expected that not only flow velocity but also gradient of velocity can be predicted accurately. Fig.6 shows the predicted examples without or with the gradient loss factor. The flow velocity at a radius = 0.7 with the horizontal axis as an angle is shown. The blue line is the correct answer, and the red line is the predicted value. The left-side figure shows the results without the gradient loss factor. There is the roughness around 2.5~3.5 rad. On the other hand, from the right-side figure with the gradient loss factor, the smooth line can be obtained.

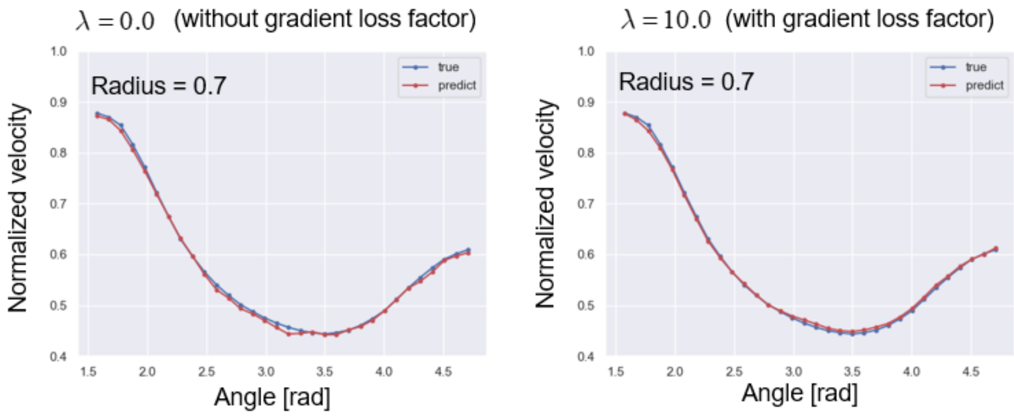


Fig.6: Comparison of the predicted result without or with the gradient loss factor

5. Results

The learning model described in Section 4 was trained using the training datasets described in Subsection 4.3. The batch size is 512, and the epoch number is 3000. First, Fig.7 shows the transition of the value of the loss function for the training data and the validation data. The vertical axis is the value of the loss function, the horizontal axis is the number of training epochs, the blue line shows the loss value for the training data, and the orange line denotes the value for the validation data. From this figure, the value also decreased for the validation data, and learning progressed without over-fitting until 3000 epochs.

Next, Fig.8 shows the examples of prediction results by the neural network. The above left-side two figures show the true and predicted stern velocity in the longitudinal direction by the machine learning model from the validation data using a color map and contour lines. We have a good matching results. The contour line of 0.6 around 4.5 rad is different, but here the value does not change much and the contour line changes by a slight difference, so the numerical error is not large. The above right-side figure shows the flow velocity value at a radius of 0.7. The blue line is the correct answer, and the red line is the predicted value. From this figure, the high accuracy of the prediction was confirmed. The figure below in Fig.8 shows the result with the largest error in the validation data. We can also confirm that the prediction is highly accurate even if the overall value of the velocity changes. However, the error is large near 4.5 rad. We regard this problem is a future subject on a practical accuracy for the examination at the initial design stage.

Although the calculation speed cannot be compared directly, we also check the calculation time of each method. When all 2,200 hull forms were analysed by CFD, 5 computers with 2 Intel Xeon Gold 6148 (20 core 2.2GHz) were used in parallel and it took about 120 hours. On the other hand, the prediction time for 2200 stern flow fields was about 18 seconds using 1 core of Intel (R) Core ™ i7-6800K. In the CFD analysis, the information other than the stern flow field in the longitude direction was also calculated, so it was not possible to make a general comparison, but it was confirmed that the calculation speed of the proposed method was very high.

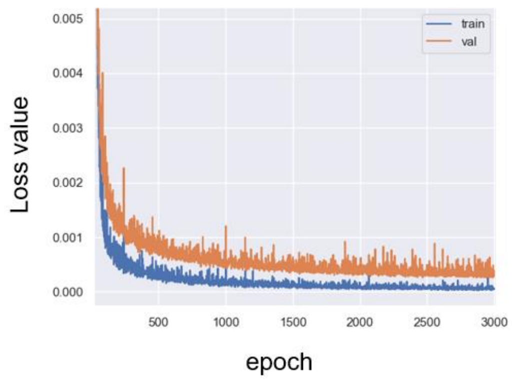


Fig.7: Examples of a stern wake results calculated by CFD

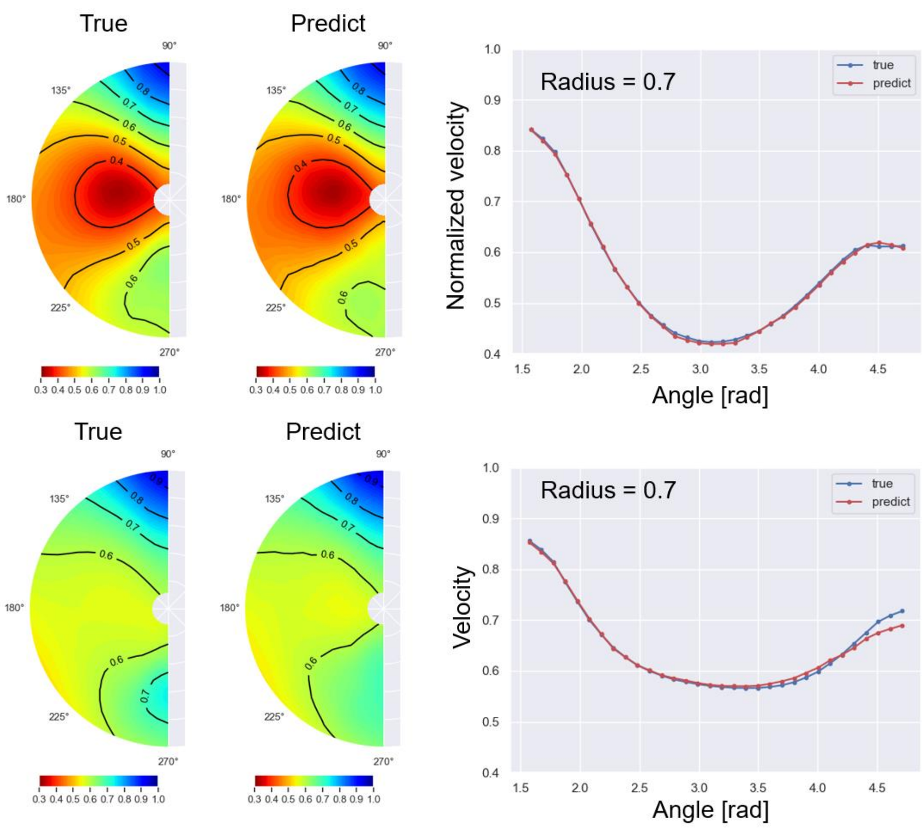


Fig.8: Examples of the prediction results by the proposed method

6. Optimization of hull form

At the early stage of the design, the optimal hull form is studied in consideration of various constraints and various purposes. For example, high propulsion performance, large dead-weight, high fuel-efficient, low-vibration are required. It is considered that the study speed can be dramatically improved by the proposed prediction method. Here, we consider the problem of optimizing the hull shape to achieve the target stern flow field, and consider hull form optimization combining this prediction method and the gradient method.

6.1. Formulation of optimization problem

In this paper, the hull form is assumed to be expressed as a linear combination using 16 basic hull forms, the design parameters are 16 weight coefficients $a_i\ (i=1\sim16)$. And, since these coefficient sums are adjusted to be 1, there are 15 parameters. The objective function is expressed as the mean square error from the target stern flow field, the optimization problem can be defined as a minimization problem that finds 16 coefficients that minimize this error with a constraint as follows:

$$\left.\begin{array}{l} \text{Find } a_i\ (i=1\sim16) \\ \text{Minimize } \mathrm{MSE}(\boldsymbol{v}_t-\boldsymbol{v}_p) \\ \text{Subject to } \sum_{i=1}^{16} a_i=1 \end{array}\right\} \quad (2)$$

Here, $\boldsymbol{v}_t, \boldsymbol{v}_p$ show the target and predicted stern flow velocity respectively.

6.2. Results of optimization

At first, 530 hull forms represented by the combinations of the basic hull forms from 0001 to 0004, 0009 to 0012 were generated besides the training and validation data shown in Table I, and its flow velocity were calculated by CFD. A solution with a small velocity gradient is selected as the target flow from 530 data, with the intention of suppressing the pressure fluctuation for low vibration. For this target flow field, the design parameters for minimizing the objective function defined in Section 6.1 were determined using the gradient method. The objective function can be calculated very quickly because this machine learning model is used. Fig.9 and Fig.10 show the optimization results. The left-side two figures in Fig.9 show the stern flow field, and we can confirm that the optimal solution close to the target flow field was obtained. From the right-side figure, it can be confirmed that the flow field is quantitatively close.

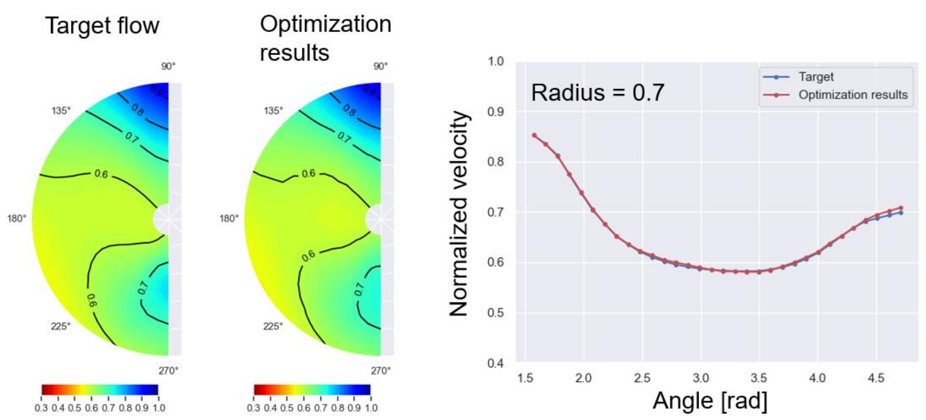
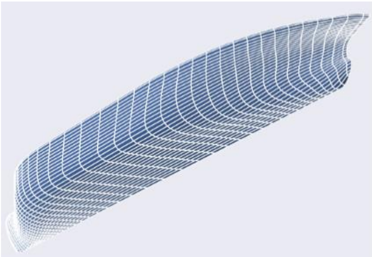
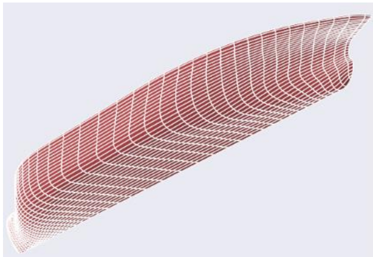
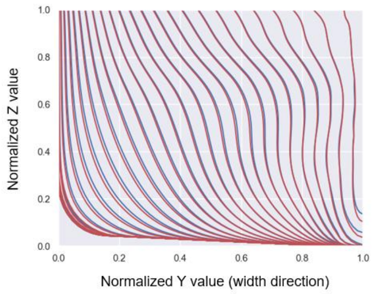


Fig.9: Examples of the prediction results by the proposed method

(a) target hull form (b) optimization results (c) hull form lines on stern

Fig.10: Examples of the prediction results by the proposed method.

Fig.10 shows the correct hull shape with the target flow field, and the hull shape obtained by optimization. The blue lines show the target hull forms, and the red lines show the hull form by the optimization calculation, and we can confirm that very similar solutions can be obtained. Also, these 1000 iterations are completed in about 6 minutes.

7. Conclusion

In this paper, as a prediction method of the flow field generated by hull form, a machine learning model was constructed with reference to the generative model. Its training was performed based on the dataset created by CFD analysis. As a result, high-precision results were confirmed for the validation data. In addition, we defined an optimization problem to find the target stern flow field, and optimized it using the gradient method. From a result of the one example, it was possible to obtain the hull shape that resulted in the desired stern flow field. In the early stage of the design, it can be expected that this estimation method and the optimization method will be combined to find the desired hull form, and to perform CFD analysis to confirm the results.

References

CEPOWSKI, T. (2020), *The prediction of ship added resistance at the preliminary design stage by the use of an artificial neural network*, Ocean Engineering 195

HIRATA, N.; HINO, T. (1999), *An Efficient Algorithm for Simulating Free-surface Turbulent Flows around an Advancing Ship*, J. The Society of Naval Architects of Japan 185, pp.1-8

ICHINOSE, Y.; TAHARA, Y. (2019), *A wake field design system utilizing a database analysis to enhance the performance of energy saving devices and propeller*, J. Marine Science and Technology 24, pp.1119–1133

KANAI, K. (2000), *Application of the Neural Network to Estimate of Ship's Propulsive Performance and Hull Form Optimization (in Japanese)*, Trans. West-Japan Society of Naval Architects, 99, pp.1-1

KIM, B.; VINICIUS, C.A.; THUEREY, N.; KIM, T.; GROSS, M.; SOLENTHALER, B. (2019), *Deep Fluids: A Generative Network for Parameterized Fluid Simulations*, Computer Graphics Forum 38(2)

MARGARI, V.; KANELLOPOULOU, A.; ZARAPHONITIS, G. (2018), *On the use of Artificial Neural Networks for the calm water resistance prediction of MARAD Systematic Series' hullforms*, Ocean Engineering, 165, pp.528-537

MATUMURA, T.; URA, T. (1998), *Preliminary Planning of Ship Hull Form based on Artificial Neural Networks (1st Report) -Estimation of Wave Marking Resistance for High Speed Ships- (in Japanese)*, J. Society of Naval Architects of Japan, 183, pp.91-100

RADFORD, A.; METZAND, L.; CHINTALA., S. (2016), *Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks*, Int. Conf. Learning Representations

TAKAGI, Y.; MIKAMI, K.; HINO, T. (2019), *Prediction of Wave Resistance Coefficient Based Ship hull images with machine learning (in Japanese)*, Japan Society of Naval Architects and Ocean Engineers, June, pp.393-398

On-board Human Operators: Liabilities or Assets?

Stig Eriksen, SDU, Odense/Denmark and SIMAC, Svendborg/Denmark, ser@iti.sdu.dk

Abstract

This paper analyses data from cargo ships in order to investigate how unmanned operation will impact the ability to detect and prevent near miss incidents from developing into marine accidents. Many claim that eliminating the presence of an on-board crew will eliminate accidents caused by human errors. Others caution that human error may not disappear with the elimination of the crew and that new incidents may occur because of it. The analysis finds that on-board human presence is attributable to some errors but that it at the same time is vital for detecting and stopping the development of incidents.

1. Introduction

Unmanned and autonomous ships are expected by many to revolutionize the maritime industry in the coming decades. With improved sensor and computing capabilities together with the elimination of the on-board crew, the ships of the future are expected to be safer and more efficient with regard to both operation and energy consumption, e.g. *Rødseth et al. (2012)*. There is, however, still considerable ambiguity in the use of the term autonomy in the maritime context. It is often claimed that autonomy does not imply unmanned operation but many of the expected benefits of autonomous operation come from the absence of crew and of the accommodation of the crew on board, *Eriksen (2019)*. One of the most frequently claimed benefits of the autonomous and/or unmanned Ships (UMS) is the elimination or at least reduction of human error. An estimated 75% - 96% of all marine accidents can be attributed to human error, *Allianz (2017)*, and it seems an obvious solution to simply remove the crew, and therefore the human element, from the Conventionally Manned Ships (CMS). Others, however, warn that it is not so easy. *Ahvenjärvi (2016)* explains that there is human interaction in all aspects of marine operation, not just the day-to-day on-board operation, and that all aspects are subject to human error. He explains that removing the crew may simply shift the human error from the ship's crew to the control room personnel, designers, programmers and maintenance personnel responsible for operating, designing and maintaining the UMS.

Even if the majority of accidents can be partly or fully attributed to human error, the onboard crew are often vital in the mitigation of consequences when accidents happen. *Wróbel et al. (2017)* examine 100 maritime accident reports and assess whether the accident would have happened on a UMS, and once it had - would its consequences have been different if there had been no one on board to counteract them. They conclude that the number of navigation-related marine accidents such as grounding or collisions may be reduced on UMS but that the consequences of the accidents will be greater due to limited possibilities for consequence mitigation. *Wróbel et al. (2017)* go on to explain that their study is limited in that they were only able to investigate incidents where accidents did happen and not when crews were successfully able to prevent an accident from happening.

It is fortunately the norm that ship operate without incidents or accidents for the overwhelming majority of the time. From time to time incidents happen that require the crew to react to prevent danger to the ship, crew or environment. Most of the time these incidents are dealt with without consequences to the ship, crew or environment and without the need to involve external parties or the authorities. The body of evidence on maritime accidents is very one-sided in that it mainly contains reports on instances when things go wrong, not when they go right. This paper will attempt to bridge that gap by examining incidents that were close to resulting in an accident but where an accident was prevented. By analysing reports on near miss incidents from ships this paper will investigate the role played by the on-board crew in the occurrence and detection of incidents. The paper will also evaluate how unmanned operation affects the ability to stop an onboard incident from developing into a maritime accident.

2. Theory

In this section the theoretical background for the analysis in this paper is presented. In Section 2.1, a brief explanation of the concept of a near miss is given and the rationale behind the reporting and processing is described. In Section 2.2, the conceptual accidents model used throughout this paper is introduced.

2.1. Near Miss

A near miss is described by *IMO (2008a)* as "A sequence of events and/or conditions that could have resulted in loss. This loss was prevented only by a fortuitous break in the chain of events and/or conditions". The loss can be human injury, environmental damage, or negative business impact. Reporting of near miss occurrences is a mandatory part of The International Safety Management (ISM) Code, *IMO (2008b)*, and serves to improve the safety of ships by learning from past experiences. Ships must report near misses to the company office, which must analyse and make or amend procedures or recommendations based on these. The object is to identify unsafe conditions and prevent the hazardous situation, event or unsafe act from occurring again, and perhaps this time resulting in a loss. The interest in reporting and investigating near miss incidents also stems from the idea that even though the near miss itself does not have consequences, there is proportionality between near misses, minor incidents and major accidents. This concept is called the "iceberg model" or "accident pyramid" and is developed by *Heinrich et al. (1980)*. An example of Heinrich's Pyramid can be seen in Fig.1.

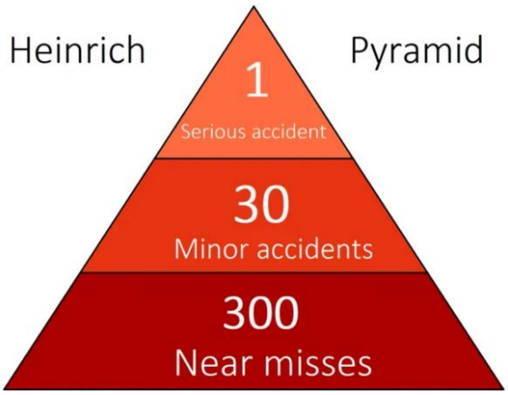


Fig.1: Heinrich's accident pyramid, *Colins (2018)*

The understanding of accident prevention has moved on since the pioneering work initially begun by *Heinrich et al. (1980)* in the 1930s. It is now widely accepted that the linear causal understanding of accidents that Heinrich proposes, and that the accident pyramid implies, is over-simplistic and does not accurately describe the complex failures of intricate systems, *Salminen et al. (1992)*. The ratio of major to minor accidents and near misses varies greatly across industries and the causes of minor incidents and major accidents are often not the same, *Hale (2002)*. Some, but not all major accidents can be predicted by minor accidents, and not all minor accidents have the potential to develop into major accidents.

Major accidents are often a consequence of a combination of unlikely circumstances and frequently come as a fundamental surprise to the operators, according to the "Normal Accident Theory", *Perrow (1999)*. It follows that the magnitude of accidents that were prevented from happening for one reason or another may not ever be realized. It may indeed never even be noticed that a major accident was close to occurring at all.

Evaluating safety performance based on the rate of near misses and/or minor incidents alone is a poor but often used practise, *Hale (2002)*. One issue is the tenuous relationship between minor and major incidents mentioned above. Another equally important issue is the discrepancy between actual and reported incidents. There are many reasons why near misses are not reported. Sharing the experience of an accident that nearly happened with the rest of the company could ideally prevent it from happening again, this time perhaps with serious consequences. There is, however, no obvious short-term incentive for the individual employee or supervisor to report the incident. The individual may also fear being blamed, disciplined, embarrassed or found legally liable, *IMO (2008a)*. If individuals feel that the company is unsupportive, complacent or insincere about addressing safety issues, they may be less inclined to use their often already busy work time on reporting.

The quality of reported near misses can also vary greatly. To incentivise reporting many companies require a minimum number of reports per month per ship. Some incidents may therefore be reported to satisfy a quota rather than because there is anything to be learned from them. The definition of a near miss can also be ambiguous. Some companies distinguish between unsafe conditions, unsafe acts, near misses, minor incidents, near accidents, etc. but others do not. Determining what qualifies as which, if any category, can be challenging. The reporting system for near misses is also sometimes used to report issues that do not qualify as near misses or are not directly related to safety. If a request for spare parts or service on equipment is not being met, for example, the ship's crew may deliberately use the near miss reporting system to circumvent the normal communication channels in the company. Safety-related issues are often referred directly to a senior safety officer rather than to the ship's superintendent.

Near misses are not a perfect representation of how exposed an organization is to major accidents but all of these reservations aside, there is good evidence for the importance of near miss reporting as a tool to improve safety, *Jones et al. (1999)*.

2.2. Bowtie method

One commonly used risk evaluation method to evaluate and understand causal relationships in high-risk scenarios is the bowtie method, an example of which can be seen in Fig.2. The method is named after the shape, which resembles a bowtie. Causes, also called threats, are on the left and Consequences, also called outcomes, are on the right. In the centre the bowtie converges in a Critical Event which is also called the top event, central event or centre event. The Critical Event can be defined as a release of hazardous substance or energy, *Markowski et al. (2009)*. Between the Causes and the Critical Event and between the Critical Event and the Consequences there are a number of Barriers. A Barrier is in place to prevent, control or mitigate the release of energy or hazardous substance. Barriers can be physical barriers or other physical equipment, or they can be non-physical, such as procedures or policies.

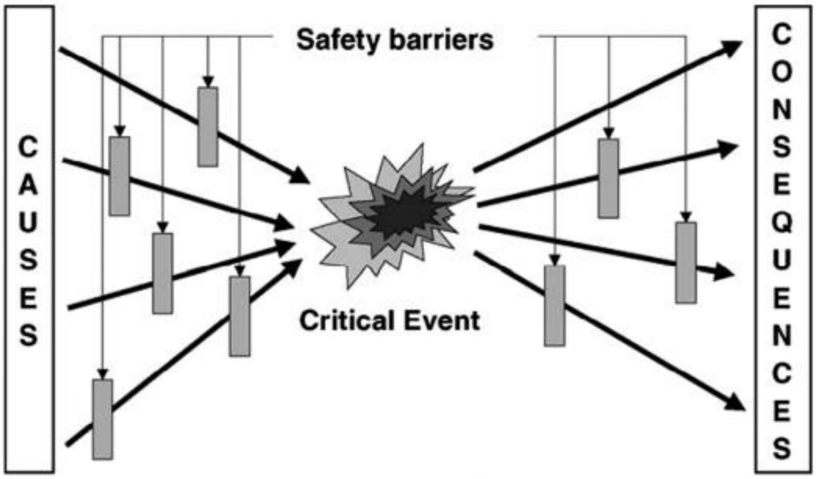


Fig. 2: Generic example of a bowtie, *de Dianous et al. (2006)*

The overall idea behind the bowtie method is that there can be several Causes of a Critical Event and that a Critical Event can have several Consequences. A Critical Event could be a fire. There can be many Causes of a fire breaking out and there can be many Consequences of the fire depending on the Barriers that are in place to control, contain or mitigate the fire. By focusing on the Critical Event, it becomes evident whether there are sufficient Barriers in place.

Although it is possible to define Causes, Critical Event and Consequences in the bowtie method, it can be difficult to distinguish clearly between the categories, *de Ruijter et al. (2016)*. Depending on the perspective chosen for the analysis, an electrical blackout on a ship, for example, could be described as all three categories. The Cause could be low lube oil level, the Critical Event could be low lube oil pressure and the blackout would then be the Consequence. If the blackout were the Critical Event, the Consequence could be grounding or collision. The grounding or collision could also be the Critical Event that had an oil leak or structural damage as a Consequence and the blackout as a Cause.

The bowtie model can be used both quantitatively and qualitatively. In this paper, it is used as a conceptual understanding of the development of accidents.

3. Data and data collection

The basis of this paper is the analysis of near misses. Safety-related information such as near misses often contains sensitive business information and is not normally publicly available, but one company was generous enough to supply near miss reports for the purpose of analysis. The name and operating segment of the company will remain anonymous, but the company operates one segments of cargo ships trading costal and worldwide. A total of 7205 near misses was initially examined. The near misses originated from 28 different vessels over a total of 126.9 years of operation between the start of 2012 to the middle of 2019.

The near miss reports consist of a brief Subject description averaging seven words and a slightly longer Description averaging 29 words. There are also accounts of the Immediate cause, Underlying cause and Corrective actions averaging 11, 20 and 16 words respectively. The length and detail of all the categories vary greatly, the shortest consisting of just one word while the longest contains more than 500 words. The date and time of the reports are available but details of position, operating mode, time of incident, weather conditions and other specifics are only occasionally included in the text, presumably if they are deemed to be of importance to the report.

4. Methodology

This section describes the method used in the analysis of near miss reports in this paper. Section 4.1. describes the operational scenario through which the analysis is done. In Section 4.2. the initial selection of near misses for analysis is described and in Section 4.3. the further categorization of the near misses is explained.

4.1. Scenario

The intention of this paper is to evaluate the UMS as being as close to the CMS as possible. Some assumptions about the on-board equipment and operation of the UMS must however be made for the purpose of the analysis. The operational scenario is based on that described in the project Maritime Unmanned Navigation through Intelligence in Networks, http://www.unmanned-ship.org/munin/partner/. In the scenario the UMS is assumed to be unmanned during sea passage in open sea, but is manned, remote-controlled or continuously monitored from shore during manoeuvring in and out of port, in narrow or heavily congested waters and while alongside berth. The shore control centre has access to all the inputs from the engine monitoring and alarm system of the UMS as well as all the bridge equipment. The control centre can start, stop and operate all the electronic or electronically actuated equipment as the on-board crew would normally do. It must also be assumed that the control centre can manipulate at least some of the valves, fire flaps, control levers, etc. that the crew normally operate manually. Some kind of video surveillance system is also assumed to be installed on board. When operating unmanned, the ship is assumed to be able to perceive and react to other vessels or objects and be able to react to the situation without outside control in the same way as a navigating officer would or to convey the information to a shore control centre that can then take over the control of the vessel. The placement, quantity, quality and maintenance regime of all machinery equipment on the UMS is assumed to be identical to that of the CMS except for remote control and monitoring capabilities. The added complexity and possible additional sources of failure resulting from the remote monitoring and operation capabilities are not considered.

4.2. Initial selection

As described in Section 2.2, it can be difficult to distinguish between Causes, Critical Events and Consequences in the bowtie model. This is especially true for near misses where the event was interrupted before it could develop into its full consequence. It is, however, not essential to distinguish between Causes, Critical Events and Consequences in this analysis. What is important is to differentiate between what will be described here as Active Events, which are those that can fall into the three categories mentioned above, and Passive Events which relate to the barriers. Active Events are non-routine events that will develop into a dangerous situation if not detected and stopped. Missing or defective barriers are Passive Events that, although they can be critical, will never in themselves develop into an accident without some other initiating event. Only Active Events are included in the analysis.

What the bowtie method described in Section 2.2. also shows is that there can be many different outcomes from one Critical Event. It is very hard to accurately estimate what the consequences would have been if they had been allowed to develop. In some cases, the event would have been close to resulting in a major accident, in other cases they may have been caught by another barrier before any consequences occurred. Only incidents that, if not detected and stopped, could plausibly result in very serious casualties or serious casualties as defined in *IMO (2008b)* “Casualty Investigation Code” are included. No ranking of criticality was done due to uncertainty about the possible consequences. Marine accidents as defined by *IMO (2008b)* are those that involve “Fire, explosion, collision, grounding, contact, heavy weather damage, ice damage, hull cracking or suspected hull defects etc.” Incidents that would “only” result in harm to single individuals, and not the ship and/or crew as a whole are not included.

4.3. Categorization

The categorization of the near misses was done in four steps. Near misses were first categorized into types of comparable incidents. There are large variations between the reported near misses and exact evaluations of the criticality and likely consequences of each incident were not possible, as described in Section 4.2. The incidents within each type were assessed to be sufficiently similar to make more generalized assessments of each type as a whole.

In the first categorization step, the outcomes from the initial selection were categorized into three near miss types; Fire, Flooding and Contact. The three types follow those of the marine accidents as described in Section 4.2. but are simplified into fewer categories since it is not possible to precisely predict what the outcome would have been had the near miss been allowed to develop into a marine accident.

The Fire category, which also includes the marine accident category Explosion, comprises incidents which could have resulted in a fire due to unattended or improper use of equipment, dangerous working practices, heavy running of equipment, short circuit or malfunction of electrical components and fuel oil leaks. Lubrication oil leaks are also included if they are specified as being near to a source of ignition such as a hot exhaust pipe. Improper storage of materials including garbage and oily rags is also included in the Fire category.

The Flooding category includes open watertight openings, corrosion or failure of material containing water and incorrect working practices leading to internal leaks.

The Contact category is an amalgamation of the marine accident categories Collision, Grounding and Contact. Incidents include broken mooring lines, miscommunication and human error during manoeuvring, faulty navigation, propulsion and steering equipment failure in confined waters, other vessels’ dangerous movements and own vessel dragging anchor.

In the second step it was noted whether the near misses were related to equipment that can be described as "crew comfort equipment" such as tumble dryers, bread toasters, cooking stoves, ovens and other galley equipment. It is highly unlikely that equipment such as this will be installed on a UMS and that it will be in use during unmanned operation if it is installed. The position or operating mode of the vessel at the time of the near miss is also noted when available in the near miss reports. Positions or operating modes are divided into the categories Sea, Anchor, Port and Manoeuvring which also include operation in restricted waters and channels. It is also noted whether the near miss is discovered by on-board Human Presence or by an existing Alarm or Measuring Point.

The third step is an evaluation of the cause of the near miss and is grouped into Human Error, Equipment Failure or External Influence, the latter meaning another ship, object or an extreme meteorological event. Only the immediate cause of the near miss is evaluated as the available data does not support a deeper analysis into the underlying causes. If an immediate cause cannot be established from the data, it is marked as unknown.

In the fourth and final step an evaluation is done of the effect of unmanned operation on the possibilities for stopping the development of the incident. This evaluation is based on the assumptions of the capabilities of the UMS described in the operational scenario in Section 4.1. as well as knowledge of the capabilities and design of machinery systems of the existing manned ship. This evaluation relies very much on the engineering knowledge of the author. The author has ship-specific knowledge of one of the ships from which the data originated as well as general knowledge about the construction and operation of marine systems in general from multiple other vessels of similar design. The evaluation is done based on the assumption that the design of the remaining ships from which the near miss data originated is not fundamentally different from that which is common in this type of vessel. This final step includes an evaluation of whether the possibilities of stopping an incident developing on a UMS would be Better, the Same or Worse compared to a CMS. A fourth category, Contain, was introduced meaning that the incident could probably be contained and not develop and cause further harm but could not be repaired until crew entered the UMS again.

5. Presentation and discussion of results

The initial selection described in Section 4.2 resulted in 481 near misses chosen for analysis. This means that only 6.7% of the total amount of 7,205 near misses met the sorting criteria. Some reported incidents were not safety-related and did not fall under the IMO definition of a near miss. Many incidents were related to incorrect work procedures and could or did only result in minor personal injuries such as cuts or slips. The majority of near misses were excluded because they only related to missing barriers such as missing personal protection equipment or broken or missing safety or monitoring equipment.

5.1. Near miss by type

Categorizing the near misses results in the distribution seen in Figure 3. The largest segment is Fire, which represented 40% of the near misses. Flooding, which represented 36%, is the second largest segment and the remaining 24% related to Contact. Of the 481 near misses analysed, 65 were related to equipment that can be described as "crew comfort equipment". All of these 65 near misses were in the Fire category and constitute 14% of the total number of near misses and 33% of the number of near misses in the Fire category.

For the comparison of the distribution of near misses to actual accidents, Fig.4 shows the distribution of global accidents. The accidents included in Fig.4 are those categorised as "very serious" and "serious" and that fall into the same three categories as the analysed near misses. The data is taken from European Maritime Safety Agency, *EMSA (2014)* and originates from the years 2011 to 2013. It is clear that there is a large discrepancy in the distribution of categories between the analysed near misses and actual reported accidents. One likely cause of this discrepancy is different ratios of minor incidents to major accidents for different segments as described in Section 2.1. It may be that there are generally more near misses per actual accident for the near miss types Fire and Flooding than for Contact. Ships often operate close to other ships or structures or with very little under-keel clearance. This is regarded as part of the normal operation but when things do go wrong there is very little margin of error. Fire and Flooding incidents are not part of the normal operation, so when a potential fire or flooding hazard occurs it is immediately rectified and reported. For Fire especially it is also likely that the bar for when an incident is considered potentially dangerous is very low. Fire is often described as the seafarer's worst nightmare and so even minor deviations are likely to be reported. For Flooding, unless it is very severe, there will typically be ample time to rectify the damage and in most cases, it will be possible to stop the near miss from developing into an accident.

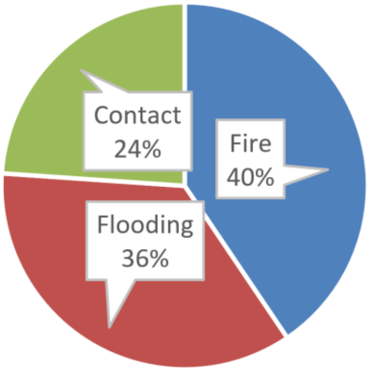


Fig. 3: Near misses by type – this analysis

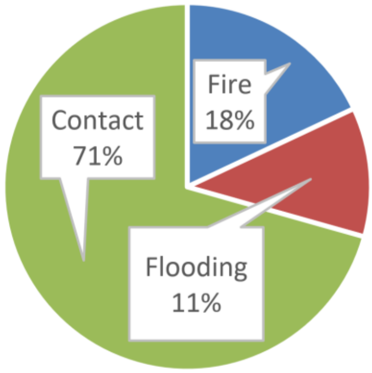


Fig. 4: Accidents by type - global values 2011-2013, *EMSA (2014)*

5.2. Position of ship at occurrence of near miss

Table I shows the reported position of the ship at the time at which the near miss occurred. For the majority of near misses relating to Fire and Flooding, no position was reported. It is likely that the particular position or operating mode of the ship was not judged to be of importance to the occurrence and possibility of rectifying the incident and was subsequently not reported.

For near misses relating to Contact, the position can always be inferred from the data, which makes sense since the traffic situation and/or geographical position is important to the understanding of the near miss. In only 8% of the situations where the ship was close to a collision with another ship or structure or close to a grounding, the vessel was in open sea. For the rest of the incidents, the vessel was either at anchor, manoeuvring into or out of port, transiting canals or other narrow passages or moored alongside a dock in port.

Table I: Position of ship at occurrence of near miss

	Sea	Anchor	Manoeuvring	Port	Unknown
Fire	14%	2%	2%	3%	79%
Flooding	30%	2%	3%	6%	59%
Contact	8%	14%	61%	17%	0%
Total	18%	5%	16%	8%	53%

5.3. Discovering near misses

Table II shows the means by which the near misses were discovered. There is some variation in the values across the categories, but the vast majority were discovered by human presence for Fire, Flooding and Contact. More near misses related to Flooding were discovered by an alarm or measuring point than for the other two categories. This is probably related to the fact that a near miss reported on Flooding almost always relates to an actual leakage as opposed to those reported on Fire for example, which often relate to immediate fire hazards but without the outbreak of an actual fire. If left undiscovered and unattended, a leak will eventually result in a bilge alarm.

Table II: Discovery of near misses

	Human presence	Alarm or measuring point	Unknown
Fire	94%	6%	0%
Flooding	77%	19%	4%
Contact	91%	5%	4%
Total	87%	11%	2%

All near misses are analysed based on the available material and it is possible that an alarm gave notification of the incident but that this was not reported in the data. This is especially relevant for near misses relating to Contacts. Alarms on radars and electronic chart displays can be customized to warn of close, or projected close proximity to other objects. This is a helpful feature in open sea where a large berth can and should be given to other ships and objects. In manoeuvring operations or when travelling in restricted water, however, ships are routinely required to operate much closer to other objects than would be regarded as safe in open sea, as also explained in Section 5.1. The evaluation of whether a situation is dangerous or safe relies on the interpretation of many different inputs about the actual situation and not least about the intended future actions of own and other vessels. Proximity alarms from radars and electronic chart displays are of very little value in these situations and are generally ignored or disabled.

5.4. Cause of near misses

Table III shows the causes of the analysed near misses across the three near miss types. Human Error is responsible for between 29% and 45% of the near misses. Equipment Failure is responsible for the majority of Fire and Flooding-related near misses but for only 14% of those related to Contact. Nearly half of Contact-related near misses, but virtually none relating to Fire and Flooding, are caused by External Influence.

Table III: Causes of near misses

	Human Error	Equipment Failure	External Influence
Fire	45%	55%	0%
Flooding	29%	69%	2%
Contact	42%	14%	44%
Total	38%	50%	11%

The number of near misses caused by Human Error across all three categories is lower than the corresponding value for incidents that resulted in actual accidents. Human Action was an Accident Event or a Contributing Factor to 65.8% of analysed marine accidents in 2018, according to *EMSA (2019)*. This discrepancy is probably due to a combination of different factors. Reporting bias must certainly be considered as one of these; individuals may be more inclined to report on equipment failures or failures resulting from external influences than on their own or their colleague's unsafe behaviour, since this could reflect badly on them personally, as described in Section 2.1. It may also be easier to define and recognize faulty equipment than unsafe behaviour.

Another factor may be that the data used in this paper did not support a deeper analysis of underlying or contributing factors of the incidents. If the data had enabled a more thorough investigation, it is possible that some near misses related to Equipment Failure would have been found to have a Human Error element due to poor maintenance or maintenance practice, wrong operation, inadequate design, etc.

The discrepancy may also be influenced by the fact that near misses in this paper are only analysed from the perspective of the vessel reporting the incident. Had an incident resulted in a collision for example, an investigation involving both vessels would have been conducted. It is likely that a large proportion of the near misses that are attributed to External Influences in this paper also have a Human Error element.

Yet another contributing factor could be that the causes of major accidents are fundamentally different from minor accidents/incidents, as explained in Section 2.1., in that the former are often a result of multiple failed preventive and mitigative barriers. One or more of these barriers are often of a human nature and therefore Human Error will be at least a contributing factor in many major accidents. Near misses are often the result of the failure of one single barrier. Other barriers, one or more of which could be of a human nature, then prevent this near miss from developing into an accident.

5.5. Stopping near misses from developing into an accident

Table IV shows the distribution of the evaluated possibility of stopping near misses developing into more serious incidents on UMS compared to CMS. Near misses that are related purely to "crew comfort equipment" have been excluded from the data in Table IV. The table clearly shows that there are large differences between the three near miss types. For Fire and Flooding it would be possible to Contain but not repair about half of the near misses. For the other half, however, the possibility of stopping the development on a UMS is evaluated as Worse.

For Contact there would be the Same possibility of stopping the incident from developing on a UMS as on a CMS for almost all the analysed near misses. This is closely related to the operational scenario, described in Section 4.1. The UMS is assumed to be either manned or remotely controlled during manoeuvring and in port. As seen in Section 4.2, the vast majority of near misses relating to Contact occur during these phases of the vessel's operation.

Table IV: Possibility of stopping development of near misses on UMS

	Same	Contain	Worse	Unknown
Fire	2%	49%	49%	0%
Flooding	0%	56%	43%	1%
Contact	97%	3%	1%	0%
Total	27%	39%	33%	1%

6. Discussion

Near misses are not a perfect proxy for the distribution and frequency of major accidents as described in Section 2.1. Most of the near misses were discarded from the analysis under the selection criteria described in section 4.2. Even among the 481 near misses that remained for analysis there may be large differences in criticality and potential severity. Some incidents may have been minutes away from developing into a serious marine accident had they not been detected and stopped. Other incidents might have been stopped by several other layers of barriers and would therefore not have had any consequences, even if they had escaped initial detection and mitigation. The estimation of criticality and severity is merely speculative, because near misses are accidents that did not happen. Near misses are, despite all these reservations, a valuable source of information about accidents that almost happened but were prevented for one reason or another. Information of this type is generally very hard to come by and has so far been missing from the discussion about unmanned operation of ships. The volume of near misses in this analysis is believed to be of a sufficient quantity to make more generalized conclusions about the nature of the detection and mitigation of the initial stages of maritime accidents. It would, however, have strengthened the analysis had near misses from other shipping companies operating other types of vessel been available.

Human Error is underrepresented in the near misses in this analysis compared to the actual accidents for all three accident types as described in section 5.4. The term human error covers a wide variety of faults, one of which is "omission", which is the failure to act. The discrepancies between near misses and actual accidents point to the role of the crew in managing to act to stop the development of incidents. Had the crew failed to act, the near miss could have resulted in an accident that would have been attributed to Human Error due to omission. The near misses attributed to Human Errors in this analysis are related to direct human actions on board as described in Section 5.4. Some of these near misses can be assumed to disappear on the UMS, but not all. Some of the near misses relate directly to the maintenance work being done on board. The majority of this work will still have to be carried out on the UMS but since it cannot be done while the ship is at sea, it will have to be compressed into the short periods when the ship is in port or close to shore. Some of the near misses attributed to Human Error will follow the maintenance work and will not disappear with unmanned operation.

Human Errors are attributable to 45% of Fire-related near misses and 29% of Flooding-related near misses in this analysis but 94% of Fire-related near misses and 77% of Flooding-related near misses are discovered by Human Presence. For almost half of the same two categories the possibility of stopping the incident developing is evaluated to be worse on a UMS than on a CMS. This has worrying implications for UMS because the majority of incidents can still be expected to occur but there are drastically limited possibilities for detecting the incidents and for stopping the incidents when they are detected. This finding supports the finding of *Wróbel et al. (2017)* in their analysis of 100 marine accidents. They find that, in general, there are more accidents that are less likely to happen on a UMS than on a CMS. For the Fire and Flooding categories in isolation however, they find the opposite to be the case. They also find that the consequences will be more severe for the majority of Fire and Flooding accidents on UMS

This analysis found that an overwhelming majority of near misses were detected by on-board human presence. This should not be taken to mean that the incidents cannot or will not be detected on UMS. Many of the near misses, especially those relating to Fire and Flooding, would eventually have been detected with the existing systems. Detecting the incidents later, however, means that they will be more developed. With more developed incidents there will be less time to stop the development of the incident into an accident. This is particularly problematic for the Fire category, where 94% of the analysed incidents were detected before an actual fire had broken out and an alarm was activated. The existing fire detection system can only warn of a fire when smoke, flame or heat has developed, at which point the incident will already have developed into, or be very close to, a marine accident. In addition to the difficulties of detecting near misses on UMS, the possibilities for stopping the development of incidents is also found to be worse for many near misses.

The possibilities for stopping incidents relating to Contact are evaluated to be the same on UMS as CMS for the vast majority of near misses. The position of the ship at the time of the near miss as described in Section 5.2. should, however, be considered because 92% of near misses related to Contact occur at times when the ship is assumed to be manned. The effect of unmanned operation on incidents relating to Contact must be considered to be very limited under the scenario described in Section 4.1. According to *EMSA (2014)*, 71% of accidents are related to Contact, but the occurrence of 92% of the Contact-related incidents in this analysis will remain unaffected by the introduction of UMS, because the ships are assumed to be manned at the time at which the incidents occurred.

In the evaluation of the possibilities for stopping the accident the Contain category was introduced, meaning that the near miss could be contained in the same way as on a CMS but not repaired while the ship was unmanned. It can be argued that this Contain category is then equal to the Same category, which is true if one looks only at the consequences of the near miss in isolation. Containing a near miss often means taking a system out of service to isolate a leak for example. There is already redundancy in many systems on CMS and stopping one system will have no direct consequences. Taking one system out of service, however, makes the ship vulnerable to failures in the remaining functioning systems. The longer the ship has to operate without the possibility of repairing the failure to the isolated system, the more exposed the ship is to experiencing failures to the remaining operational systems.

The possibility of stopping the development of incidents on UMS was not found to be greater than on CMS for any of the analysed near misses. This might seem very critical, but based on the operational scenario described in Section 4.1. it is not surprising. In the comparison, the UMS and CMS were assumed to be as identical as possible. There are really no options for stopping or mitigating accidents on a UMS that would not also be possible on a CMS. There are, however, many manual actions relating to damage control and emergency repairs that would only be possible on a CMS. Under the criteria described in Section 4.1. it would in fact be hard to imagine a scenario where the possibility for stopping the development of an incident would be better on a UMS than on a CMS.

No UMS of the scale of a modern cargo ship is in operation anywhere in the world today but a UMS would almost certainly not be designed simply as an unmanned version of today's CMS. With no actual operational data available for the UMS, however, the CMS must be used as a reference, which results in several uncertainties. The capabilities of the UMS for stopping the development of an incident are not known. Based on the results of this and other analyses, it seems likely that the UMS will have to be equipped with more redundancy in the mechanical systems and more automated or remote-controlled equipment to be able to detect and handle failures and breakdowns. These improved capabilities would favourably influence the analysis of the ability of UMS to detect and stop the development of failures. On the other hand, the added complexity resulting from these extra systems would be the cause of new yet unknown near misses. Incidents occurring because of unmanned operation represent a substantial unknown in analyses comparing UMS and CMS. Most of the reported and analysed incidents and accidents could have been prevented with hindsight and it can always be claimed that the systems on UMS will be designed so that these failures do not happen. It must be remembered, however, that the system was not designed to fail on the CMS either. The systems were designed and built to perform their task but failed to do so anyway. Construction and operation faults are almost always unexpected events that occur despite the intention of the designer and builder. Systems fail to perform as expected or are subjected to conditions that the designer did not intend. This will also be the case on UMS.

6. Conclusion

In this paper, 481 near misses are analysed in order to examine what role on-board human operators play in the occurrence and detection of the initial stages of marine accidents. The effect of unmanned operation on the ability to stop the development of incidents into accidents is also examined. Near misses are not a perfect representation of the distribution of marine accidents. Many near misses do not have the potential to develop into a marine accident and not all marine accidents can be predicted by near misses. Out of a total of 7,205 near misses, only 481 were evaluated as likely to result in a serious marine accident and were further analysed.

The near misses were divided into three types; Fire, Flooding and Contact. Compared to actual marine accidents, the near miss types Fire and Flooding were over-represented and Contact was under-represented. The explanation for this discrepancy is believed to be a combination of factors relating to the criteria for when incidents are reported. The discrepancy is also seen as evidence of the on-board human operators' role in preventing these types of incident from developing from a near miss without consequences to a serious marine accident.

It was found that 87% of near misses were discovered by humans, i.e. the on-board crew. In many cases, existing systems would probably have detected the incident eventually but only after the incident had become more severe. This is an indication of the need for significantly more monitoring and failure detection equipment on ships under unmanned operation.

For near misses related to Fire and Flooding, it was evaluated that it would be possible to contain about half of the incidents without further consequences during unmanned operation. For the other half, the possibilities for stopping the development of the accident without the presence of on-board human operators would be worse. In almost all near misses related to Contact, the possibility of stopping the development of the incident would be the same on an unmanned ship as on a manned ship. This evaluation relates very much to the finding that the vast majority of 92% of the near misses related to Contact occur when the ship is manoeuvring, at anchor or in port. The vessels are assumed to be manned during these operation phases, under the operational scenario chosen for this analysis. Unmanned operation can therefore only be expected to have little impact on Contact-related accidents. The majority of actual marine accidents are Contact-related.

Across the near miss types, 38% of incidents were found to be caused by direct on-board human error. Not all incidents related to human error can be expected to disappear with unmanned operation. At least some of the human errors must be assumed to occur following the maintenance work that, where not possible to do at sea, must be concentrated in the short periods where the ship is accessible in port or close to shore. Some near misses and accidents are direct causes of on-board human error that could be prevented if tasks could be automated, while others would be caused by the introduction of unmanned operation. Whether the introduction of unmanned operation will result in a net decrease or increase in the number of incidents and accidents is still uncertain and will depend entirely on the technical capabilities of the unmanned ships of the future. On today's conventional ships, however, with the existing technical systems, humans are vital in both the detection of incidents and the ability to stop the development of incidents into accidents.

Acknowledgments

This paper was written in connection with the PhD project entitled "Autonomous Ships from the Perspective of Operation and Maintenance". The project has been generously funded by The Danish Maritime Fund, Lauritzen Fonden, A/S D/S Orient's Fond and Svendborg International Maritime Academy. The author would like to thank PhD Project Supervisor Marie Lützen and Co-Supervisor Mads Bruun Larsen for their support and contributions.

References

AHVENJÄRVI, S. (2016), *The Human Element and Autonomous Ships*, TransNav, Int. J. Marine Navigation and Safety of Sea Transportation 10**,** pp.517-521

ALLIANZ (2017), *Safety and Shipping Review 2017,* Allianz

COLINS, D. (2018), *What is the Heinrich Pyramid in Safety Management?*, https://www.pro-sapien.com/blog/heinrich-pyramid/

DE DIANOUS, V.; FIÉVEZ, C. (2006), *ARAMIS project: A more explicit demonstration of risk control through the use of bow–tie diagrams and the evaluation of safety barrier performance*, J. Hazardous Materials 130, pp.220-233

DE RUIJTER, A.; GULDENMUND, F. (2016), *The bowtie method: A review*. Safety Science 88**,** pp.211-218

EMSA (2014), *Annual Overview of Marine Casualties and Incidents 2014,* European Maritime Safety Agency, Lisboa

EMSA (2019), *Annual Overview of Marine Casualties and Incidents 2019,* European Maritime Safety Agency, Lisboa

ERIKSEN, S. (2019), *Autonomous Ships – Changing Perceptions and Expectations,* 18th Conference on Computer and IT Applications in the Maritime Industries, Tullamore, pp.33-50

HALE, A. (2002), *Conditions of occurrence of major and minor accidents: Urban myths, deviations and accident scenarios,* Tijdschrift Voor Toegepaste Arbowetenschap 15

HEINRICH, H.W.; PETERSEN, D.; ROOS, N. (1980), *Industrial Accident Prevension: A Safety Management Approach*. New York

IMO (2008a), *Guidance on Near-Miss Reporting*, International Maritime Organization

IMO (2008b), *MSC-MEPC.3/Circ.3. Casualty-Related Matters Reports on Marine Casualties and Incidents*, International Maritime Organization

JONES, S.; KIRCHSTEIGER, C.; BJERKE, W. (1999), *The importance of near miss reporting to further improve safety performance*, J. Loss Prevention in the Process Industries 12, pp.59-67

MARKOWSKI, A.S.; MANNAN, M.S.; BIGOSZEWSKA, A. (2009), *Fuzzy logic for process safety analysis*, J. Loss Prevention in the Process Industries 22**,** pp.695-702

PERROW, C. (1999), *Normal accidents : living with high-risk technologies*, Princeton University Press

RØDSETH, Ø. J.; BURMEISTER, H.-C. (2012), *Developments toward the unmanned ship*, DGON ISIS 2012, Berlin

SALMINEN, S.; SAARI, J.; SAARELA, K.L.; RÄSÄNEN, T. (1992), *Fatal and non-fatal occupational accidents: identical versus differential causation*, Safety Science 15, pp.109-118

WRÓBEL, K.; MONTEWKA, J.; KUJALA, P. (2017), *Towards the assessment of potential impact of unmanned vessels on maritime transportation safety*, Reliability Engineering & System Safety 165**,** pp.155-169

Towards Safety Regulations for the Design of Autonomous Ships

Jiri de Vos, Delft University of Technology, Delft/The Netherlands, jiri.devos@tudelft.nl
Robert Hekkenberg, Delft University of Technology, Delft/The Netherlands, r.g.hekkenberg@tudelft.nl

Abstract

Many current ship design regulations are in place to safeguard the lives of seafarers. With the possible advent of unmanned autonomous ships, these regulations no longer serve this purpose and may, if kept in place unaltered, unnecessarily limit a designer's freedom to develop an efficient ship. Using the concept of equivalent safety, it is explored which regulations are the most urgent candidates for reconsideration. To support the discussion, a casualty analysis is performed in order to demonstrate for which casualty types, ship types and ship sizes the impact of loss of human life on the overall safety level is largest.

1. Introduction

Ever since the start of maritime transport, it has been the intention that ships should, among others, be safe to use for crew and passengers. Over the centuries, more and more regulations have been introduced and revised to improve the safety of ships, often as a response to ship accidents and disasters. However, the regulations cannot ensure that a ship is completely safe. Nonetheless, regulations serve to ensure that the safety level of a ship is sufficiently high, such that the remaining risks are deemed acceptable.

Within the boundaries of the regulations, the maritime industry strives to provide transport as efficiently as possible. In most cases, regulations set requirements to design aspects that have an influence on the safety level of a ship, so called prescriptive regulations. Examples of such requirements are a minimal initial metacentric height (GM_0) to reduce the probability of capsizing or the installation of fireproof walls to minimize the consequences of a fire. However, these prescriptive regulations also constrain the design and the transport efficiency.

Apart from constraining conventional designs, prescriptive regulations may also prevent the use of innovative designs, even though these could be as safe as and more efficient than the original conventional design. In order to stimulate the development of innovative designs, designers may challenge existing regulations. One approach is to demonstrate that the innovative design has an equivalent level of safety compared to conventional designs (MSC.1/Circ.1455, 2013). In this article we explore which regulations can or should be reconsidered first in order to enable efficient unmanned autonomous ships.

1.1 Autonomous ships

In recent years, the research on autonomous ships has increased significantly. The feasibility of autonomous ships has been explored in several projects, among which the MUNIN project, *MUNIN (2016)*, the AAWA project, *Rolls-Royce (2016)*, and the YARA Birkeland demonstrator, Kongsberg (2017). As a response to the increasing research by the industry, IMO is currently performing a regulatory scoping exercise (MSC 99/WP.9, 2018) in order to assess how existing regulations may be affected by autonomous ships.

An important incentive for the realisation of autonomous ships is to be able to reduce the crew significantly or even to allow a ship to sail without people on board. This can be beneficial for the future, since a shortage of nearly 150,000 additional officers is expected by 2025, *BIMCO & ICS (2015)*, and the cost of a crew is significant, especially for smaller ships. Also, it is speculated that introducing autonomous ships will reduce the number of accidents, since it is widely accepted that 80% of the accidents are due to human error. However, the exact influence of increased autonomy on the number of incidents is still unknown, since accident statistics do not show how many accidents have been prevented by humans on board, nor is it known which percentage of incidents can be avoided through autonomous navigation. In this paper it will be conservatively assumed that there will be no noticeable effect on the number or severity of accidents when ships will become autonomous. However, a definite consequence of sailing without people on board, is that lives cannot be lost on the ship. This leads to a safety benefit, since the risk of losing life is no longer present.

Besides the explorative projects mentioned above, most research performed in the industry focusses on route planning and collision avoidance. Examples of this are the work of *Beser and Yildirim (2018), Huang et al. (2020)* and *Ramos et al. (2019).* Only a few have described the need for new regulations. Especially DNV GL and Lloyd's register have expressed their belief in the need of a new regulatory framework in their view on the future, *Lloyd's Register et al. (2017), Vartdal et al. (2018).* These two classification societies, as well as Bureau Veritas, have described what they believe the new regulatory framework should look like, *Bureau Veritas (2017), DNV GL (2018), Lloyd's Register (2017).* However, how these new regulations can be obtained and evaluated is not described in detail.

1.2 Equivalent safety

As mentioned above, innovative designs can be accepted if equivalent safety is upheld. The same concept is used by the classification societies that shared their concepts for a new regulatory framework. DNV GL has described this as "autonomous ships shall be as safe as conventional ships" and other instances have published similar statements. The concept of equivalent safety for ships has been used in previous work by the authors as well *Vos and Hekkenberg (2020)* and a brief explanation of the concept will be given next.

IMO defines safety as "(…) the absence of unacceptable levels of risk (…)" (MSC.1/Circ.1455, 2013). As a result, it can be said that something is safe as long as it is subjected to an acceptable level of risk. Subsequently, two ships will be equally safe when they are subjected to the same level of risk.

Risk is the likelihood that an undesirable event will occur multiplied by the consequences of that event. This means that risk is measurable and can be calculated. Generally, the overall level of risk is determined by performing a risk analysis. In a risk analysis, the possible undesirable events are determined and per event the risk is calculated. These risks are then added together to form the overall level of risk.

In the case of equivalent safety, it must be ensured that the overall level of risk that the innovative design is subjected to is not higher than the overall level of risk the conventional design is subjected to. When comparing unmanned autonomous ships with conventional ships, the overall level of risk of the first will not include the risk of loss of life. Assuming that the ship and its behaviour will not change, the unmanned autonomous ship will thus be safer, since it is subjected to a lower level of risk.

Subsequently, it can also be argued that for unmanned autonomous ships the risk of one or more of the remaining undesirable events (i.e. all events apart from loss of life) are allowed to slightly increase, as long as the overall level of risk remains equal to or lower than the overall level of risk of a conventional ship. Consequently, the regulations that are in place to limit the remaining risks can be loosened proportionally for unmanned autonomous ships.

1.3 Challenging regulations

Summarizing the above, it is the intention to explore which regulations should first be reconsidered to keep an equivalent safety level but remove unnecessary high-impact restrictions in design freedom. The exploration will be based upon a casualty analysis using accident data of cargo ships from 2000 to 2018. The purpose of the casualty analysis is to show when lives are lost at sea. The type of events where most lives are lost will be the events where the largest beneficial changes in risk will occur when ships become unmanned. It can be expected that the regulations that are associated with the largest risk decrease are prime candidates to be challenged. The analysis will also address for which ship types and sizes the effect of reconsidering the regulations can be expected to be largest, i.e. those where the contribution of the risk of loss of life to the overall level of risk is largest.

As stated before, this paper will conservatively assume that the probability of an accident will not decrease due to the introduction of autonomous technology. Therefore, the casualty analysis will focus on the severity of accidents that occur. Subsequently, the discussion will focus on the regulations that are intended to limit the consequences of an accident.

The paper is organized as follows: in Section 2 the structure of the casualty analysis is described in detail. Thereafter, in Section 3 the results of the casualty analysis are presented. In Section 4 it will be discussed which regulations should first be reconsidered, based on the results of the casualty analysis. In Section 5 the conclusions are presented.

2. Casualty Analysis

The IHS SeaWeb® database is used for the analysis of casualty data. The casualty data consists of all serious accidents concerning cargo ships from 2000 to 2018. Serious accidents are defined by IHS Markit as those accidents where the vessel incurred significant damage and/or was withdrawn from service. Although the definition of serious events is ship-centred, 99% of all lives lost are allocated to serious accidents and, therefore, this study will focus only on the serious accidents.

The dataset has not been limited by vessel size, but ships built before 1980 are excluded. This limitation is adopted from the statistical analysis performed by *Eliopoulou et al. (2016)*. It is the intention to use ships in the analysis that are employed with similar shipbuilding technology. According to Eliopoulou et al. less radical changes in employed shipbuilding technology are observed after 1980, compared to before 1980. Although it would be best to compare future ships with technology that is used in present shipbuilding only, further limiting the number of recorded accidents will increase the uncertainty of the statistical analysis.

The analysed dataset only includes cargo ships, since these ships can become truly unmanned. For a ship to be truly unmanned, there cannot be any passengers on board as well. The ships are divided into five categories, using the StatCode 5 ship type coding system as is used by IHS Markit. The resulting five categories are General Cargo Ships, Bulk Carriers, Container Ships, Tankers and Other Cargo Ships. In earlier work (Vos & Hekkenberg, 2020) differences in the probability of losing life per ship type have been observed and it is expected that these can be explained by the difference average size per ship type. Therefore, the differences per ship type will be evaluated in the casualty analysis as well.

Furthermore, to support the analysis per ship type, the contribution of the ship size to the probability of losing lives will be evaluated as well. It is expected that the impact of removing the crew on the overall level of risk of a ship will be largest for smaller ships, has already been stated in the same earlier work *Vos and Hekkenberg (2020).* For larger ships, the value of the ship and cargo becomes the dominant contributor to risk. Also, it has been speculated that the probability of losing lives is larger for smaller ships. However, this speculation has thus far not been validated due to a lack of available casualty data. Following these two arguments, it is expected that the contribution of risk of losing life to the overall level of risk will be higher for smaller ships. Therefore, the impact of removing crew is expected to be largest for smaller ships.

The ship size will be represented by the ship's length. A number of important regulations, such as those for intact stability and damage stability, use length as a measure of ship size. As will be discussed in Section 4, these regulations are important for increasing shipping efficiency and, therefore, the same measure for ship size will be used in this paper.

The casualties in the IHS SeaWeb® database are subdivided in eight categories. The definitions of these categories as provided by IHS Markit are as follows:

- Collisions: Incident as a result of striking or being struck by another ship, regardless of whether under way, anchored or moored. This category includes collision with drilling rigs/platforms, regardless of whether in fixed position or in tow.
- Contact: Incident as a result of striking an external substance – but not another ship (see collision) or the sea bottom (see stranded) – except where the contact is only momentary, and the vessel does not come to a standstill.
- Fire/Explosion: Incident as a result of fire and/or explosion where it is the first event reported. It therefore follows that casualties including fires and/or explosions after collision, stranding etc., would be categorised under 'collision' or 'stranded' etc.
- Foundered: Ships which sank as a result of heavy weather, springing of leaks, breaking in two etc., but not as a consequence of any of the other categories listed.
- Hull/Machinery Damage: Hull/machinery damage or failure which is not attributable to any other category.
- Missing: After a reasonable period of time, no news having been received of a ship and its fate being therefore undetermined.
- Stranded: Incident as a result of the ship coming to a standstill on the sea bottom, sandbanks or seashore, etc., as well as entanglement on underwater wrecks.
- War-loss/Hostilities: Incidents causing loss of or damage to a ship as a result of a hostile act.

Most of the categories have a clear distinction between them. However, the line between hull/machinery damage and the other categories is not always clear. The starting point for categorizing an accident as hull/machinery damage is that hull or machinery failure needs to be the first reported event. However, most of the severe 'hull/machinery damage' accidents are accidents where the ship took water and subsequently foundered. The cause of the ship taking water is not always described in detail and 'took water' or 'developed list' are often given as initial event, while the accident is classified as 'hull/machinery damage'.

3. Results of casualty analysis

3.1 Results per casualty type

The complete set of casualty records covers 12090 accidents, as shown in Fig.1. The figure provides a global overview of the casualty types that cargo ships have been involved in. Most recorded accidents are 'hull/machinery damage', followed by 'collisions' and 'stranded'. The remaining casualty types occur significantly less often. Especially 'missing' and 'war-loss/hostilities' only occurred a few times between 2000 and 2018.

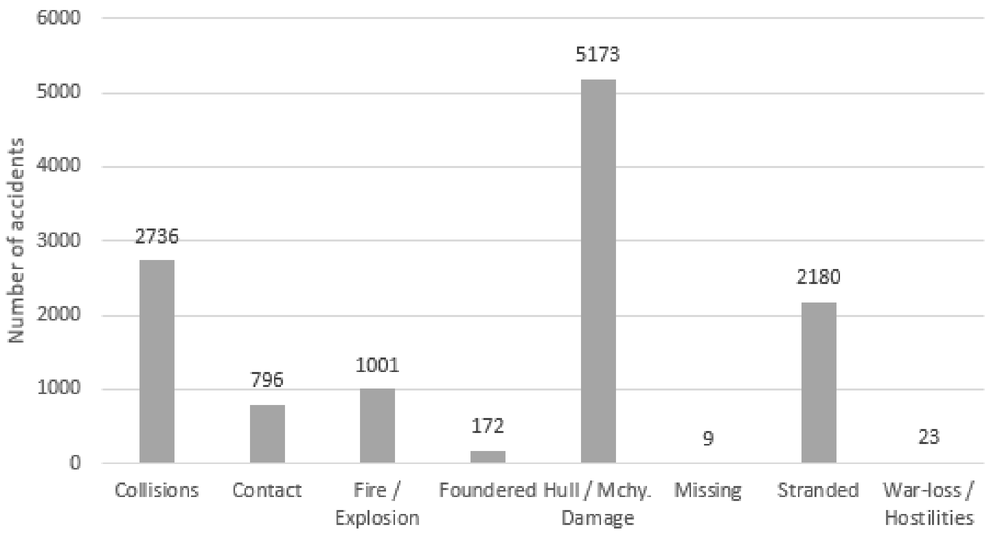


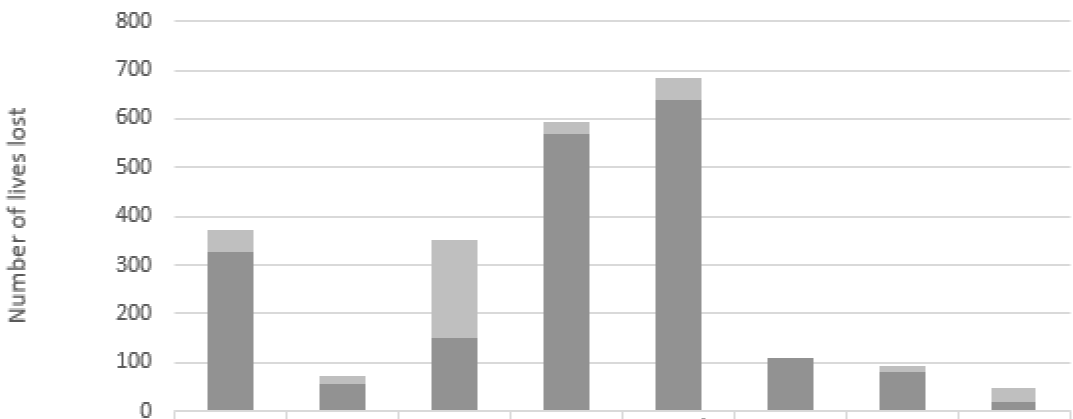
Fig.1: Number of accidents per category of casualty type.

In Fig.2 the total number of killed or missing people per category of casualty type are presented. In this figure a distinction has been made by lives lost on ships that survived the casualty and ships that were lost due to the casualty. This distinction shows that significantly more lives are lost on ships that do not survive the accidents. Exceptions are accidents concerning fire/explosion and war-loss/hostilities.

Fig.2 also shows that most lives are lost during accidents involving collision, fire/explosion, foundering and hull/machinery damage. As can be seen in figure 1, these are not necessarily the casualty types that occur most often. Especially the categories ‘foundered’ and ‘stranded’ stand out. Foundering occurs less often than other categories, but is still the second largest cause for loss of life. In contrast to ‘stranded’, which is the third most occurring casualty type, but does not lead to large numbers of lives lost.

The fact that a significant amount of lives are lost during foundering accidents, can be explained by the definition of the category ‘foundered’. For a ship to be foundered, it has to sink, often due to heavy weather. The circumstances of such an accident can drastically decrease the possibility of the crew to save themselves. Stranding accidents result in a ship loss far less often. Also, the ship loss is not necessarily due to sinking; the ship can also be declared as lost due to the damage it took during the stranding accident. Therefore, the circumstances during stranding accidents are more likely to allow the crew to bring themselves to safety.

Although Fig.2 does not show this explicitly, most of the lives lost allocated to ‘hull/machinery damage’ are associated with accidents where the ship subsequently foundered. This is of importance, because it shows that in general most lives are lost when the ship took water and subsequently sank. The exception to this statement are the fire/explosion accidents, where the fire and explosion events itself contribute to the risk of losing lives.



	Collisions	Contact	Fire / Explosion	Foundered	Hull / Mchy. Damage	Missing	Stranded	War-loss / Hostilities
Ship survived	45	15	202	23	43	0	12	28
Ship lost	327	57	151	570	639	108	80	19

Fig.2: Total number of killed or missing people per category of casualty type, subdivided into lives lost when the ship survived or when the ship was lost.

Fig.3 combines the total number of casualties and the total number of killed or missing people. The figure shows the average number of lives lost per accident per casualty type. The y-axis has been limited to 0.5 average number of lives lost per accident due to extreme differences between the categories. For each category the exact number is provided as well.

The average number of lives lost during ‘missing’ accidents stands out. However, it can be expected that if the fate of the ship is unknown and labelled as ‘missing’, that the crew went missing as well. Therefore, this number merely says something about the average crew size of ships that went missing.

War-loss/hostilities accidents are also associated with a high average number of lives lost per accident. However, for this category in particular, the cause of loss of life is not necessarily ship-related. As a result, not all accidents can be related to regulations that focus on the design of the ship. Because of this and the low number of accidents allocated to this category, this category is not explored further.

The average number of lives lost during foundering accidents is the third that stands out. This is in line with the remarks about Fig.2. Since only a low number of foundering accidents are recorded and a high number of lives are lost, it can be expected that the average number of lives lost per accident is high as well.

Besides the three categories that stand out, the highest average number of lives lost per accident can be found for fire/explosion accidents. This implies that the risk of loss of life is a relatively large contributor to the overall level of risk associated with this category.

Fig.3 also confirms the earlier remark about stranding accidents. The high number of accidents, but low number of lives lost results in a low average number of lives lost per accident. As a result, it is the least lethal of the considered categories.

As for 'collisions' and 'hull/machinery damage', these types of accidents occur often, but they are not always accompanied by loss of life. The high number of accidents that are serious for the ship, but are without fatalities, causes the average number of lives lost per accident to be lower. However, a significant number of these accidents still lead to the loss of a part of or the entire crew. As a result, the average number of lives lost per accident is still significantly higher than stranded accidents.

Contact accidents are similar to collision accidents, but they occur less often and are less severe. This can also be seen in the lower average number of lives lost per accident. This can mainly be explained by the fact that most contact accidents occur in or near a port and at low speed. Also, aid can be provided quickly by third parties in these areas.

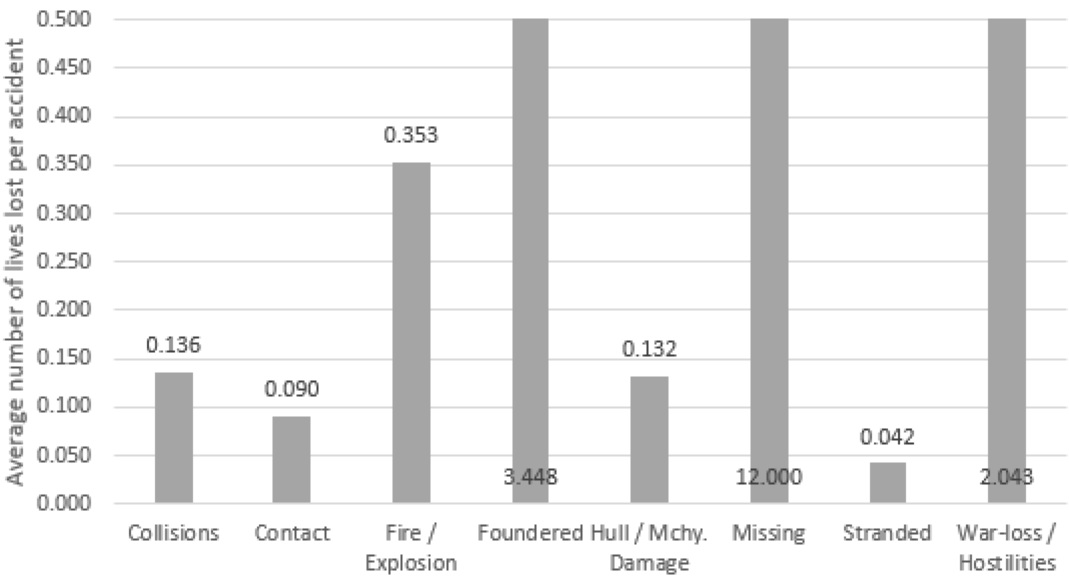


Fig.3: Average number of lives lost per accident per category of casualty type. Note that the vertical axis has been limited to 0.5 average number of lives lost per accident.

Concluding the remarks about Figs.2 and 3, most lives are lost when the entire ship is lost. A ship loss is most often the result of water ingress. Water ingress can be the result of collision, contact, heavy weather (e.g. foundering accidents) or any other breach of the watertight integrity (e.g. fire/explosion or hull/machinery damage). The second most important cause of loss of life are fire/explosion accidents.

3.2 Results per ship size

As mentioned in Section 2, it is expected that the probability of losing life is larger for smaller ships. In this section, the probability of losing life will be evaluated using ship length as a representation of the ship size.

First of all, Fig.4 shows the number of accidents recorded as a function of the length of the ship. The trend of the figure can mostly be explained by the number of ships that sailed around for each interval from 2000 to 2018. Most ships had a length between 80 and 120 ms. The number of ships decreases steadily when the ships become larger, with an exception for ships between 160 and 200 m. Another peak is present for this interval, which can be allocated to an increased number of bulk carriers able to access most smaller ports, so-called handymax ships.

The interval that stands out most is the interval from 40 to 80 m. The number of ships at risk for this length interval is close to the number of ships of the interval 80 to 120 m. However, for ships between 40 and 80 m significantly less accidents have been reported. This may suggest significant under-reporting of accidents for ships under 80 m.

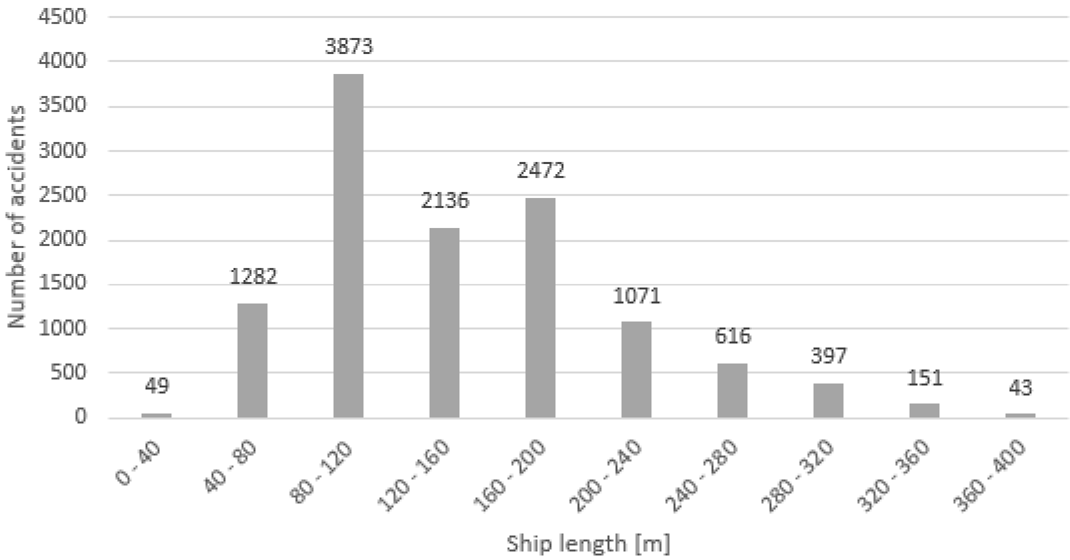


Fig.4: Number of accidents sorted by the length of the ship involved

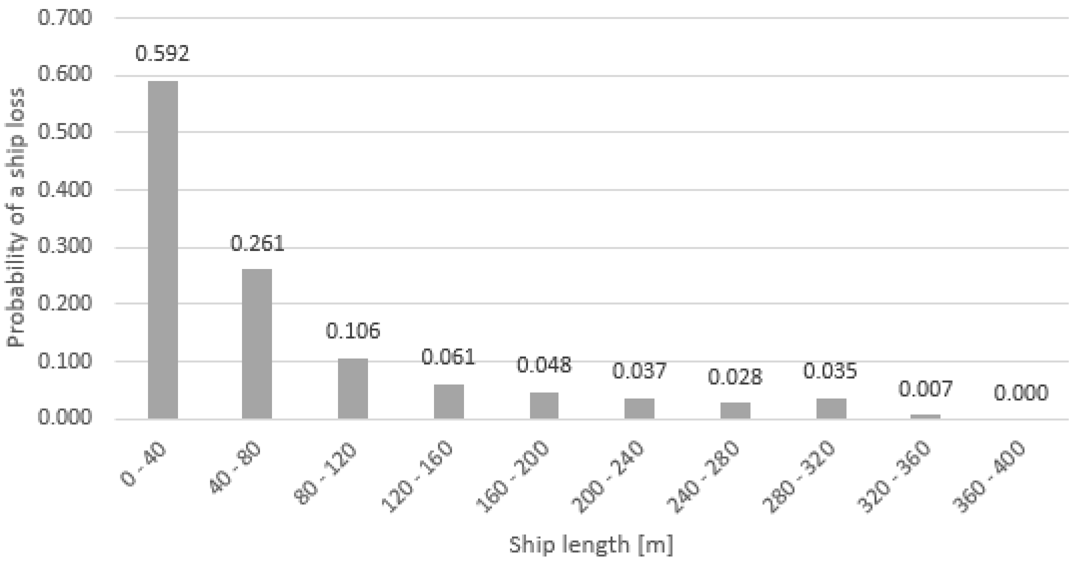


Fig.5: Probability on a ship loss sorted by length of the ship involved

In general, most accidents occur with ships of a length of under 200 m. Moreover, these are also the most severe accidents. Fig.5 shows the probability on a total ship loss when involved in an accident. It follows that the highest probabilities are associated with ships of under 200 meters in length and the probability increases for smaller ships.

Figs.6 and 7 respectively show the total number of lives lost per length interval and the average number of lives lost per accident per length interval. Fig.7 also shows that the highest average number of lives lost are associated with ships that have a length of 120 m and under. However, the average increases again for ships over 240 m.

The low number of accidents for the intervals above 200 m makes them sensitive for small changes in the data. For example, 22 of the 31 lives lost in the interval 320 to 360 m were on board a single ship of nearly 322 m. Would this ship have been just 2 m shorter, it is unlikely that the safety of the ship will be different. However, this would have made a significant difference in the numbers shown in the figures in this section. For the two intervals between 240 and 320 m, most lives lost can be linked to such individual large accidents as well, while for the interval from 200 to 240 m no such accident exist.

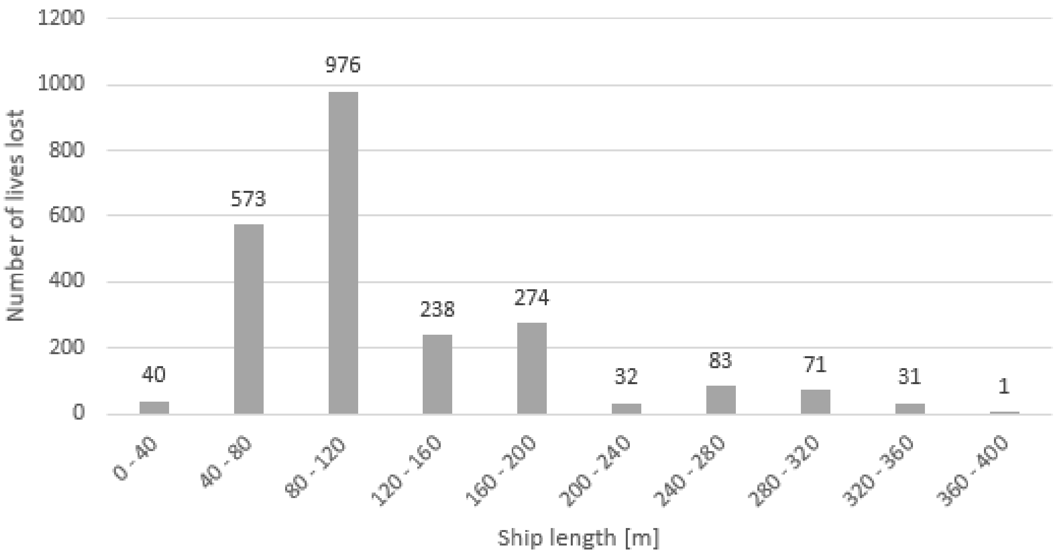


Fig.6: Total number of killed or missing people sorted by the length of the ship involved

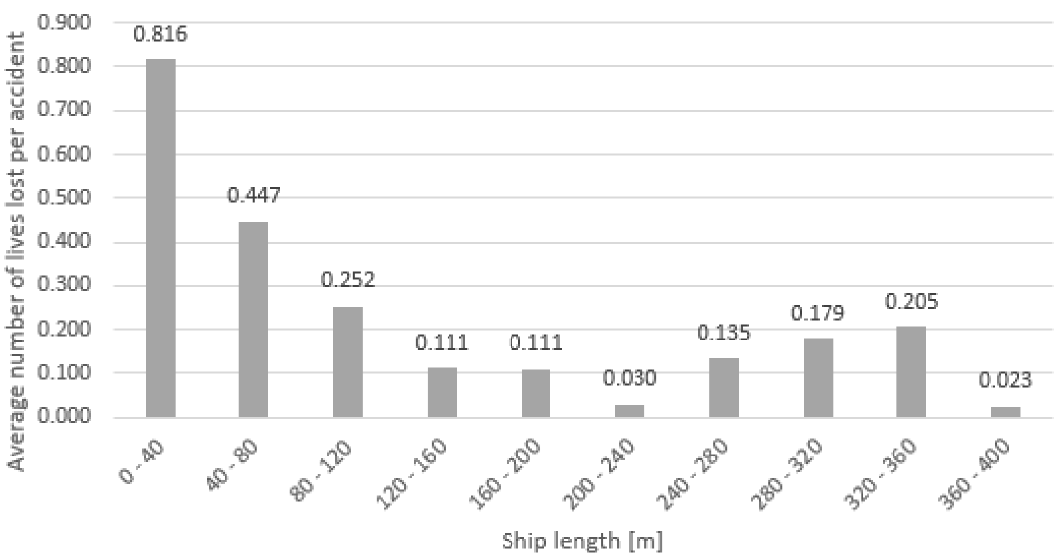


Fig.7: Average lives lost per accident sorted by the length of the ship involved

Concluding this section, in general most lives are lost on ships under 200 m. Also, this section confirms the expectation that the probability of losing life is higher for smaller ships. The highest average number of lives lost per accident can be found for ships under 120 m. Therefore, the risk of losing life will have the largest impact on the overall level of risk for ships well under 200 m. As a result, the reconsideration of regulations should focus on small ships, since for small ships the impact of removing crew will be largest.

3.3 Results per ship type

For the casualty analysis, five categories of ship types are used, as presented in Section 2. In this section, the differences per ship type will be presented. Per ship type, Table I shows the number of recorded accidents, the average length of the involved ships, the standard deviation of this length and the average lives lost per accident.

The ship types with the highest average number of lives lost per accident are general cargo ships. The average in this category is almost twice as high as in the other categories. As expected, this can mainly be explained by the fact that general cargo ships are the category with the smallest average vessel size. The standard deviation is small as well, indicating that around 95% of the ships are under 160 m. The average number of lives lost per accident for general cargo ships is in line with the average number of lives lost per accident for smaller ships, as can be seen in Fig.7.

Concerning the other ship types, most of them have about half the average number of lives lost per accident as general cargo ships, except for container ships. For these ship types the average length is higher and the length of the ships is more spread out. Therefore, the average number of lives lost for these ship types is also in line with the results of Fig.7.

As mentioned above, only container ships show a significantly smaller average number of lives lost per accident. This might be explained by the differences in design and operational profile compared to the other ships. These differences can lead to a higher reserve buoyancy for container ships, providing a safer environment for the crew on board container ships. This is supported by a lower probability of a ship loss for container ships compared to other ship types.

Table I: Casualty details per ship type

Ship Type	Number of accidents	Average length	Standard deviation	Average number of lives lost per accident
	[-]	[m]	[m]	[-]
General cargo ships	4569	102	30	0.297
Bulk carriers	2582	198	46	0.162
Containerships	1567	203	73	0.029
Tankers	2438	152	68	0.149
Other cargo ships	934	138	48	0.148
Total	12090	148	66	0.192

4. Discussion

Summarizing the results of the casualty analysis, the reconsideration of regulations should take the following into account. Most lives are lost on ships well under 200 m, with the highest average number of lives lost per accident associated with ships under 120 m. The most common ship type for the smaller ships is general cargo ships, which shows the highest average number of lives lost per accident compared to other ship types. Furthermore, first and foremost lives are lost during accidents where ingress of water is involved. The second largest number of lives are lost in events where fire or explosions are involved.

Following the results, this section will discuss in more detail which regulations are the prime candidates for reconsideration.

Regulations can be divided into two types of categories, those that intend to prevent a casualty and those that intend to reduce the consequences of the casualty. An example of the first type would be the definitions of a stand-on vessel and a give-way vessel to prevent a collision. An example of the second type would be the damage stability requirements to reduce the consequences of damage by reducing the probability on a total ship loss. It is not the intention that the reconsideration of regulations increases the number of accidents that occur. Among others, because this might also increase the risk for manned ships nearby the unmanned autonomous ship (e.g. an increased probability of a collision will have a negative effect on the safety level of nearby manned ships as well). In order to ensure that the safety level of third-party ships is not negatively affected, it is recommended to reconsider the regulations that intend to reduce the consequences of a casualty.

Furthermore, Section 3.1 already showed that water ingress (resulting in a ship loss) and fire or explosions are the most pressing causes of loss of life at sea. Also, following the introduction, the incentive for reconsideration of regulations is to increase the efficiency of autonomous ships. As a result, the regulations that are prime candidates for reconsideration are the regulations associated with water ingress and fire or explosion for which it is expected to increase the efficiency of autonomous ships.

It is expected that changing the regulations in place for reducing the consequences of a fire or explosion can reduce the building cost of the ship. Most of these regulations are associated with required building material, or the installation of fire suppressing measures, such as sprinklers.

To specifically reduce the consequences of water ingress due to damage, for instance as a result of collision or contact, damage stability regulations are applicable. An example of damage stability regulations is the required subdivision index, which has been evaluated in previous work.

For reducing the consequences of water ingress in general, multiple regulations are in place. Examples of these regulations are the International Code on Intact Stability and the Convention on Load Lines. Both of these regulations focus on ensuring sufficient reserve buoyancy. However, a downside of these regulations is a reduced loading capacity.

The regulations mentioned above often depend on the length of the ship, such as the required subdivision index or the required minimal freeboard height. As Section 3.2 showed, the risk of losing life depends on the length of the ship as well. Repeating the result from Section 3.2, the reconsideration of regulations should focus on smaller ships of (well) under 200 m, since the average number of lives lost per accident is highest for these ships.

5. Conclusion

The intention of this paper was to explore which regulations should first be reconsidered to keep an equivalent safety level but remove unnecessary restriction in design freedom. A casualty analysis showed that most lives are lost when the ship is lost due to water ingress or during accidents concerning fire/explosion. As a result, the prime candidates for reconsideration are regulations associated with reducing the consequences of a fire, explosion or water ingress.

The casualty analysis also showed that the risk of losing life depends on the length of the ship; the contribution of the risk of losing life to the total risk level will be higher for smaller ships. Therefore, the reconsideration of regulations should focus on smaller ships of (well) under 200 meter. Furthermore, general cargo ships are the ships with the highest risk of losing life, since general cargo ships are generally small ships.

Acknowledgments

The research leading to these results has been conducted within the NOVIMAR project (NOVel Iwt and MARitime transport concepts) and received funding from the European Union Horizon 2020 Program under grant agreement n° 723009.

References

BESER, F.; YILDIRIM, T. (2018), *COLREGS based path planning and bearing only obstacle avoidance for autonomous unmanned surface vehicles,* Procedia Computer Science 131, pp. 633-640

BIMCO & ICS (2015), *Manpower Report – The global supply and demand for seafarers in 2015 – Executive Summary,* http://www.ics-shipping.org/docs/default-source/resources/safety-security-and-operations/manpower-report-2015-executive-summary.pdf?sfvrsn=16

BUREAU VERITAS (2017), *Guidelines for Autonomous Shipping*, Bureau Veritas

DNV GL (2018), *Class Guideline – Autonomous and Remotely Operated Ships*, DNV GL

ELIOPOULOU, E.; PAPANIKOLAOU, A.; VOULGARELLIS, M. (2016), *Statistical analysis of ship accidents and review of safety level,* Safety Science 85, pp.282-292

FRIJTERS, T. (2017), *Future Ships,* Master Thesis, Delft University of Technology

HUANG, Y.; CHEN, L.; CHEN, P.; NEGENBORN, R.R.; GELDER, P.H.A.J.M. VAN (2020), *Ship collision avoidance methods: state-of-the-art,* Safety Science 121, pp.451-473

KONGSBERG (2017), *Final design of "Yara Birkeland" revealed – model commences testing at SINTEF Ocean,* Kongsberg Maritime, https://www.kongsberg.com/maritime/about-us/news-and-media/news-archive/2017/final-design-of-yara-birkeland-revealed--model-commences-testing-at-sintef/

LLOYD'S REGISTER; QINETIQ; UNIVERSITY OF SOUTHAMPTON (2017), *Global Marine Technology Trends 2030,* Lloyd's Register Group Ltd, QinetiQ and University of Southampton

LLOYD'S REGISTER (2017), *Code for Unmanned Marine Systems*, Lloyd's Register

MSC 99/WP.9 (2018), *Regulatory Scoping Exercise for the use of Maritime Autonomous Surface Ships (MASS) – Report of the Working Group*, IMO, 23 May 2018

MSC.1/Circ.1455 (2013), *Guidelines for the approval of alternatives and equivalents as provided for in various IMO instruments,* IMO London, 24 June 2013

MUNIN (2016), *Research in maritime autonomous systems – project results and technology potentials,* Final Report, http://www.unmanned-ship.org/munin/wp-content/uploads/2016/02/MUNIN-final-brochure.pdf

RAMOS, M.A.; UTNE, I.B.; MOSLEH, A. (2019), *Collision avoidance on maritime autonomous surface ships: Operators' tasks and human failure events*, Safety Science 116, pp.33-44

RØDSETH, Ø.; BURMEISTER, H.C. (2015), *D10.2: New Ship Designs for Autonomous Vessels,* MUNIN

ROLLS-ROYCE (2016), *Autonomous Ships: The next step*, Rolls-Royce plc

VARTDAL, B.J.; SKJONG, R.; ST.CLAIR, A.L. (2018), *Remote-Controlled and Autonomous Ships,* Position Paper, DNV GL

VOS, J. DE; HEKKENBERG, R.G. (2020), *Assessment of the Required Subdivision Index for Autonomous Ships based on Equivalent Safety,* Int. Seminar on Safety and Security of Autonomous Vessels (ISSAV) and European STAMP Workshop and Conf. (ESWC), Helsinki, pp.14-23

The Potential of Virtual Reality (VR) Tools and its Application in Conceptual Ship Design

Jose Jorge Garcia Agis, Ulstein International AS, Ulsteinvik/Norway, jose.jorge.agis@ulstein.com
Per Olaf Brett, Ulstein International AS, Ulsteinvik/Norway & Norwegian University of Science and Technology, Trondheim/Norway, per.olaf.brett@ulstein.com
Ali Ebrahimi, Ulstein International AS, Ulsteinvik/Norway & Norwegian University of Science and Technology, Trondheim/Norway, ali.ebrahimi@ulstein.com
Diogo Kramel, Norwegian University of Science and Technology, Trondheim/Norway, diogo.kramel@ntnu.no

Abstract

State-of-the-art VR tools can be used for simulating critical ship operations at the conceptual design stage to evaluate the likely implications of set design parameters and to understand how external factors such as environmental conditions impact the concept design solution. A comparison of different available free-source VR software used for such simulations is performed in this paper. Our study shows that some recent VR tools are unnecessarily complex and cumbersome to be used effectively in conceptual ship design processes. Sometimes and in certain user-case situations, the ship design industry needs to find a better balance between the precision and quality of display and the resource intensity to develop the right VR-based simulations. Very often they focus more on the visualization element - to give a WOW-effect- and less on aspects relating to the efficient use of the tools within acceptable time-to-market deadlines and overall resources consumption.

1. Introduction

In recent years, the impact of virtual reality (VR) has increased exponentially in many business areas including engineering, medicine, education, design, and training. The maritime industry is also influenced by such a digital revelation. Starting in academic environments as a way to generate knowledge, *Haddara and Xu (1998)*, and expanding lately to production, *Karpowicz and Simone (1987)*, training purposes, *Pawlowski (1996)*, rule development, *Glen and Galea (2001)*, conceptual design, *Chaves and Gaspar (2016)*, *Erikstad et al. (2015)*, and operational management, *Ludvigsen et al. (2016)*. Virtual prototypes can be further implemented in virtual reality environments to give a more authentic perspective of a ship design process. However, despite being considered as a mature industry more than 20 years ago, *Jayaram et al. (1997)*, the implementation of VR is still limited and we can see only a few application examples in ship production, *Fernández and Alonso (2015)*, and the shipping industry in general, *Li et al. (2019)*. Thus, there is still a huge potential to introduce, apply and benefit from the new digital technologies in the maritime industry.

The concept of virtual reality expands beyond three-dimensional (3D) visualization, including in-depth perception criteria involving engagement and manipulation with the scene, *Wann and Mon-Williams (1996)*. The perception of presence and the effect of user's experience are a unique feature of virtual reality, *Spencer et al. (2019)*, "allowing the user to step into the design", *Cassar et al. (2019)*. Virtual reality is often seen as a means for "efficient management of information", *Schiavon et al. (2019)* through visualization. It is an additional step in the communication staircase between the ship designer and the other stakeholders, going from verbal communication and mimics, to sketches, general arrangements and full 3D renderings. It opens up an avenue to present information to experts and non-experts in ship design that can be more easily understood. For example, the perception of vessel movement and accelerations the seakeeping capability of a vessel, for a given sea state condition. For someone lacking the tacit perception of movements and accelerations, it will be difficult to evaluate if a working deck is designed for safe operations at 3.0 m Hs (significant wave height) or not. Or to decide what is a comfortable width for the corridors onboard the vessel. In general, virtual prototyping enables designers and operators to simulate changes and alternatives during the design or operational phases before their real implementation. Hence, it brings the opportunity of reducing errors and increasing design and operating efficiency, *Keane et al. (2017)*, and improve the effectiveness of the overall conceptual design process by reducing the perception of uncertainty, *Garcia (2020)*. In addition to the benefits as a complementary design tool, VR based simulators are quickly becoming an indispensable part of modern professional training in the maritime industry.

Industry and academia, *Cassar et al. (2019), Jayaram et al. (1997), Spencer et al. (2019)*, identify three areas where virtual reality can contribute to the ship design field: (a) improving the communication among stakeholders and enhance collaboration; (b) visualize the design from multiple perspectives and get a better understanding of the vessel itself, especially concerning the production and operation of the vessel; and (c) validate the vessel design solution in its operating environment. VR is also a tool that could foster innovation. Yet, to make VR effective in ship design and during the shipbuilding process, it has to fulfil three characteristics, *Martin (2019)*: (i) it has to be integrated across the entire lifecycle of the project, from concept design to production and in operation throughout the lifetime of the vessel; (ii) it must be "simple and intuitive" for use; and (iii) it must have "seamless" integration with other design tools and information sources of the vessel design, production process and vessel life cycle.

Virtual reality technology has multiple potential applications within the ship design value chain, from its conceptualization to its operation and scrapping, including its production phase. Yet, it is feasible to integrate all them in a common platform? Or each activity will require its own, purposely customized solution? If each department or activity part of the ship life's cycle has a customized VR tool, according to *Martin (2019)*, the benefits of effectiveness will not be fully exploited. This is one of the principal challenges faced within the design community, the integration of specialized software. Is the output of Software A feasible as input for Software B? Can we use the hull lines from our computer fluid dynamics (CFD) tool into the computer-aided design (CAD-2D) design tool? The seamless integration of design work among the tools used in the ship design and production process is paramount to achieve higher effectiveness. Avoiding repetition work by having to implement the same design changes in multiple tools is essential for the effective integration of new tools in the design process.

As of today, most of the VR technology requires extensive knowledge of software engineering and visual graphics, which are not widespread skills in the ship design community. This results in steep and long learning curves that make it difficult to implement in the ship design processes, *Cassar et al. (2019)*. Alternatively, it might require the integration of an additional role in the design process, the VR designer, taking partly away the naval architect the drawing function from, and leaving him or her more as a project coordinator. Simple and intuitive VR tools will allow the integration of VR as part of a naval architect's daily activities.

To sum up, for VR technology to be fully implemented in ship design processes in the future, it must represent an attractive balance between costs and benefits. After all, the original objective of virtual reality in design was to provide significant cost and time savings, *Jayaram et al. (1997)*. To this respect, the computer-game market has contributed to reduce the cost of developing VR displays to a fraction of their initial and original cost, *Wann and Mon-Williams (1996)*.

Hence, this paper investigates how state-of-the-art VR technology applies to the development and improvement of conceptual ship designs, and assess the cost-benefit status of the technology for applicability in conceptual ship design processes.

2. Virtual Reality in a ship design firm

The overall objective of this research is to explore the use of virtual reality technology to improve and strengthen the conceptual ship design process and support the design and operation of smarter, safer and greener vessels. Yet, we recognise that there are other areas where a ship design firm could take advantage of these technologies, Fig.1, including downstream design activities (Basic and Detail engineering), marketing and sales, production planning, construction and life cycle services. Overall, we have summarized them in three applications: (a) a design tool to evaluate new designs and design features; (b) a communication tool for marketing and sales processes; and (c) a simulator for training of crew and passengers during and before operations. Fig.1 represents the different activities taking place in the life-cycle of a vessel, from conceptualization to scrapping. The activities in blue represent those carried out by Ulstein, while in red are those activities carried out by other third parties. In the lower part of Fig.1, we have included widgets of foreseeable applications of VR technology within a vessel's lifecycle.

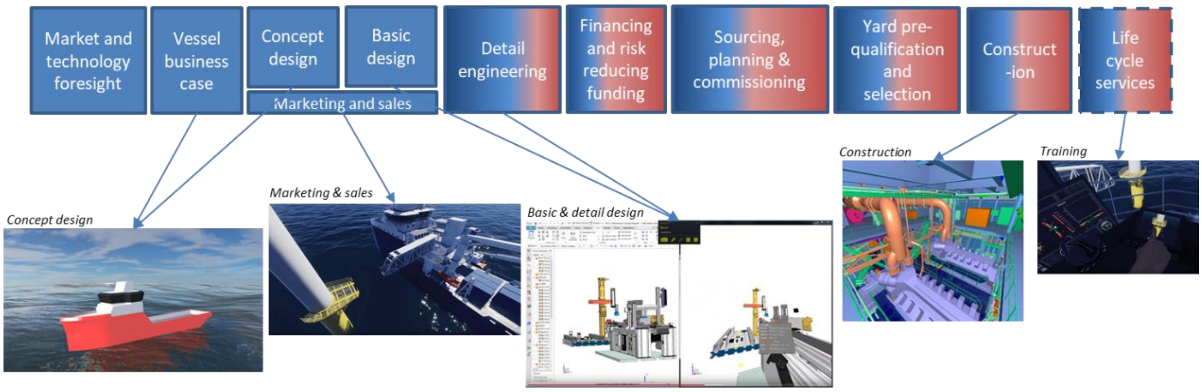


Fig.1: Integration of VR capabilities in Ulstein's existing service and activity portfolio

The VR design tool is considered as a supporting element during the Conceptual and Basic design, Detail engineering and Construction phase of a new building process. The tool could be used for both, getting a better understanding of the vessel in its operational context (the sea) and supporting the definition of the general arrangement of the vessel by giving a better perception of volumes, sizes and interaction of equipment and their functionality. One of the challenges ship design firms are facing today is that during the different design phases, they use different 3D modelling tools and methodologies, including e.g. Rhinoceros, Siemens NX and Cadmatic, respectively. Such 3D modelling tools are not always easily integrated and complemented.

As a design laboratory, the use of VR technology could encourage innovation by reducing uncertainty in the development of novel conceptual ship designs. VR environments supported by gaming software could provide a good platform to simulate the operation of new concept designs and therefore, design smarter, safer and greener vessels. Overall, the design laboratory could be used as a proof of concept, and an alternative to more expensive and resource-demanding processes.

VR technology can also support marketing and sales activities as an effective communication tool for complex objects like ships. It represents a better framework for communication with customers and other stakeholders that do not necessarily have a ship design background. The model developed here could be used as a platform to implement the equipment and solutions of 3rd party suppliers, such as furniture, topside equipment, fishing gear, etc. As a visualization tool, ship design companies use existing VR technologies for enhancing and strengthening their marketing and sales activities as complements to existing 3D renderings. Yet, it is recognised that the "wow" effect generated by the application of VR glasses is limited. In this respect, a video simulation of the VR environment may bring more possibilities. This shall facilitate the distribution of the "message" without requiring specific hardware (goggles). Besides, it could be used for more than one person at the time, and it could be easily integrated into the webpage of the company, or on a screen at a conference or seminar.

Fig.2 (left) is a snapshot of a tender operation being illustrated in Rhinoceros. Fig.2 (right) represents the operation of the gangway from the view of the gangway driver in VR. Both applications are being used for improving the design of service operation vessels (SOV). The definition and visualization of the operations give the users a better perspective of how the final vessel will look-like and perform and make changes if necessary. Changes at this stage are substantially quicker and cheaper to implement compared to later stages in the detailing phase or event during the construction phase.

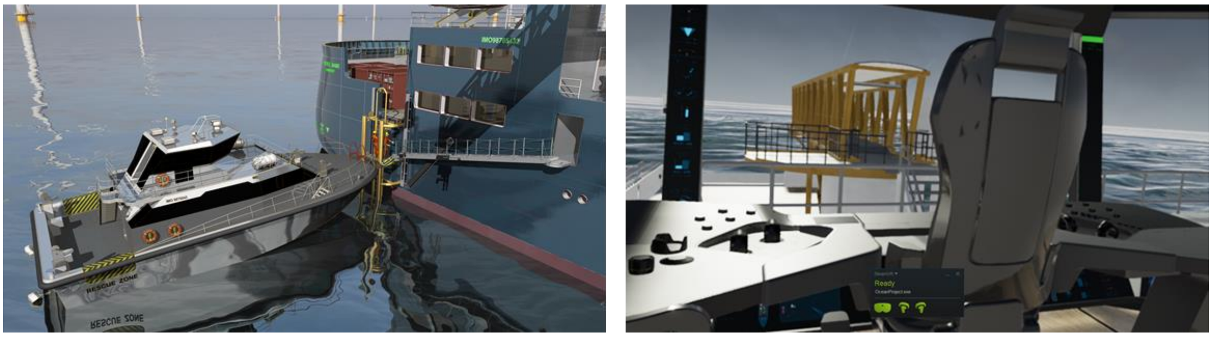


Fig.2: Tender operation renders Rhinoceros (left) and gangway operation VR (right)

Further, existing VR technologies could be used for training purposes with marine and hotel crew, but also with technical personnel and passengers. Marine and hotel crew could practice emergencies, evacuations, tendering operations, mooring and other important procedures relating to the operation of the vessel. The shipping companies could also, when and where relevant, exploit these technologies for training personnel and passengers in critical operations such as transfers of passengers between the mother vessel and tender boats, kayaks, zodiacs or helicopter landings, etc. Based on this training, cruise companies could decide whether a given operation is safe or not, when the externalities changes, like weather and sea conditions. Thus, the simulator could be an additional service that ship design companies offer to shipping companies for training purposes as well as an entertainment service provision for passenger onboard cruise vessels, for example.

Although the descriptions above might suggest the development of three alternative tools, they can be seen as alternative sub-applications of a single tool, which may be adapted to the needs of each application and the expertise and knowledge of the users. This is a pre-requisite for the effective integration of VR technology in the ship design process, as highlighted by *Martin (2019)*. However, with the purpose of this paper, the design tool application has been prioritized.

3. Tool exploration

The task of this project was to investigate currently-available software solutions for quickly exporting 3D ship models to a VR simulation environment, in which the response to different sea states can be computed and conceptual ship designs validated. The paradigm of this project manifests itself in the trade-off between realist physics and processing time. Despite the ever-growing evolution of processing power in modern computation, the complexity involving fluids interacting with rigid bodies remains a challenging task, that does not necessarily fit the needs of the conceptual design phase where response time is paramount. Nevertheless, to avoid complex fluid simulation and applying unnecessarily heavy mathematics, simpler models can be applied. Possible solutions include the use of closed-form expressions to estimate heave, pitch, and roll, *Jensen et al. (2004)*; or simply interpreting the hydrostatic forces of the vessel in a quasi-static sea state.

Thus, the challenge to tackle the problem at hand of this work was to find what solution could provide a good-enough answer and representation of the simulations without compromising the quality of the solution, and adapting to short response times. From the set of existing software solutions earlier identified, there was variation in terms of the approach, technology, methodology, licensing and their adaptability. Finding a better application was to understand how each of them fit different criteria that will ultimately deliver the expected performance requirements of this study. The criteria to evaluate the different solutions consisted of: (i) the ability to be integrated with the current 3D modelling tools, (ii) the ability to built up a library with different functionality, models and scenes; and the versatility to add new features, (iii) the ability to represent animation - the motion of objects over time without necessarily respect specific physics, (iv) the ability to simulate vessel movements and vessel operations, (v) its integration with a virtual reality environment and compatibility with Ulstein's existing hardware, and (vi) the need for domain knowledge and respective learning process curve.

The set of software under consideration are all alternatives found in the public domain between equivalences to computer fluid dynamics (CFD) software to purely cinematic animation software tools. Based on this criteria – processing speed vs data accuracy, the software under evaluation was subdivided into three groups, Fig.3: (a) modelling software with support for simulation through plug-ins (Rhinoceros), (b) game engines (Unreal Engine and Unity) and (c) animation software (Blender and SimLab Composer).

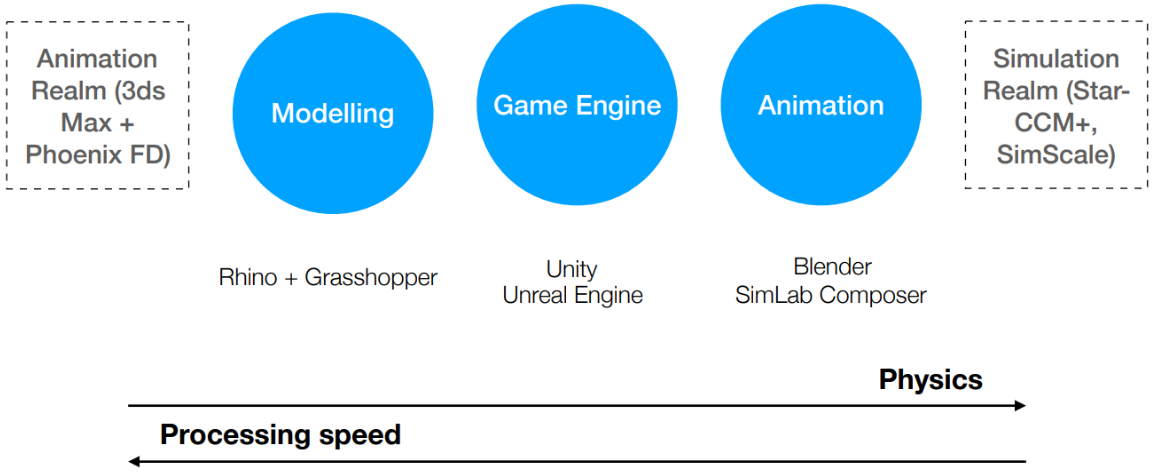


Fig.3: Overview of software evaluated concerning their capacities and capabilities

A promising solution, and supposedly simplistic, is to build a model completely integrated with the current 3D modelling software, Rhinoceros. Using plugins, such as Grasshopper - a graphical algorithm editor integrated with Rhinoceros, it is possible to develop scripts that will simulate the interaction among objects and systems of the vessel. However, the simplicity of this visual application limits its applicability. In the end, the convenience of having all the features available in one software might be increasing the complexity simply due to the need of using several plugins (for rendering, animation, virtual reality and physics).

Applications used for animating 3D models, such as Blender and SimLab Composer are interesting for integrating the whole process of modelling (and/or importing), animating and simulating, rendering and exporting to virtual reality, in the same platform. Not only that, but they offer extensive documentation and are relatively easy to use. However, the plugins available for simulating objects interacting with fluids are scarce, and those available, demand a considerable amount of time to be processed. Due to this fact, it was necessary to render the simulation before visualising. One option would be to apply DualPhysics, a plugin available for Rhinoceros that applies smooth particle hydrodynamics capable of representing physical iterations with high accuracy. However, the quality of the results comes with a price, as it takes several hours (up to 24h in some cases) to render a few seconds of a scene.

Game engines such as Unity or Unreal Engine do not include many features in its standard version. However, the users can benefit from an extensive online library and forum community, from which it is possible to add assets, scenes and algorithms required for the intended simulations. An alternative is to develop own scripts since game engines use generally common programming languages, such as JavaScript, C#, and C++. The use of game engines accelerates the processing time, allowing the user to simulate the scene in real-time and export it to platforms with much lower processing power, such as smartphones.

All the options evaluated above showed sufficient capabilities for rendering and exporting the simulation to several other platforms, including VR. Thus, the main differences that can be spotted are how realistic the fluids can be represented and how adaptable they are to changes. And in both aspects game engines are one step ahead. The fact that scripts can be programmed in C#, JavaScript or

Python, allows the user to include, in theory, any type of physics with different degrees of accuracy. Moreover, the learning curve and its implementation are significantly shorter than competing alternatives – based on our experiences. Fig.4 shows the result from our study and research on the subject with the time required for installation and implementation of Unity and Blender respectively. The data presented in Fig.4 represents the time spent by a person the first time he or she interacts with the software.

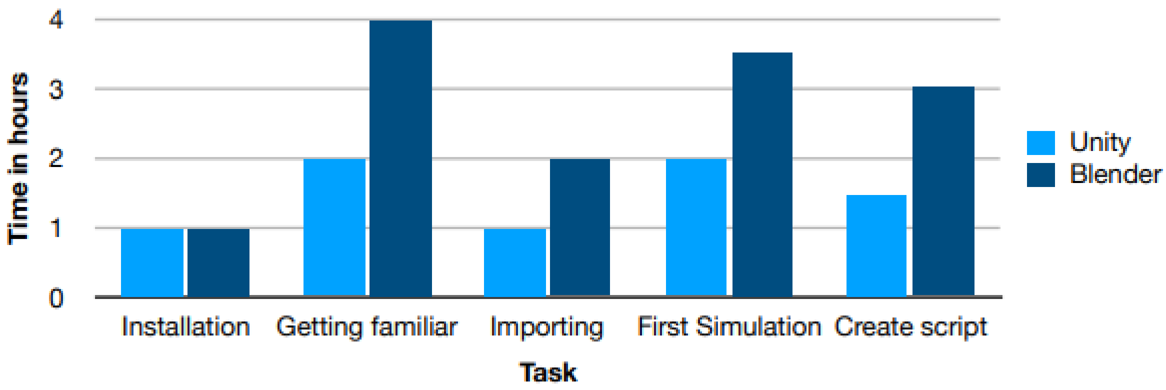


Fig. 4: Learning curve for Blender and Unity

After experimenting with the five solutions presented above, they were individually evaluated on a Likert scale of 1 (poor) to 5 (excellent) by an expert group of potential users of the applications. The results are summarised in Table I.

Table I: Evaluation of software solutions based on the selected criteria.

Attributes and Features	Rhino & Grasshopper	Blender	Unity/UE
Cost	5 (Free)	5 (Free)	4 ($75/month + assets)
Connection with modeling	5	3 (Plug-in)	3 (Plug-in)
Connection with Siemens NX	4	2	2
User-friendliness	2	3.5	5
Realistic Physics Engine	2.5	3	4
Libraries	3	4	5
Flexibility and adaptation	3	4	5
Virtual Reality compatibility	3.5	4	5
Potential process automation	5	4	4
Script language	2 (Grasshopper)	4 (Python)	5 (JS or C#)
Sum	**35**	**36.5**	**42**

Based on the results of the evaluation criteria, the ranking of VR technology solutions stands as follows: (1) Unity, (2) Blender and (3) Rhinoceros with Grasshopper. The same conclusion was found by the University College London, *Cassar et al. (2019)*, that identified Unity as the most attractive alternative for conceptual ship design applications.

Game engines have a more extensive application for handling fluids and are more extensively used for industrial applications in the simulation of ship operations. Because of those reasons, the following user-case study case was developed with Unity3D. A major limitation of software from the game industry such as Unity is the connection with other CAD tools, *Cassar et al. (2019)*. Changes implemented in CAD are not implemented directly to Unity and vice-versa. So, if we apply a change to the CAD model, this will have to be re-inserted in Unity, and all attributes (e.g. colours and textures) will have to be implemented again from scratch. Lack of compatibility between the applications is, therefore, reducing the efficiency of the application of VR simulations in the conceptual ship design process.

4. User-case studies

The case study carried out consisted of simulating new concept designs in different sea states to understand the effects and implications of weather conditions (wind, current and waves) in the operability of the vessels. Operability is in this case study evaluated based on the amount and probability of green water on deck.

The main challenge in this exercise was to balance the trade-off between the realistic motion of the vessel and good rendering without overwhelming computations. To overcome this dichotomy, we opted to develop and implement a simplified method based on Froude-Krylov forces for estimating the forces acting on the surface of the vessel. Sea waves are modelled as a flat mesh where the vertices move in only in the vertical axis following the results of sinusoidal waves. For simulations, the user may decide to operate with a regular wave sea or superimpose two or more regular waves to recreate an irregular sea. The size of the mesh size and the number of waves will influence the performance of the simulation model. This element is essential to the balance between the accuracy of results and resource allocation to each specific analysis.

The vessel was also represented as a mesh. Even though the entire vessel is imported as *NURBS* surfaces, it is only the vessel mesh that will physically interact with the sea mesh. The mesh and model of the vessel are further characterised with a displacement [mass] and a centre of gravity with an associated position in the XYZ coordinates. The vessel mesh is an important part of the simulation process. A coarse sea mesh is not recommended for high amplitudes since the boundaries between vessel and sea surface will not be well represented. For the vessel, it is important to have a symmetrical mesh. Based on our analyses, the mesh should have around 500 polygons for a reference vessel of 60 meters length. For complex geometries, special care needs to be taken to not increase unnecessarily the number of polygons. Simplifications are encouraged, using proxies to simplify the geometry if needed. This will reduce the computational power and time required for the simulations.

The interaction between objects happens in the interface between hull and sea meshes. For each time step, the submerged part of the hull mesh is determined. For each vertex submerged, the hydrostatic force is determined as Eq. (1). Where "h" is the water column below the surface, "g" the gravity, "S" the submerged surface of the hull, and "n" perpendicular vector to an acting point on the hull surface.

$$d\vec{F} = \rho g h d S n \qquad (1)$$

The pressure will act perpendicular to the surface, allowing the script to calculate the force acting over the hull. Viscous forces are accounted as damping factors. Working with time steps of fractions of seconds, the simulated system represents a realistic reflection of what it would be the real operation. Fig.5 (left) presents an extract of one simulation of the hull mesh of a platform supply vessel (PSV) operating in head seas with regular waves. The same vessel design is presented in Fig.5 (right), in this case operating with quarterly seas.

The model developed and tested as part of this study provides a simulation of the vessel in six degrees of freedom (6DOF) – heave, pitch, roll, sway, yaw and surge. The user-case studied as a basis to verify the results of our model was a PSV design. The results obtained from the model were compared with closed-form expressions for heave and roll provided by the online application Ship Motion Simulator (SMS) developed by NTNU, *Chaves and Gaspar (2016)*. Both results are displayed in the time domain. The comparison is made in terms of the motion amplitude, which was calculated by the difference between the maximum and minimum position of the vessel in that degree of freedom.

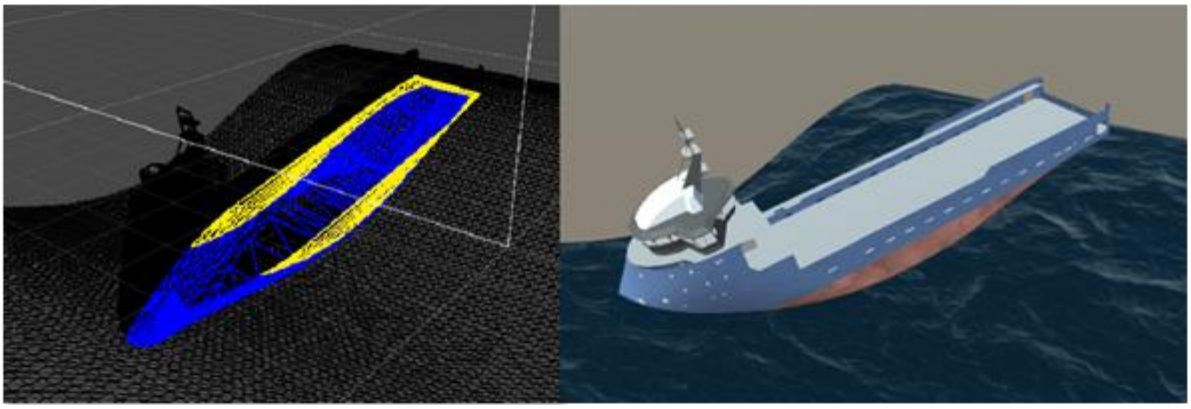


Fig. 5: Screenshots of simulation analysis for ULSTEIN PX121

Table II: Result comparison for operations in head seas (θ=0°)

Wave parameters	SMS heave	Unity heave	SMS pitch	Unity pitch
A = 1m, T = 11 s	0.8m	0.9m	1.8°	2°
A = 2m, T = 10s	1.4m	1.4m	4o	4.5°
A = 3m, T = 9s	1.9m	2.1m	7o	8.1°

Table III: Result comparison for operations in quarter seas (θ=45°)

Wave parameters	SMS heave	Unity heave	SMS roll	Unity roll
A = 1m, T = 11 s	0.97m	1.2m	1.8°	1.0°
A = 2m, T = 10s	1.94m	1.8m	3.5°	1.0°

The results show minor deviations between the results observed in Unity and those calculated based on closed-form expressions. The results for heave and pitch motions are within ±15%. In general, the results obtained in Unity overestimate both, the heave and pitch motions for head seas. The values for roll motion in head seas are, however, remarkable different. In general, our Unity model under-estimates roll motions for quarter seas.

Case study I: Platform supply vessel

The first case study consists of the simulation of a medium-sized platform supply vessel. The vessel design was verified by simulating operations in a variety of sea state conditions, including variations of wave height and wave direction, considering regular and irregular waves. Further, the vessel was also simulated during sailing at 5.0 and 10.0 kn speed, and during dynamic positioning (DP) at 0.0 kn speed. The objective of these simulations was to evaluate the probability of green water on deck. Green water on deck represents a risk for the operation and the integrity of the cargo in operations involving platform supply vessels and might stop the vessel from performing its intended work offshore.

Fig.6 shows an example of the simulations carried out for the platform supply vessel. The figure includes two screenshots of the simulations carried when the vessel was sailing at 5.0 kn speed on waves of 6 m height.

The objective of this simulation was to validate the design of the stern of the vessel with regards to the potential of flooding of water to the main deck when operating in dynamic positioning at stern seas. The design of the stern of Ulstein PX121 was revised after the first vessels were put into operation.

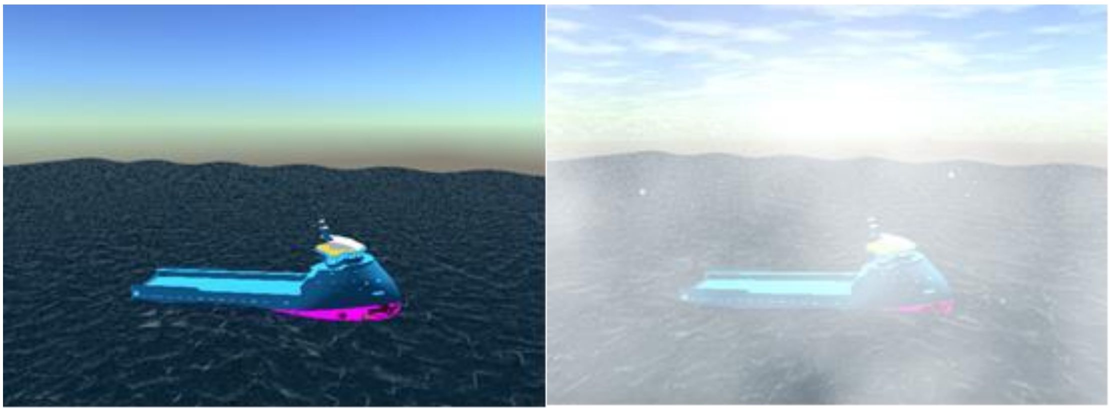


Fig.6: ULSTEIN PX121 sailing on head waves of 6 m height, sunny (left) and a snowstorm (right)

The latest vessels built of this design for operation in the North Sea were modified, including a raised stern to reduce the water intake when operating at stern seas. The comparison of the original stern design and the later modified version are presented in Fig.7. The purpose of the simulation was to evaluate the response of the vessel with the modified stern as compared to the original design.

Fig.7: ULSTEIN PX121 – original stern design (left) and modified stern design (right)

Case study II: Factory stern trawler

As part of the concept design development of a new factory stern trawler, we carried out simulations to evaluate the behaviour of the vessel during operations in different sea states. The design under consideration will operate under a quota regimen that limits the gross tonnage (GT) of the vessel to 1200 GT. The gross tonnage is a measure of the enclosed volume of the vessel. Hence, the ship designer is challenged with a trade-off between allocating volume for cargo, and volume for protection of deck operations. Reducing the height of the sidewalls, the designer frees up more volume for cargo-carrying capacity, improving the commercial performance of the vessel. However, this may compromise the safety of deck operations, where the crew has to handle the trawl. This is one of the reasons why it is not uncommon to see large amounts of green water on deck on these vessels.

To take a more informed decision, we decided to use this concept design as a case study for the under-development simulation tool. By simulating the operation of the vessel in different sea states, designers can verify the design and get real-time feedback of their decisions and modify accordingly the design of the vessel to better balance commercial performance and safety. This work had to be done in the early design phase before basic design work took place.

In this study, we have focused on wave heights above 6 m, as factory stern trawlers operate frequently in those conditions. The most critical operation simulated was with the vessel operating at speeds between 3 and 5 kn. This represents trawling speeds for whitefish species. Fig.8 is an extract of a simulation of the vessel design sailing with quarter sea waves of 8 m height. Similarly, Fig.9 (left) and right is extracts of simulations operating with 12 m waves on the side and heard seas respectively – extreme but yet realistic working conditions.

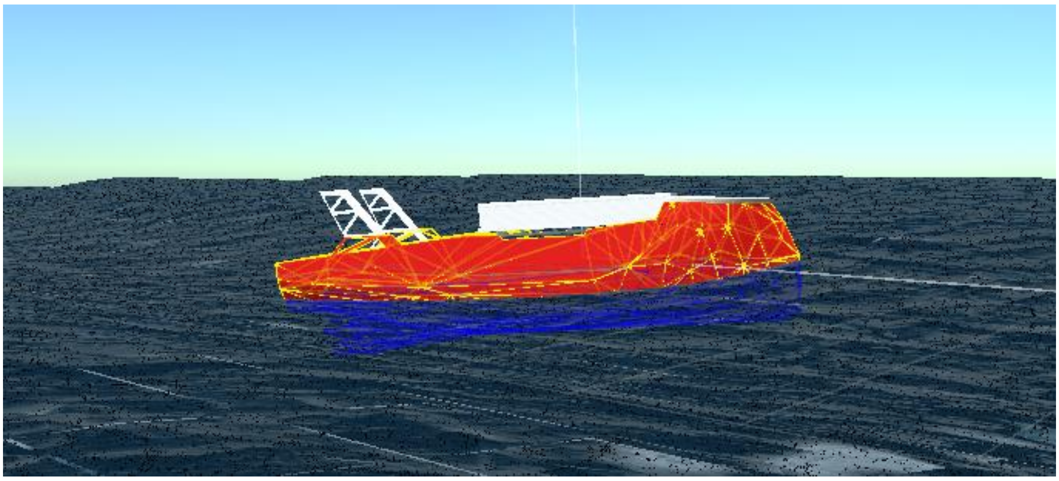


Fig.8: Ulstein trawler design sailing on quarter seas of 8 m height

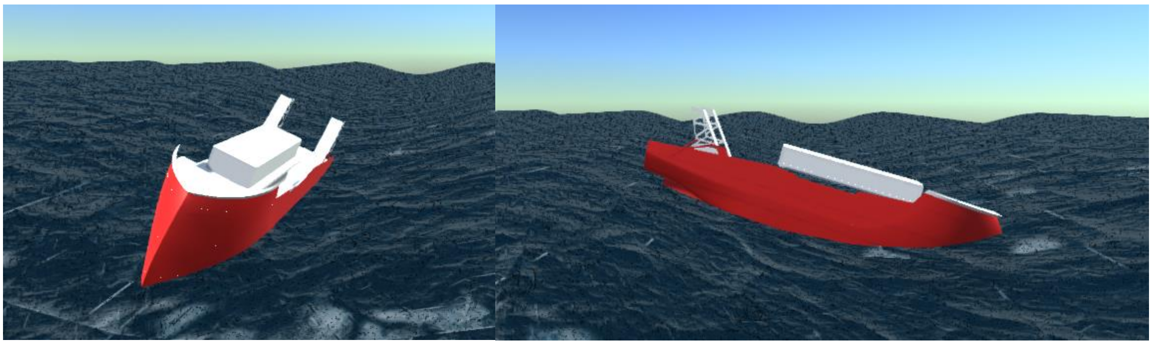


Fig.9: Ulstein trawler design sailing on side seas (left) and head seas (right) with 12 m height waves

5. Discussion

There are multiple applications for VR technology within the ship life cycle, including the design, production and operational phases. Within the ship design phase, we have recognised three applications: (a) a design tool to evaluate new designs and design features; (b) a communication tool for marketing and sales processes; and (c) a design tool to verify the vessel design and serve as communication tool towards ship producers. Overall, we foresee a large potential in utilizing virtual reality (VR) in the early phases of ship design development as a validation tool of concept ship design solutions. Although there are no off-the-shelf solutions to cater for the initial expectations, we have explored alternative software solutions and their capacity and capability to cater for the challenge proposed.

From the expansion of existing applications like the integration of Grasshopper into Rhinoceros to more advanced animation tools like Blender, we have identified a handful of VR tools that could easily and preferably be applied in ship design, without introducing complex and heavy mathematically based customised simulation tools. However, not all of them meet the critical needs of the conceptual design phase: good enough results and quick response times. Among the gaming software evaluated, Unity was ranked as the most attractive alternative, considering factors such as costs, flexibility, existing library, facility of use, etc. The accuracy of vessel movements and the perception of reality achieved by using Unity with a moderate dedication (less than 2 hours) are acceptable to a designer in the conceptual ship design phase. Yet, there is a limitation concerning the simulation of the roll motion. However, to use the tool for further design verification stages, it would require further research and development and more investments in time and resources to streamline the simulation tool. The same applies if the tool is used with customers and their application domain.

As of today, VR solutions stand alone as complimentary design tools for ship designers, and in most of the cases are not completely integrated into the ship design process. Virtual reality so far has found its most frequently used in the marketing and sales environment. The simulation of bridge operation has also represented an important business opportunity for developers of VR technologies. Ulstein sees the coupling of VR technologies and gaming software as an important addition to its existing and more traditional and main-stream design toolbox for enhancing design expedience (response to customers) while enhancing the design of better ships. The tool is not seen, initially, as a replacement of existing tools, but a compliment.

6. Conclusion

Based on the findings of this "Forstudie" (pre-study), Ulstein concludes that the use of VR in conceptual ship design procedures has value as a design validation tool. The reliability levels, time and resource consumption of applying game engines are considered acceptable to implement the simulation tool in future conceptual design processes, where the designer considers they can be of interest and relevance for the project. The importance of its application increases as higher is the uncertainty and novelty of the vessel design solution. Furthermore, Ulstein recognizes that for its further implementation and use in marketing and sales activities, additional resources need to be allocated to this activity, including some competencies that may not be available at the time within the company.

This paper shows how such applications can help designers in the early ship design phase to make better compromises on main ship dimensions, arrangements and functionality to enhance the lifelong operability in different commercial and operational future settings. The advantages of VR technology and tools-application in sales and marketing of new ship designs are also explored and argued in this paper. Even though authors like *Cassar et al. (2019)* suggest that VR helps in the early stages of vessel general arrangement, we see also its applicability towards later stages of the design cycle. Even expanding into the operation of the vessel as a training tool and for further marketing of the vessel.

This work has revealed that very often the application of VR tools is complex, time-consuming and user-case-limited, resulting in an expensive "nice-to-have" feature. This work concludes that many of the early versions of VR technology and applications are immature and not really fit for purpose. Their benefit is not retained throughout the vessel concept design phase outside the initial WOW-effect. The maritime industry needs to identify, experiment and exploit further benefits from applying this VR technology and its tools before it can become a useful complementary design means and become fully integrated into the commercial ship design industry.

This paper concludes that there still exists a puzzle as to how to build an effective feed-forward and feedback loop relating to VR tools applied in conceptual ship design processes. This paper questions whether we have set out on a wrong foot in the development of appropriate VR tools for conceptual ship design – so far the technology and tools identified and tested are still too complex, expensive, cumbersome to use…

Acknowledgements

This research has been partially funded by Innovation Norway under a forprosjekt scheme.

References

CASSAR, C.; SIMPSON, R.J.; BRADBEER, N.; THOMAS, G. (2019), *Integrating Virtual Reality Software into the Early Stages of Ship Design*, ICCAS Conf., Rotterdam, pp.24-26

CHAVES, O.; GASPAR, H.M. (2016), *A Web Based Real-Time 3D Simulator for Ship Design Virtual Prototype and Motion Prediction*, COMPIT Conf., Lecce, pp.410-419

ERIKSTAD, S.O.; GRIMSTAD, A.; JOHNSEN, T.; BORGEN, H. (2015), *VISTA (Virtual Sea Trial by Simulating Complex Marine Operations): Assessing Vessel Operability at the Design Stage*, IMDC Conf., Tokyo, pp.107-123.

FERNÁNDEZ, R.P.; ALONSO, V. (2015), *Advances in Engineering Software Virtual Reality in a Shipbuilding Environment*, Advances in Engineering Software 81, pp.30-40

GARCIA, J.J. (2020), *Effectiveness in Decision-Making in Ship Design under Uncertainty*, Norwegian University of Science and Technology (NTNU)

GLEN, I.F.; GALEA, E.R. (2001), *Ship Evacuation Simulation: Challenges and Solutions*, SNAME Trans. 109, pp.121-139

HADDARA, M.R.; XU, J.S. (1998), *On the Identification of Ship Coupled Heave-Pitch Motions Using Neural Networks*, Ocean Engineering 26(5), pp.381-400

JAYARAM, S.; CONNACHER, H.I.; LYONS, K.W. (1997), *Virtual Assembly Using Virtual Reality Techniques*, CAD Computer Aided Design 29(8), pp.575-84

JENSEN, J.J.; MANSOUR, A.E.; OLSEN, A.S. (2004), *Estimation of Ship Motions Using Closed-Form Expressions*, Ocean Engineering 31, pp.61-85

KARPOWICZ, A.S.; SIMONE, V. (1987), *An Application of Computer Simulation Methods in Ship Production Process*, Computers in Industry 9(1), pp.37-51

KEANE, A.; BRETT, P.O.; EBRAHIMI, A.; GASPAR, H.M.; GARCIA, J.J. (2017), *Preparing for a Digital Future - Experiences and Implications from a Maritime Domain Perspective*, COMPIT Conf., Cardiff

LI, X.; ROH, M.I.; HAM, S.H. (2019), *A Collaborative Simulation in Shipbuilding and the Offshore Installation Based on the Integration of the Dynamic Analysis, Virtual Reality, and Control Devices*, Int. J. Naval Architecture and Ocean Engineering 11(2), pp.699-722

LUDVIGSEN, .B.; JAMT, L.K.; HUSTELI, N.; SMOGELI, Ø. (2016), *Digital Twins for Design*, COMPIT Conf., Lecce, pp.448-456

MARTIN, J.S. (2019), *Virtual Reality: Tool or Toy for Shipbuilding?*, ICCAS Conf., Rotterdam

PAWLOWSKI, J.S. (1996), *Hydrodynamic Modelling for Ship Manoeuvring Simulation*, Int. Conf. Marine Simulation and Manoeuvrability, Copenhagen, pp.625-640

SCHIAVON, M.; COSUTTA, A.; AMBROSIO, L.; KEBER, M.; ZINI, A.; JEZ, M. (2019), *Virtual Reality in Shipbuilding: Three Use Cases in a Cruise Ship Design Process*, ICCAS Conf., Rotterdam

SPENCER, R.; BYRNE, J.; HOUGHTON, P. (2019), *The Future of Ship Design: Collaboration in Virtual Reality*, COMPIT Conf, Tullamore

WANN, J.; MON-WILLIAMS, M. (1996), *What Does Virtual Reality NEED?: Human Factors Issues in the Design of Three-Dimensional Computer Environments*, Int. J. Human Computer Studies 44(6), pp.829-847

Prediction of Vessel Domain and Safe Areas for Navigation Based on Machine Learning

Armin Pobitzer, SINTEF Ålesund AS, Armin.Pobitzer@sintef.no
Jon Schonhovd Dæhlen, SINTEF Ocean AS, jon.daehlen@sintef.no
Svein Peder Berge, SINTEF Ocean AS, Svein.Berge@sintef.no
Severin Sadjina, SINTEF Ålesund AS, Severin.Sadjina@sintef.no

Abstract

As autonomy is starting to get foothold in the maritime sector, safety will be an important aspect for all maritime stakeholders. In particular, the insurance sector needs to establish risk management for the rapid development of technical solutions as the navigation task will be replaced with computers and sensors. The PROXIMA project aims at utilizing data from on-board sensors on ships in a more efficient manner to get a better situational awareness of marine traffic, as well as obstacles on the vessel domain. The first stage of the project has focused on sensor fusion of radar data with a video stream, to extract relevant features such as ships, islets and drifting debris, and estimating its size and past track. This paper presents a method utilizing machine learning to predict future tracks of vessels that may encounter the domain of the vessel. An Automatic Identification System (AIS) data base forms the basis for the training data set, and a novel approach is presented for data selection and feature engineering. The resulting prediction model can be used to determine a zone (vessel domain) where the observed vessel is likely to be in the next minutes. Five different use cases have been used to study the robustness of the proposed method.

1. Introduction

With larger vessels and around 90% of all trade happening by sea, maritime safety is becoming increasingly important. Humans are, for most parts, in charge of monitoring and decision making, but this introduces high costs and the potential for errors. The contemporary increasing introduction of autonomy in the maritime segment requires on-board control systems to become situational aware making decisions in a maritime traffic pattern. In particular, it needs to predict the trajectory of nearby traffic to find which other vessels should influence its decisions, i.e. which traffic will reach the vessels domain. Further, to meet demands related to cyber-security, operation in an "off-line" state, not receiving other information than on-board sensors can provide, becomes vital.

As of December 31th 2004, The International Maritime Organization's has made transmitting AIS data compulsory for all passenger vessels, all vessels of more than 500 gross tonnage and all vessels of more than 300 gross tonnage on international voyages are required to use an AIS sender at all times, *IMO (2001, 2015)*. All AIS units transmits three different categories of messages:

- Static (MMSI, IMO number, AIS antenna location, etc.)
- Voyage related (Destination, planned track, draught, etc.)
- Dynamic (Current position, current speed, etc.)

While the voyage-related and static messages are sent with 6-minute intervals, the dynamic messages are with an interval of 3 minutes and 2 seconds, depending on the state of the vessel. Thus, AIS provides plentiful near real-time information on vessel movement and has unlocked great potential for surveillance, route planning and optimization, anomaly detection, and traffic pattern analyses. But the huge amount of data also requires automation to turn it into actionable insights.

AIS-data incorporates information suited to stitch together an estimate of the intentions and future track of vessels, e.g. using the destination and planned track fields. However, AIS may be considered unreliable as it can be disabled by jamming, the transmitted data from a sender may, intentional or not, be incorrect, or a sender can be set up to imitate a vessel.

On-board sensors such as radar and video cameras can provide an overview of objects close to the ship, as well as their tracks and estimated sizes. The accuracy and level of detail available from such sensors may vary depending on type, position and number installed, however in this paper its assumed that the current geographic position of a vessel along with is speed, heading, estimated length and breadth, and a past track consisting of waypoints are available.

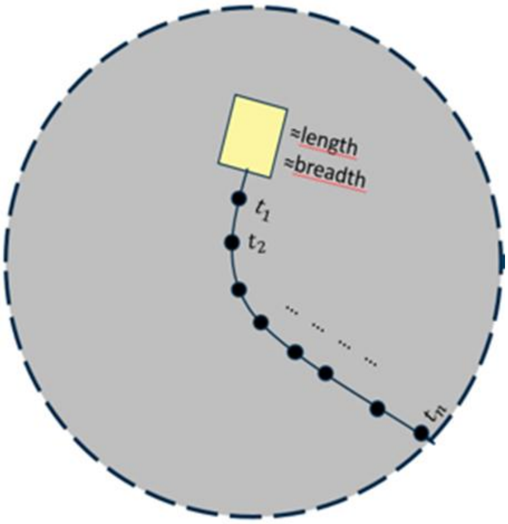


Fig.1: Illustration of an observation from vessel on-board sensors

Further, the behaviour of the observed vessels (traffic) is likely to be dependent on the vessel observing (reference vessel), considering the traffic separation scheme, as well as its own size relative to other near-by vessels. This information seems highly likely to be relevant for a prediction algorithm. Based on this, this paper proposes to use machine learning techniques to train an algorithm from historic AIS-data, using only data from on-board sensors, to be able to predict the short-term future path of vessels on the reference vessel domain.

2. Literature survey

2.1. Pre-processing of AIS data

AIS data exhibits large variability and volume and thus demanding to process. Highly irregular trajectory segments make it challenging to extract information automatically. Some common issues include, *Mao et al. (2018)*:

- Variable message frequency (gaps in trajectories and missing information)
- Redundancy (AIS messages received multiple times)
- Wrong timestamps (from multiple receivers)
- Vessels can erroneously share the same identifier
- Some information is entered manually (voyage information) and is often unreliable or missing
- Wrong values (often applies to Speed Over Ground (SOG) and Course Over Ground (COG))

Further analyses and prediction tasks must handle these issues and turn raw AIS data into more digestible representations. Broadly speaking, research groups have used one of the following approaches for AIS data pre-processing:

- Grid-based methods, e.g. *Lei et al. (2016), Sun and Zhou (2017)*
- Approaches based on entire trajectories, e.g. (Mao et al., 2018), however these may not leverage the information contained in their segments optimally (Lee et al., 2007; Petrou et al., 2020b; Tampakis et al., 2019)

- Point-based approaches and approaches based on sub-trajectories, e.g. (Pallotta et al., 2013b; Petrou et al., 2020b).

Studies on using big-data frameworks for storing and handling large amounts of AIS-data to be used in traffic analysis and prediction has been conducted, *Petrou et al. (2020b), Wijaya and Nakamura (2013b), Zhu (2011).*

Clustering is often used to identify and categorize vessel movements. Algorithms such as TRACLUS, *Lee et al. (2007)*, is designed to partition a trajectory into smaller segments, and group segments based on similarity. Clustering is also often used to detect outliers and noise. *Liu et al. (2014)* divides the trajectories of the AIS-data into "stopping" and "moving" sub-trajectories, depending on of the vessel speed is above a given threshold, reasoning that a low vessel speed indicates areas such as harbours, thus being able to treat data in these areas differently when analysing collision risk. *Qi and Zheng (2016*) clusters trajectories into distinct classes, then using Support Vector Machine (SVM) to predict and classify new trajectories. Gives a statistical measurement of the likelihood of a new predicted trajectory belonging to a certain class is advantageous in safety critical systems. *Kim and Jeong (2015)* applies SVM to find a reference route through a harbour, the authors suggesting this to be used to detect abnormal behaviour.

Several studies focus on detecting abnormal behavior, such as raising an alarm in a Vessel Traffic Control Center if a vessel deviates from an anticipated trajectory. Defining normal behavior is closely related to predicting future behavior, thus methods may be applied for both cases, *Laxhammar (2011).* Such anomaly detection methods can be used to disregard (sub)trajectory outliers according to some similarity criterion. In *Hagen (2018)* they combine historical AIS data with data from SafeSeaNet Norway. The project is also described in *Engen (2019),* and the purpose is both anomaly detection, and to estimate time of arrival. *Pallotta et al. (2013b)* presents the TREAD algorithm, designed to monitor AIS data for a given region (bounding box). It clusters the position messages instead of the trajectories, forming entry and exit-points where traffic is statistically likely to enter and leave the bounding box, and common waypoints within. Abnormal behaviour can then be detected based on deviations from these points. *Tu et al. (2020b)* specifically identify and exclude self-intersections and sharp turns from their training data. AIS-data often contains trajectory segments that are irrelevant or undesired for the task at hand: Wavering, U-turns, self-intersections, collision avoidances, interactions with other vessels, and movement influenced by traffic regulations may all be considered outliers in this context and should, to some extent, be excluded from model training for prediction tasks.

Sheng and Yin (2018b) select only the most representative points along trajectories (using cut-offs for speed, heading, and segment length) claiming that this helps with the extraction of information in high-density regions. *Tu et al. (2020b)* create uniform samples using a sliding window approach and normalize GPS positions (rotate and center) to avoid overfitting for their trajectory prediction task. *Chen et al. (2018)* apply the Least-squares Cubic Spline Curves Approximation to AIS data tracks to represent a set of tracks as uniform cubic B-spline curve. *Zhao and Shi (2019b)* presents an algorithm for trajectory similarity measurement based on Dynamic Time Warping (DTW), to cope with different time between the AIS messages without resampling.

2.2 Trajectory Prediction in various domains

Active research areas for short-term trajectory prediction that seem most closely related to the maritime problem statement are road traffic, human movement, and aviation.

Road traffic and human movement prediction are interesting to consider for their focus on the interactions between different agents. Intent, social interactions, and large-scale behavior often must be modelled, *Pei et al. (2019b), Ridel et al. (2018). Pei et al. (2019b)* use Long Short-Term Memory Recurrent Neural Networks (LSTM) to learn individual movement patterns connected by a "Social Affinity Map" pooling layer which is shared across the agents and learned jointly to model complex social interactions. *Xue et al. (2019)* use an encoder-decoder setup to encode pedestrian and vehicle trajectories, predict, and decode into a pedestrian trajectory prediction. Compared to vessel trajectory prediction, road traffic prediction intervals are much shorter (often only a few seconds) and road agents are typically moving according to constraints (roads, lanes), *Altché and de La Fortelle (2017)*. Human movement, in addition, exhibits abrupt changes.

There is also a lot of recent work in the aviation domain to predict trajectories, Ayhan and Samet (2016a), Liu and Hansen (2018b), Petrou et al. (2020b), Wang et al. (2017), Yang et al. (2019), Zhao et al. (2019). Some work focuses on airplanes on the ground, but typically this is a 4D problem. Weather conditions must often be included, *Liu and Hansen (2018b)* introducing additional uncertainty. Several studies discretize the space into a grid of cubes. Estimated time of arrival is often the main concern (for scheduling around airports) as is long-term prediction over the journey.

Table I: Methods used for trajectory predictions across different domains. Please note that the total count of methods is not representative for the whole field of research as only a selected number of publications has been chosen outside the maritime domain.

Model	Maritime	Aviation	Road	Human	Total
LSTM	(Gao et al., 2018b; Li et al., 2019a; D. D. Nguyen et al., 2018; Tang et al., 2019a; Zhao & Shi, 2019a)	(Liu & Hansen, 2018a; Yang et al., 2019; Zhao et al., 2019)	(Altche & De La Fortelle, 2018; Xue et al., 2019)	(Pei et al., 2019a)	11
DBScan	(Cazzanti & Pallotta, 2015; Daranda, 2016; Engen, 2019; Li et al., 2019a; Pallotta et al., 2013a; Sheng & Yin, 2018a; Zhao & Shi, 2019a)	(Wang & Delahaye, 2017)			7
Neighbor-based	(Dalsnes et al., 2018; Hexeberg, 2017a; Petrou et al., 2020a; Wijaya & Nakamura, 2013a)	(Petrou et al., 2020a)			5
NNs	(Bomberger et al., 2006; Borkowski, 2017a; Laxhammar, 2011; Mao et al., 2018; Tu et al., 2020a; Valsamis et al., 2017a)	(Wang & Delahaye, 2017)			7
Gaussian Mixture Model	(Dalsnes et al., 2018)	(Liu & Hansen, 2018a)			2
Gaussian Process Regression (GPR)	(Patterson et al., 2019; Rong et al., 2019; Tu et al., 2018)				3
Markov Chain	(Liu et al., 2019b; Pallotta et al., 2014)	(Ayhan & Samet, 2016b)			3
Kernel Density Estimator	(Pallotta et al., 2013a; Ristic et al., 2008)				2
Linear Regression	(Kim & Jeong, 2015; Qi & Zheng, 2016; Valsamis et al., 2017a)				3
Random Forest	(Valsamis et al., 2017a)				1

2.3 Short-term Vessel Trajectory Prediction in the Maritime Domain

Short-term prediction of vessel trajectories is a challenging task. Normality is hard to define because vessels are relatively unconstrained in their movements with no "topological support". Typical parabolic and slow trajectories also make prediction methods from other domains—road vehicles and human movement, for example—hard to adapt.

Trajectory prediction approaches may be physics-based, data-driven, or hybrid approaches. Physics-based approaches are based on mathematical equations directly describing the involved physics (albeit approximatively). Their major drawbacks for the purpose of short-term trajectory prediction are that it is hard to find the right initializations and model parameters and that they often need (close to) ideal conditions. Data-based approaches, especially machine learning, can be very powerful but require enough data of enough quality and a capable model fitting the situation. Hybrid methods may couple models from the two first approaches, for example a curvilinear model (physics-based) with Kalman filtering, *Tu et al. (2018).*

Hexeberg (2017b) predicts future trajectory positions for a vessel in a pure data-driven manner by retrieving AIS-messages for vessel that has been in a geographic proximity, with a similar course (nearest neighbours). Two strategies are then presented:

- Single Point Neighbour Search Method: A future position is estimated using the median or mean of the courses of the nearest neighbours, and the process is repeated for the predicted position until enough position for a predicted trajectory is obtained.
- Neighbour Course Distribution Method: A set of courses are randomly drawn from the nearest neighbours, and a new predicted position is calculated for each of them. Further, for each of the new positions, its nearest neighbours are found, and new predicted positions found based on their courses. The process is repeated, giving a tree structure of possible outcomes, claimed to cope with crossing ship lanes and uncertainty quantification. *Dalsnes et al. (2018)* further develops the method to use a Gaussian Mixture Model (GMM) for predicting new positions.

Recent years has seen an increasing popularity of machine learning methods, especially deep learning approaches, *Borkowski (2017b), Gao et al. (2018a), Li et al. (2019b), Nguyen et al. (2018), Tang et al. (2019b), Tu et al. (2020b*). One caveat when comparing the efficacy of methods is that hardly any two studies focus on the same geographical region, *Graser et al. (2019).*

- GPR and GMM methods enjoy a good track record and are analytic. But they are slow and scale poorly.
- (Extended) Kalman filtering is a well-studied classical method. It is sensitive to the initial state and model assumptions.
- SVM is a Linear Regression method that can be very powerful, but may be slow and hard to tune.
- Artificial Neural Networks (ANN) and Recurrent Neural Networks (RNN) offer good and stable performance and need no prior assumptions. They can be slow to train, require a lot of data, and suffer from convergence issues. LSTMs have especially gained a lot of popularity due to their ability to model complex sequential data and long-term temporal correlations. Extreme Learning Machine (ELM) is a single layer ANN which is claimed to learn faster.

The performance of machine learning approaches will depend to a large extent on the process of selecting and engineering features. One key ingredient is the construction of concise and uniform feature vectors. For vessel trajectory prediction from AIS data, *Liu et al. (2019a), Pallotta et al. (2013b), Tu et al. (2020b)* use GPS location, course and speed. *Pallotta et al. (2013b), Tu et al. (2020b)* also use static information (such as vessel type) and *Tu et al. (2020b)* include a categorical feature representing geographic areas (four areas were distinguished in total). *Liu et al. (2019a)* include time of day and seasonal effects to predict trajectories of fishing vessels over several hours.

Table II: Methods used for trajectory classification and predictions in maritime domain

Literature	Prediction horizon	Approach
Petrou et al. (2020a)	1-32 min	Neighbour-based
Hexeberg, (2017a)	2-30 min	Neighbour search
Valsamis et al. (2017b)	2-40 min	Linear Regression, NN, Random Forest
Dalsnes (2018)	5-15 min	One of the proposed methods predicts position based on randomly drawn course and speed. GMM
Borkowski (2017a)	6 min	NN with averaging as meta estimator
Tang et al. (2019a)	10 min	LSTM
Li et al. (2019a)	10-15 min	LSTM
Rong et al. (2019)	10-30 min	GPR
Wijaya and Nakamura (2013a)	10 min	kNN
Ristic et al. (2008)	10-72 min	Motion transition model is given by the Gaussian density under nearly constant velocity motion model.
Bomberger et al. (2006)	15 min	NN with associative learning
Tu et al. (2020a)	15-60 min	ELM on each cluster with kNN as meta estimator
Tu et al. (2018)	15-60 min	Fuzzy ARTMAP, GPR, SVM, NN, ELM, GMM
Liu et al. (2019b)	≤ 10 h	Markov chain
Pallotta et al. (2013a)	(Long-term, up to entire voyage)	KDE, statistical methods
Nguyen et al. (2018)	(Long-term, up to entire voyage)	RNN (gated recurrent units)
Mao et al. (2018)	20-40 min	ELM
Zhao and Shi (2019a)	15-60 min	LSTM
Gao et al. (2018b)	Unknown	Bidirectional LSTM
Kim and Jeong (2015)	Unknown	Support Vector Machine
Laxhammar (2011)	Unknown	Similarity-based Nearest Neightbour Conformal Anomaly Detector
Sheng and Yin (2018b)	Unknown	Modified DBSCAN
Qi and Zheng (2016)	Unknown	Support Vector Machine

3. Predicting surrounding traffic with contextual awareness

The work presented in this paper focuses on producing short-term (2 minutes) prediction of vessel trajectories based on its latest AIS message and trajectory for the past 8 minutes. The goal of this exercise is to provide a steppingstone for situational awareness as needed for autonomous vessels. Some attempts to include context information can be found in literature: *Pallotta et al. (2013b)* aims at accounting for seasonal variations and changing routing schemes, while *Hexeberg et al. (2017)* suggests including surrounding vessels predicted trajectories to improve estimates.

Although the field of maritime trajectory prediction and traffic analysis based on AIS-data is well established, to the best of our knowledge, there exists no research taking the perspective of a vessel, predicting the future of a traffic situation from the view of a reference vessel, accounting for this vessel's influence on the traffic situation due to its size and manoeuvrability. The here presented approach follows this idea, restricted to vessels in the immediate proximity of a given reference vessel. The underlying idea is that in order to produce meaningful predictions for proximal awareness, there is no need to be able to predict arbitrary vessel trajectories, but it suffices to model the behaviour of vessels interacting with the reference vessel. We capture the interaction implicitly through custom data selection and pre-processing, before using a data-driven prediction scheme. We demonstrate the feasibility of our approach by applying it to the vessels in service of the coastal route Bergen-Kirkenes in Norway, a costal traffic route providing daily, year-round service for 34 ports between the city of Bergen in the south-west and the town of Kirkenes on the Norwegian-Russian border in the north-east.

3.1. Data collection

Since we are only interested in relative behaviour, we are only interested in AIS messages that is derived from positions within a certain radius from our reference vessel. Since AIS data is sent in intervals of 2-10 seconds, extracting all messages within a given radius for each message belonging to the reference vessel becomes a prohibitive task for larger database. Hence, we discretise the search space geographically. In order to do so, we

1) collect all available messages from the reference vessel,
2) overlay each position with a square of adequate size (chosen search radius x sqrt(2)),
3) merge the squares into one large polygon, which we call the cover,
4) split this polygon into smaller fragments, based on a search grid

Fig.1 illustrates the process for a synthetic example.

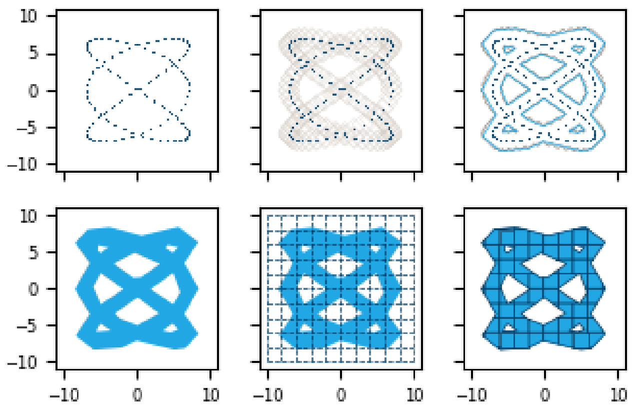


Fig.1: Illustration of the geographical discretization of the search space. Top row: raw positions (left), overlaid with squares (centre), and the merged polygon (grey) and a after simplifying the outline (light blue, the cover). Bottom row: The cover (left), overplayed with the search grid (centre), and split into fragments (right).

Now, we can restrict ourselves to query each of the bounding boxes of the fragments. While this step eliminates the geographical redundancy in the queries of the database, we expect that a large number of the messages from each search fragment are unrelated to the reference vessel, since it does not occupy the fragment at all times. To eliminate this temporal overhead, we construct a two-dimensional histogram for each of the fragments, indicating when the fragments is occupied by the reference vessel. We call these histograms therefore occupancy histograms. We generate one bin per day and one per hour, resulting in number of days x 24 discretisation of the timespan covered by the database. Examples of such occupancy histograms can be found in Fig.6.

Then we add all histograms from neighbouring fragments (within the selected search radius) to the fragments own histogram in order to capture all messages within the search radius. Finally, we collect all messages from each fragment and day for time spans with non-zero occupancy histogram bins.

3.2. Data Processing

Each of the retrieved AIS messages consists of a time stamp, its MMSI, geographic position, speed and course over ground and heading. In addition, we calculate the distance to the reference ship, the relative course and speed over ground, and the angular position as observed from the reference ship. Fig.2 illustrates the performed calculations.

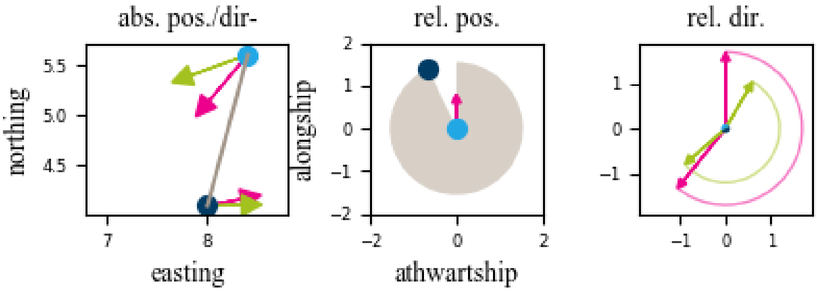


Fig.2: Illustration of the data processing step. The left figure shows the vessel (dark blue) and its reference vessel (light blue), and their heading (green) and course over ground (pink). The grey line indicates the distance between the vessels. The centre figures shows the angular position (illustrated through a wedge) of the vessel relative to the reference vessel. The angle is measured from the reference vessels course over ground. The fright figure shows the vessels course over ground and heading relative the reference vessels course over ground and heading, respectively. All angles are positive in clockwise direction following usual maritime conventions.

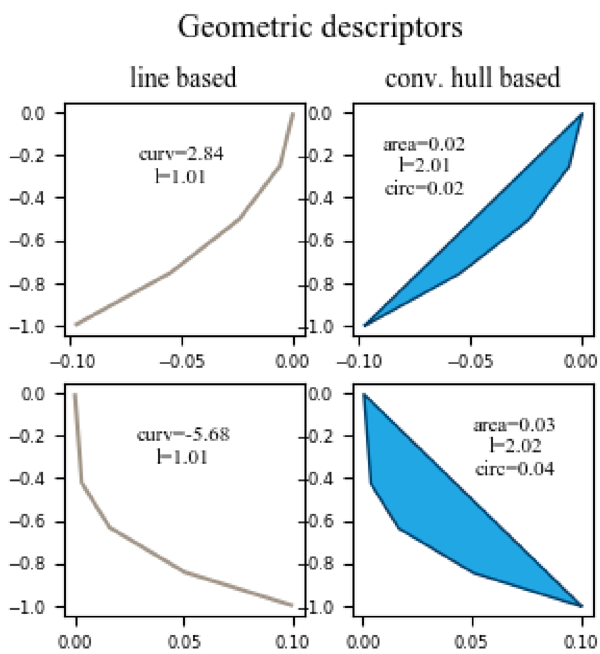


Fig.3: Illustration of the geometric descriptors derived from the normalised path pathlets. The first column shows the line-based measures, i.e. length and curvature, the second columns shows the descriptors based on the convex hull, i.e area, circumference and circularity. Circularity is calculated as $\boldsymbol{2\,\pi\, area/l^2}$. Circularity reaches it maximum of 1 for a circle.

Furthermore, we generate "tracklets" from the messages sent by the same vessel in the 8 minutes preceding the current AIS message and from the messages sent by the same vessel in the coming next 2 minutes after the current AIS message. The tracklets are normalized by transforming their coordinates into the vessel's reference frame. Wegeometric descriptor from the "past" tracklets, namely their length, curvature (measured as average difference between successive line segment orientations), the area and circumference of the convex hull of the tracklet, and their circularity (ratio between convex hull are and squared circumference times 2 pi). Fig.3 illustrates the geometric descriptors used.

Finally, we add context descriptors that represent factors besides the vessel and the reference vessel that can influence the vessel behavior, namely the distance to land, distance to the closest fairway, and the number of other vessels within the search radius. Fig.4 illustrates these descriptors for ca. 800 000 messages from the Trondheim area in Norway.

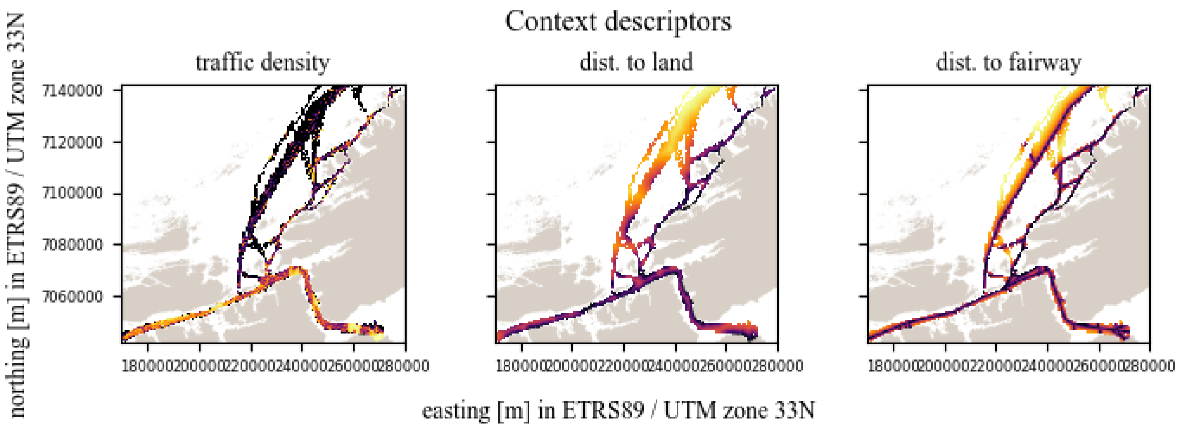


Fig.4: Context descriptors for approx. 800 000 AIS messages from the Trondheim ares. Colours indicate the average value for each location. Darker colours indicate lower values, while yellow indicates high values.

3.3. Future path prediction

For each new AIS message from a vessel within the search radius from the reference vessel, we retrieve the 10 nearest neighbours from the training dataset. The distance between the data point representing the new vessel and a data point from the training set is calculated using the Euclidian distance between a vector holding relative heading and course over ground with respect to the reference vessel, distance between the two, the geometric and contextual descriptor described above. Each entry is rescaled such that the range [-1, 1] represents the interquartile range over the training data of the corresponding component. Note that we deliberately refrain from using explicit information about the geographic area a message stems from. While local conditions strongly influence ship behaviour, using such parameters increase the demand for training data dramatically. Even using all vessel on the costal route Bergen-Kirkenes, we could ideally only collect one sample per position per day.

For each of the found nearest neighbours, we retrieve their future tracklets and translate and rotate then according to the real-world position and course over ground of the vessel. Finally, generate the convex hull of the future tracklets to represent area the vessel is estimated to occupy within the next two minutes. Hence, a second polygon is generated by growing the first by 10 m. to account for this inherent uncertainty.

3.4. Case study

Our case study uses the twelve vessels of the coastal route Bergen-Kirkenes as reference vessel. While they are not sister ships and hence not of equal dimensions, they regular schedule and sailing routes should suffice to warrant that the behaviour of vessels encountering one of them does not change significantly depending on which of them they encounter.

First, we extract all AIS messages for each of the twelve vessels has sent during 2014 and construct the polygon described above, using a search radius of 2 km. Fig.5 illustrates the process for the AIS messages sent from MS Kong Harald.

We merge the polygons (one per ship), and split the resulting merged polygon into fragments using a 1x1 km search grid. The figure below shows the process. After retrieving the AIS messages according to this spatiotemporal discretisation, we perform the data augmentation steps according to the section above. The final dataset consists of ~ 800 000 rows. Finally, we apply SciKit-Learn to build a k-NN search tree for efficient retrieval of the nearest neighbours needed for the prediction step. Next, we generate the occupancy histograms, as detailed in Fig.5. Fig.6 shows selected histograms from the Trondheim area.

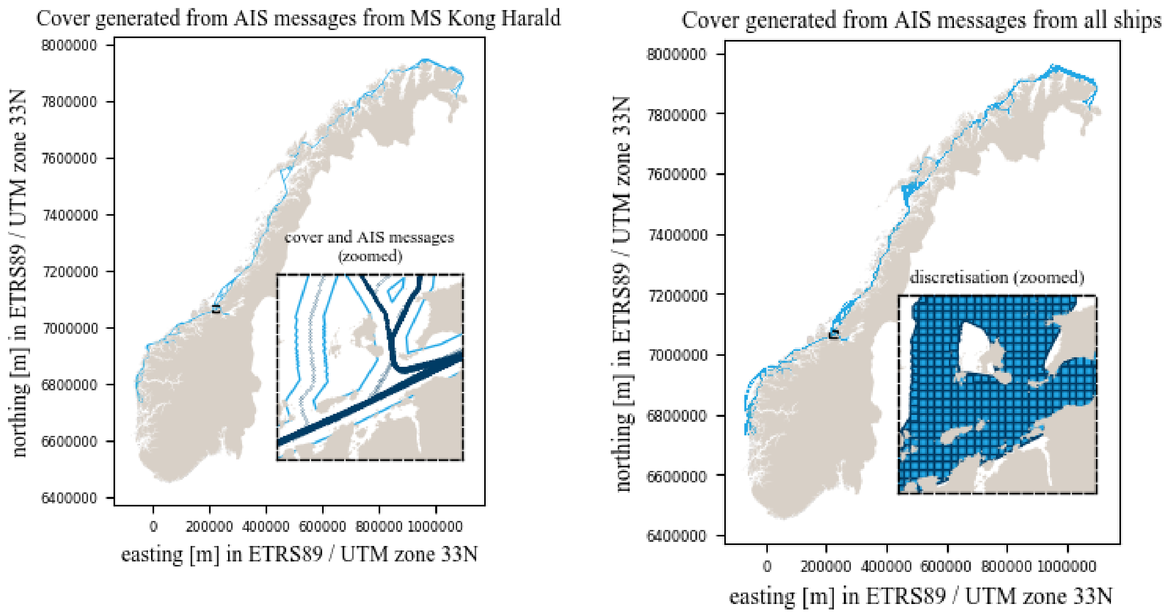


Fig.5: Left: The generation of the cover polygon for AIS messages from one specific vessel (MS Kong Harald). Right the merged cover from all reference vessels and its discretization into fragments.

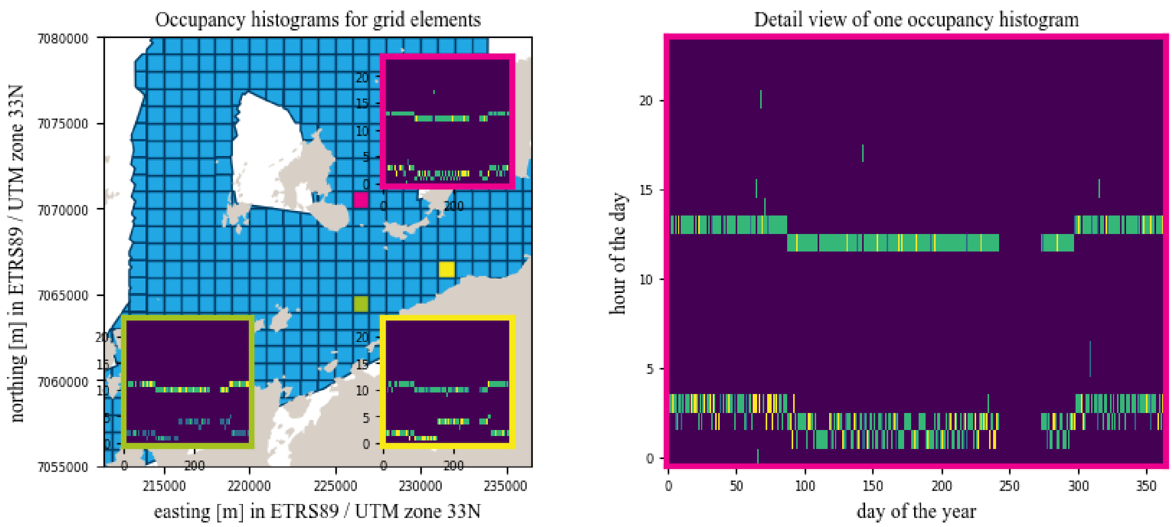


Fig.6: Occupancy histogram. Left: Occupancy histograms for three selected fragments right outside the Trondheimsfjord. Right: Detail view for one of them. Note the clear regular schedule, which corresponds nicely to the known pattern of one north- and one southbound departure from Trondheim each day, with a stop-over of several hours. The broken horizontal stripes are due to daylight saving time. The large number of unoccupied cells (blue) shows overhead retrieving all data from the fragment would generate.

3.5. Evaluation

We evaluate the performance of the presented approach based on five test cases in the Trondheim area. Two south of Trondheim with a south- (1) and northbound (2) reference vessel, respectively, one north of Trondheim with a northbound reference vessel (3) , one with the reference vessel entering the Trondheimsfjord (4), and one where the reference vessel is leaving Trondheim harbour (5).

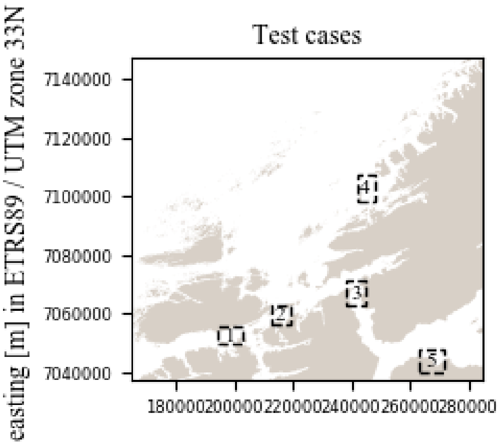


Fig.7: Overview of the test cases for evaluation of the approach

The plots of the predictions for the different test case, Fig.8, show that the predicted area (green and yellow polygons) a vessel will occupy within the next two minutes corresponds reasonably well with the ground truth (pink line), and that the approach handles situations with sparse and dense traffic as well as fast, slow and stationary vessels (speed is proportional to length of future tracks, stationary vessels are missing past tracklet, c.f. for harbour case).The predictions are tighter transit areas (case 1,2), while vessel in manoeuvring produce wider prediction polygons (3, 4).

4. Summary and future work

In this article a method for predicting short-term vessel behaviour in the proximity of a reference vessel has been introduced. The approach comprises a novel data retrieval and prepossessing and a k-NN prediction approach that uses relative measures only, reducing the demand on training data. The approach has shown promising results in a qualitative evaluation based on test cases. Future work will comprise a rigorous quantitative evaluation of the performance.

Furthermore, there is potential in experimenting with more advanced geometric descriptors and incorporating the physical dimensions of the vessels to predict, since they impact the manoeuvrability of vessel. In the current approach, all nearest neighbours contribute equally to the prediction, despite having different distances to the vessel to predict. I.e. the best neighbour contributes equally much as the worse. Investigating how to incorporate this inherent ranking of neighbours, and quantifying the prediction error is necessary when looking towards real-world applications.

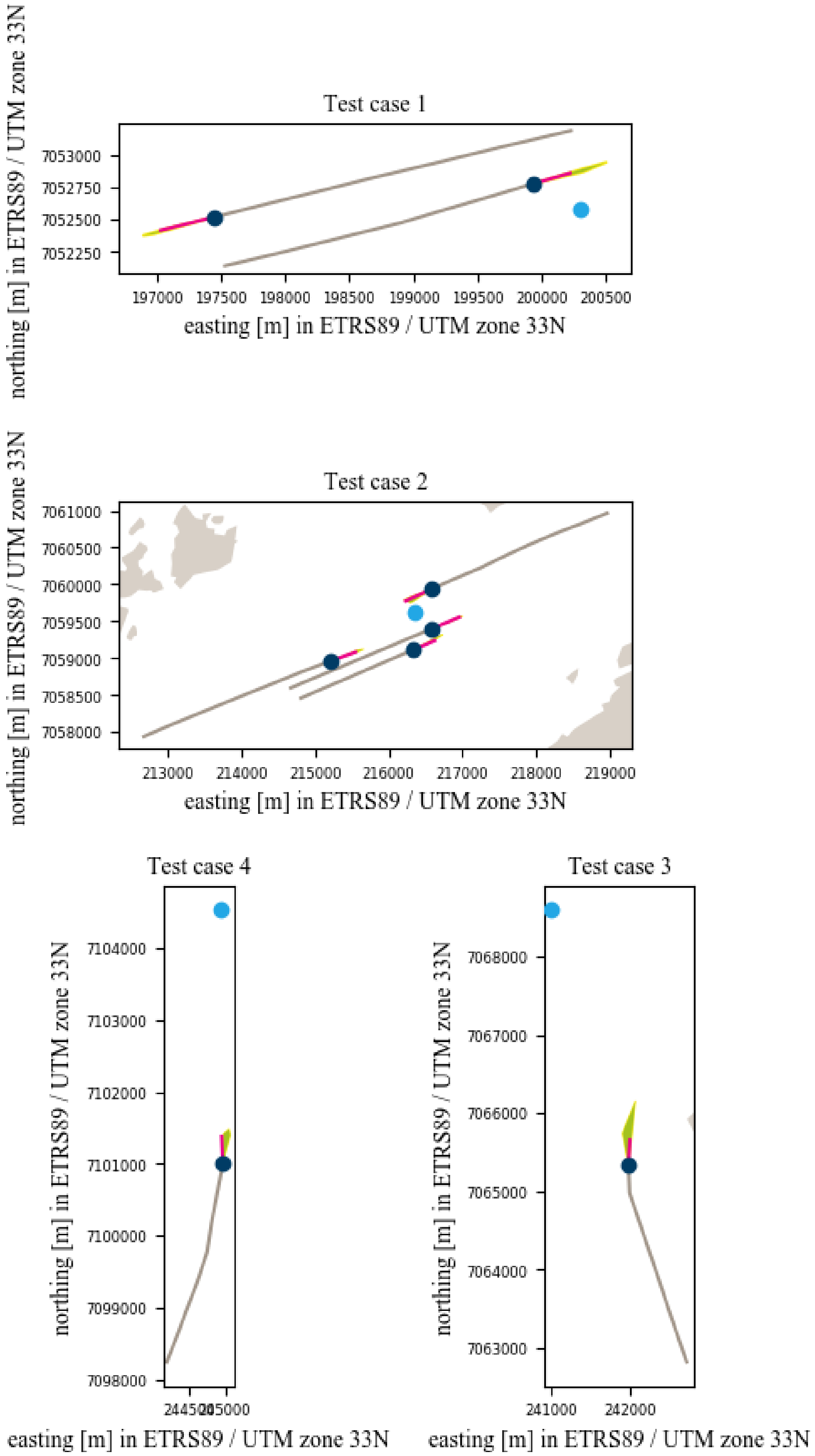
Acknowledgements

We thank our colleagues Esten Ingar Grøtli and Johannes Tjønnås at SINTEF Digital AS for valuable input, and Numan Muhammad at SINTEF Ålesund AS for assistance with data processing.

This paper is a part of the dissemination activities for the Innovation Project "Integrated maritime sensor platform for detection and classification of safety-critical objects around a ship – PROXIMA" project number 282334, supported by the Research Council of Norway.

The definitions of the fairways along the Norwegian coast, Maritime transport – Fairways, is a data provided by the Norwegian Coastal Administration under Norwegian Licence for Open Government Data though the national website for map data and other location information in Norway, Geonorge. Also, the AIS-data was provided by the Norwegian Coastal Administration.

The landmask for Norway is derived from the dataset Norwegian Counties and Municipalities 2020 (clipped by coastline), provided by The Norwegian Mapping Authority under Creative Commons 0 license through the national website for map data and other location information in Norway, Geonorge.


Test case 1
7053000
7052750
7052500
7052250
197000
197500
198000
198500
199000
199500
200000
200500
easting [m] in ETRS89 / UTM zone 33N
northing [m] in ETRS89 / UTM zone 33N
Test case 2
7061000
7060500
7060000
7059500
7059000
7058500
7058000
213000
214000
215000
216000
217000
218000
219000
easting [m] in ETRS89 / UTM zone 33N
northing [m] in ETRS89 / UTM zone 33N
Test case 4
7104000
7103000
7102000
7101000
7100000
7099000
7098000
northing [m] in ETRS89 / UTM zone 33N
easting [m] in ETRS89 / UTM zone 33N
Test case 3
7068000
7067000
7066000
7065000
7064000
7063000
241000
242000
northing [m] in ETRS89 / UTM zone 33N
easting [m] in ETRS89 / UTM zone 33N

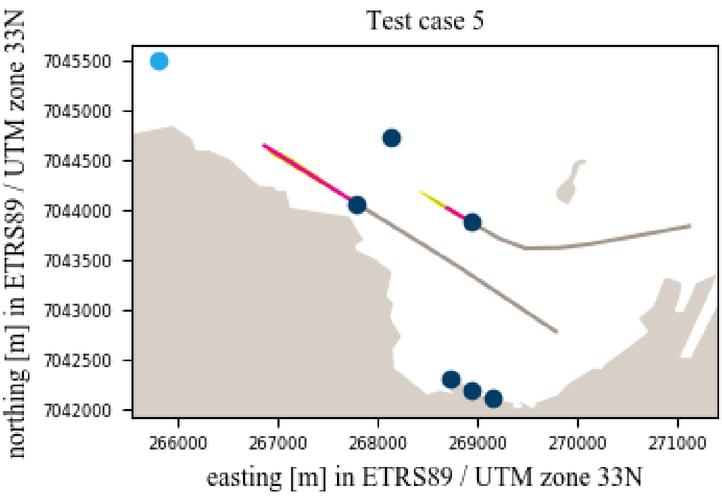


Fig.8: Evaluation of test case. The dark blue dots represent the positions of the vessels for which the area occupied during the next two minutes is to be estimated. The light blue dots indicate the position of the reference vessel. The green polygon indicates the estimated area, the yellow polygon the estimated area adjusted for positional uncertainty of the AIS data. The pink line indicates the true track for the coming 2 minutes. The here proposed method performs well in all test cases. Vessels during turning (case 3 and 4) produces wider estimates.

References

ALTCHE, F.; DE LA FORTELLE, A. (2018), *An LSTM network for highway trajectory prediction*, IEEE Conf. Intelligent Transportation Systems, *pp.*353–359

ALTCHÉ, F.; DE LA FORTELLE, A. (2017), *An LSTM network for highway trajectory prediction*, *IEEE Conf. Intelligent Transportation Systems*, pp.353-359

AYHAN, S.; SAMET, H. (2016a), *Aircraft Trajectory Prediction Made Easy with Predictive Analytics*, 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, pp.21-30

AYHAN, S.; SAMET, H. (2016b), *Aircraft trajectory prediction made easy with predictive analytics*, 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, pp.21-30

BOMBERGER, N.A.; RHODES, B.J.; SEIBERT, M.; WAXMAN, A.M. (2006), *Associative learning of vessel motion patterns for maritime situation awareness*, 9th Int. Conf. Information Fusion, FUSION

BORKOWSKI, P. (2017a), *The ship movement trajectory prediction algorithm using navigational data fusion*, Sensors 17(6)

BORKOWSKI, P. (2017b), *The Ship Movement Trajectory Prediction Algorithm Using Navigational Data Fusion*, Sensors *17*(6)

CAZZANTI, L.; PALLOTTA, G. (2015), *Mining maritime vessel traffic: Promises, challenges, techniques*, OCEANS, Genova, pp.1-6

CHEN, Z.; XUE, J.; WU, C.; QIN, L.; LIU, L.; CHENG, X. (2018), *Classification of vessel motion pattern in inland waterways based on Automatic Identification System*, Ocean Eng. 161, pp.69-76

DALSNES, B.R.; HEXEBERG, S.; FLÅTEN, A.L.; ERIKSEN, B.H.; BREKKE, E.F. (2018), *The Neighbor Course Distribution Method with Gaussian Mixture Models for AIS-Based Vessel Trajectory Prediction*, 21st International Conference on Information Fusion (FUSION), pp.580-587

DALSNES, B.R.; HEXEBERG, S.; FLÅTEN, A.L.; ERIKSEN, B.O.H.; BREKKE, E.F. (2018), *The Neighbor Course Distribution Method with Gaussian Mixture Models for AIS-Based Vessel Trajectory Prediction*, 21st Int. Conf. Information Fusion (FUSION), pp.580-587

DALSNES, B.R. (2018), *Long-term Vessel Prediction Using AIS Data*, Norwegian University of Science and Technology, Trondheim

DARANDA, A. (2016), *A Neural Network Approach to Predict Marine Traffic*, Baltic J. Modern Computing 4/3, pp.483-495

ENGEN, C.M. (2019), *Nøyaktig prediksjon av grunnstøtinger og fartøysbevegelser,* Norconsult

GAO, M.; SHI, G.; LI, S. (2018a), *Online prediction of ship behavior with automatic identification system sensor data using bidirectional long short-term memory recurrent neural network*, Sensors 18(12)

GAO, M.; SHI, G.; LI, S. (2018b), *Online prediction of ship behavior with automatic identification system sensor data using bidirectional long short-term memory recurrent neural network*, Sensors 18(12)

GRASER, A.; SCHMIDT, J.; DRAGASCHNIG, M.; WIDHALM, P. (2019), *Data-driven Trajectory Prediction and Spatial Variability of Prediction Performance in Maritime Location Based Services*, 15th Int. Conf. Location-Based Services, pp.129

HAGEN, M.B. (2018), *Ruteprediksjon av maritim trafikk med nevrale nettverk*, https://ntnuopen.ntnu.no/ntnu-xmlui/handle/11250/2564787

HEXEBERG, S. (2017a), *AIS-based Vessel Trajectory Prediction for ASV Collision Avoidance*. Norwegian University of Science and Technology

HEXEBERG, S. (2017b), *AIS-based Vessel Trajectory Prediction for ASV Collision Avoidance*. https://ntnuopen.ntnu.no/ntnu-xmlui/handle/11250/2452108

HEXEBERG, S.; FLÅTEN, A.L.; ERIKSEN, B.O.H.; BREKKE, E.F. (2017), *AIS-based vessel trajectory prediction*, 20th Int. Conf. Information Fusion (Fusion), pp.1-8

IMO (2001), *Resolution A.917(22) Guidelines for the Onboard Operational use of Shipborne Automatic Identification Systems (AIS)*, http://www.imo.org/blast/blastDataHelper.asp?data_id=24565&filename=A917(22).pdf

IMO (2015), *Resolution A. 1106(29) Revised guidelines for the operational use of shipborne automatic identification systems (AIS)*, http://www.imo.org/en/KnowledgeCentre/IndexofIMOResolutions/Assembly/Documents/A.1106(29).pdf

KIM, J.S.; JEONG, J.S. (2015), *Pattern Recognition of Ship Navigational Data Using Support Vector Machine*, Int. J. Fuzzy Logic and Intelligent Systems 15(4), pp.268-276

LAXHAMMAR, R. (2011), *Anomaly detection in trajectory data for surveillance applications*, http://urn.kb.se/resolve?urn=urn:nbn:se:oru:diva-17235

LEE, J.G.; HAN, J.; WHANG, K.Y. (2007), *Trajectory clustering: A partition-and-group framework*, ACM SIGMOD Int. Conf. Management of Data, pp.593-604

LEI, P.R.; TSAI, T.H.; PENG, W.C. (2016), *Discovering Maritime Traffic Route from AIS network*, 18th Asia-Pacific Network Operations and Management Symposium (APNOMS), pp.1-6

LI, W.; ZHANG, C.; MA, J.; JIA, C. (2019a), *Long-term vessel motion predication by modeling trajectory patterns with AIS data*, 5th Int. Conf. Transportation Information and Safety

LI, W.; ZHANG, C.; MA, J.; JIA, C. (2019b), *Long-term Vessel Motion Predication by Modeling Trajectory Patterns with AIS Data*, 5th Int. Conf. Transportation Information and Safety, pp.1389-1394

LIU, B.; DE SOUZA, E.N.; MATWIN, S.; SYDOW, M. (2014), *Knowledge-based clustering of ship trajectories using density-based approach*, IEEE Int. Conf. Big Data (Big Data), pp.603-608

LIU, C.; GUO, S.; FENG, Y.; HONG, F.; HUANG, H.; GUO, Z. (2019a), *L-VTP: Long-Term Vessel Trajectory Prediction Based on Multi-Source Data Analysis*, Sensors 19(20)

LIU, C.; GUO, S.; FENG, Y.; HONG, F.; HUANG, H.; GUO, Z. (2019b), *L-VTP: Long-term vessel trajectory prediction based on multi-source data analysis*, Sensors 19(20)

LIU, Y.; HANSEN, M. (2018a), *Predicting Aircraft Trajectories: A Deep Generative Convolutional Recurrent Neural Networks Approach*, https://arxiv.org/abs/1812.11670

LIU, Y., & HANSEN, M. (2018b), *Predicting Aircraft Trajectories: A Deep Generative Convolutional Recurrent Neural Networks Approach*, http://arxiv.org/abs/1812.11670

MAO, S.; TU, E.; ZHANG, G.; RACHMAWATI, L.; RAJABALLY, E.; HUANG, G.B. (2018), *An Automatic Identification System (AIS) Database for Maritime Trajectory Prediction and Data Mining*, ELM-2016, pp. 241-257

NGUYEN, D.D.; LE VAN, C.; ALI, M.I. (2018), *Demo: Vessel trajectory prediction using sequence-to-sequence models over spatial grid*, 12th ACM Int. Conf. Distributed and Event-Based Systems, pp.258-261

NGUYEN, D.D.; LE VAN, C.; ALI, M.I. (2018), *Vessel Trajectory Prediction using Sequence-to-Sequence Models over Spatial Grid*, 12th ACM Int. Conf. Distributed and Event-Based Systems, pp.258-261

PALLOTTA, G.; HORN, S.; BRACA, P.; BRYAN, K. (2014), *Context-enhanced vessel prediction based on Ornstein-Uhlenbeck processes using historical AIS traffic patterns: Real-world experimental results*, 17th Int. Conf. Information Fusion (FUSION), pp.1-7

PALLOTTA, G.; VESPE, M.; BRYAN, K. (2013a), *Vessel pattern knowledge discovery from AIS data: A framework for anomaly detection and route prediction*, Entropy 15(6), pp.2218-2245

PALLOTTA, G.; VESPE, M.; BRYAN, K. (2013b), *Vessel Pattern Knowledge Discovery from AIS Data: A Framework for Anomaly Detection and Route Prediction*, Entropy 15(6), pp.2218-2245

PATTERSON, A.; LAKSHMANAN, A.; HOVAKIMYAN, N. (2019), *Intent-Aware Probabilistic Trajectory Estimation for Collision Prediction with Uncertainty Quantification*, http://arxiv.org/abs/1904.02765

Pei, Z., Qi, X., Zhang, Y., Ma, M., & Yang, Y. H. (2019a). Human trajectory prediction in crowded scene using social-affinity Long Short-Term Memory. *Pattern Recognition*, *93*, 273–282.

PEI, Z.; QI, X.; ZHANG, Y.; MA, M.; YANG, Y.H. (2019b), *Human trajectory prediction in crowded scene using social-affinity Long Short-Term Memory*, Pattern Recognition 93, pp.273-282

PETROU, P.; TAMPAKIS, P.; GEORGIOU, H.; PELEKIS, N.; THEODORIDIS, Y. (2020a), *Online Long-Term Trajectory Prediction Based on Mined Route Patterns*, Lecture Notes in Computer Science Vol. 11889 LNAI

PETROU, P.; TAMPAKIS, P.; GEORGIOU, H.; PELEKIS, N.; THEODORIDIS, Y. (2020b), *Online Long-Term Trajectory Prediction Based on Mined Route Patterns,* Multiple-Aspect Analysis of Semantic Trajectories, pp.34-49

QI, L.; ZHENG, Z. (2016), *Trajectory Prediction of Vessels based on Data Mining and Machine Learning,* J. Digit. Inf. Manage 14(1), pp.33-40

RIDEL, D.; REHDER, E.; LAUER, M.; STILLER, C.; WOLF, D. (2018), *A Literature Review on the Prediction of Pedestrian Behavior in Urban Scenarios,* 21st Int. Conf. Intelligent Transportation Systems, pp.3105-3112

RISTIC, B.; LA SCALA, B.; MORELANDE, M.; GORDON, N. (2008), *Statistical analysis of motion patterns in AIS data: Anomaly detection and motion prediction*, 11th Int. Conf. Information Fusion, FUSION

RONG, H.; TEIXEIRA, A.P.; GUEDES SOARES, C. (2019), *Ship trajectory uncertainty prediction based on a Gaussian Process model*, Ocean Engineering

SHENG, P.; YIN, J. (2018a), *Extracting shipping route patterns by trajectory clustering model based on Automatic Identification System data*, Sustainability 10(7)

SHENG, P.; YIN, J. (2018b), *Extracting Shipping Route Patterns by Trajectory Clustering Model Based on Automatic Identification System Data*, Sustainability 10(7)

SUN, L.; ZHOU, W. (2017), *Vessel Motion Statistical Learning based on Stored AIS Data and Its Application to Trajectory Prediction*, 5th Int. Conf. Machinery, Materials and Computing Technology

TAMPAKIS, P.; PELEKIS, N.; DOULKERIDIS, C.; THEODORIDIS, Y. (2019), *Scalable Distributed Subtrajectory Clustering*, IEEE Int. Conf. Big Data (Big Data), pp.950-959

TANG, H.; YIN, Y.; SHEN, H. (2019a), *A model for vessel trajectory prediction based on long short-term memory neural network*, J. Marine Engineering and Technology, 4177

TANG, H.; YIN, Y.; SHEN, H. (2019b), *A model for vessel trajectory prediction based on long short-term memory neural network*, J. Marine Engineering and Technology, pp.1-10

TU, E.; ZHANG, G.; MAO, S.; RACHMAWATI, L.; HUANG, G.B. (2020a), *Modeling Historical AIS Data For Vessel Path Prediction: A Comprehensive Treatment*, https://arxiv.org/abs/2001.01592

TU, E.; ZHANG, G.; MAO, S.; RACHMAWATI, L.; HUANG, G.B. (2020b), *Modeling Historical AIS Data For Vessel Path Prediction: A Comprehensive Treatment*, http://arxiv.org/abs/2001.01592

TU, E.; ZHANG, G.; RACHMAWATI, L.; RAJABALLY, E.; HUANG, G.B. (2018), *Exploiting AIS Data for Intelligent Maritime Navigation: A Comprehensive Survey from Data to Methodology*, IEEE Transactions on Intelligent Transportation Systems 19(5), pp.1559-1582

VALSAMIS, A.; TSERPES, K.; ZISSIS, D.; ANAGNOSTOPOULOS, D.; VARVARIGOU, T. (2017a), *Employing traditional machine learning algorithms for big data streams analysis: The case of object trajectory prediction,* J. Systems and Software 127, pp.249-257

VALSAMIS, A.; TSERPES, K.; ZISSIS, D.; ANAGNOSTOPOULOS, D.; VARVARIGOU, T.

(2017b), *Employing traditional machine learning algorithms for big data streams analysis: The case of object trajectory prediction*, J. Systems and Software

WANG, Z.; DELAHAYE, D. (2017), *Short-term 4D Trajectory Prediction Using Machine Learning Methods,* Sesar Innovation Days 2017

WANG, Z.; LIANG, M.; DELAHAYE, D. (2017), *Short-term 4D Trajectory Prediction Using Machine Learning Methods,* 7th SESAR Innovation Days, https://hal-enac.archives-ouvertes.fr/hal-01652041

WIJAYA, W.M.; NAKAMURA, Y. (2013a), *Predicting ship behavior navigating through heavily trafficked fairways by analyzing AIS data on apache HBase,* 1st Int. Symp. Computing and Networking, CANDAR, pp.220-226

WIJAYA, W.M.; NAKAMURA, Y. (2013b), *Predicting Ship Behavior Navigating through Heavily Trafficked Fairways by Analyzing AIS Data on Apache HBase*, 1st Int. Symp. Computing and Networking, pp.220-226

XUE, P.; LIU, J.; CHEN, S.; ZHOU, Z.; HUO, Y.; ZHENG, N. (2019), *Crossing-Road Pedestrian Trajectory Prediction via Encoder-Decoder LSTM*, IEEE Intelligent Transportation Systems Conf. (ITSC), pp.2027-2033

XUE, PEIXIN, LIU, J., CHEN, S., ZHOU, Z., HUO, Y., & ZHENG, N. (2019). Crossing-Road Pedestrian Trajectory Prediction via Encoder-Decoder LSTM. *2019 IEEE Intelligent Transportation Systems Conf. (ITSC)*, 2027–2033

YANG, K.; BI, M.; LIU, Y.; ZHANG, Y. (2019), *LSTM-Based Deep Learning Model for Civil Aircraft Position and Attitude Prediction Approach,* Chinese Control Conf. (CCC), pp.8689-8694.

YANG, K.; BI, M.; LIU, Y.; ZHANG, Y. (2019), *LSTM-Based Deep Learning Model for Civil Aircraft Position and Attitude Prediction Approach,* Chinese Control Conf. (CCC), pp.8689–8694

ZHAO, L.; SHI, G. (2019a), *Maritime Anomaly Detection using Density-based Clustering and Recurrent Neural Network,* J. Navigation 72(4), pp.894-916

ZHAO, L.; SHI, G. (2019b), *A Novel Similarity Measure for Clustering Vessel Trajectories Based on Dynamic Time Warping,* J. Navigation 72(2), pp.290-306

ZHAO, Z.; ZENG, W.; QUAN, Z.; CHEN, M.; YANG, Z. (2019), *Aircraft Trajectory Prediction Using Deep Long Short-Term Memory Networks*, CICTP 2019, pp.124-135

ZHU, F. (2011), *Mining ship spatial trajectory patterns from AIS database for maritime surveillance,* 2nd IEEE Int. Conf. Emergency Management and Management Sciences, pp.772-775

Towards Remote Inspections of Maritime Vessels Using Drones Instrumented with Computer Vision and Hyperspectral Imaging

Erik Stensrud, DNV GL, Oslo/Norway, erik.stensrud@dnvgl.com
Torbjørn Skramstad, NTNU, Trondheim/Norway, torbjorn.skramstad@ntnu.no
Muhammad Bashir, Jotun, Sandefjord/Norway, Muhammad.Bashir@jotun.no
Joseph Garrett, NTNU, Trondheim/Norway, joseph.garrett@ntnu.no
Geir Hamre, DNV GL, Trondheim/Norway, geir.hamre@dnvgl.com
Kristian Klausen, Scout Drone Inspection, Trondheim/Norway, kristian.klausen@scoutdi.com
Bahman Raeissi, DNV GL, Oslo/Norway, bahman.raeissi@dnvgl.com
Petter Rossvoll, Idletechs, Trondheim/Norway, petter.rossvoll@idletechs.com
Jing Xie, DNV GL, Oslo/Norway, jing.xie@dnvgl.com
André Ødegårdstuen, DNV GL, Oslo/Norway, andre.odegardstuen@dnvgl.com

Abstract

DNV GL has performed production surveys in enclosed spaces using drones since 2016, demonstrating cost savings and increased personnel safety. It is a vision to develop autonomous inspection drones to reduce the need to enter tanks. We expect that this will reduce survey duration and survey preparation costs for the clients and be a major safety improvement for surveyors. Several drone capabilities are required to enable visual close-up inspection and non-destructive testing in enclosed, GPS-denied, and poorly lit environments. In this study, we report the most recent status from an ongoing research project managed by DNV GL and including several industry partners. We highlight technical challenges and preliminary results on drone navigation functionalities, computer vision for detection of cracks, and the use of hyperspectral imaging to detect rust mechanisms and the chemical composition of coatings.

1. Introduction

This paper reports the most recent results of an ongoing R&D project managed by DNV GL and including the industry partners Jotun, Scout Drone Inspection, Idletechs, Norsk Elektro Optikk, and the university partner NTNU (The Norwegian University of Science and Technology). The project is partly funded by the Norwegian Research Council for 3 years and started mid-2018.

Stensrud et al. (2019) provided the motivation and the background for this R&D project, so this is not repeated. The previous paper also presented the current survey regime; it presented the vision, goals, and benefits of a future, remote inspection regime; it presented drone usage scenarios for survey and inspection; so, none of this is repeated either.

2. The inspection drone – capabilities and results

This section presents some of the results from experimental tests conducted with the drone system from Scout Drone Inspection. They are divided into the following three categories:

1. Navigation and control: A key aspect is the ability of the drone be operated safely and know its position relative to the tank.
2. Automatic Ultrasonic Thickness Measurements (UTM)
3. Operational aspects: This includes cloud transmission and other operational aspects.

Finally, a test-case is described where the drone was operated at an FPSO to test points 1 and 3.

2.1 Navigation and Control

The navigation and control system of the drone must ensure that the drone is able to operate safely, while also being able to provide a precise location of where the drone is in the inspection asset.

Fig.1: Scout 137, an inspection drone specialized to tank inspections, http://scoutdi.com

The drone depicted in Fig.1 is equipped with a 3D laser scanner to provide the raw measurements required to solve these problems. The laser scanner provides measurements at 10 Hz, which the onboard computer processes real-time to solve the problem of Simultaneous Localization and Mapping (SLAM). This is a formulation of a problem where the robot tries to find its location in an unknown environment. In addition, the laser provides data to an on-board collision avoidance system. By checking if any object is detected within a certain distance from the drone, the drone can be stopped from coming too close. The system then overrides the user input to prevent a collision.

An example of the map created by the drone can be seen in Fig.3. Quantitative comparisons are not yet available.

2.2 Automatic Thickness Measurement - UTM

As was indicated in *Stensrud et al. (2019)*, manual control of a drone doing UTM (ultrasonic thickness measurement) requires an expert pilot that has to keep the drone steady and keep the required force on the object to be measured. In that paper, it was reported that the pilot was guided by an altitude sensor, requiring the pilot to control the forward and sideways motion. This system has now been expanded to allow the measurement to be taken fully automatically. Initial tests conducted in a controlled environment showed how data from the on-board laser can aid in this maneuver. The full test-setup is described below.

A prototype version of the Scout drone was equipped with an off-the-shelf UTM-device designed to be integrated on drones. To minimize weight, much of the housings and probe holders were re-designed to better fit the current needs. Likewise, to limit the scope of the test, UTM gel was manually applied to the probe without the on-board tank dispenser system.

Furthermore, a prototype GUI element was created on a tablet that allowed an operator to perform the following operations with mostly high-level commands:

- Takeoff: The drone would hover 1m over the ground
- Free movements: By using a joystick, the operator could position the drone in front of the object to be measured.
- Measure: By the press of a button, the drone would automatically approach the wall, do the measurement, and go back to its initial position. The resulting measurement is displayed in the GUI
- Land: After the mission, the drone could be commanded to land at its current position.

Fig.**2** is a snapshot of a video demonstrating the semi-automatic UTM capability of the drone.

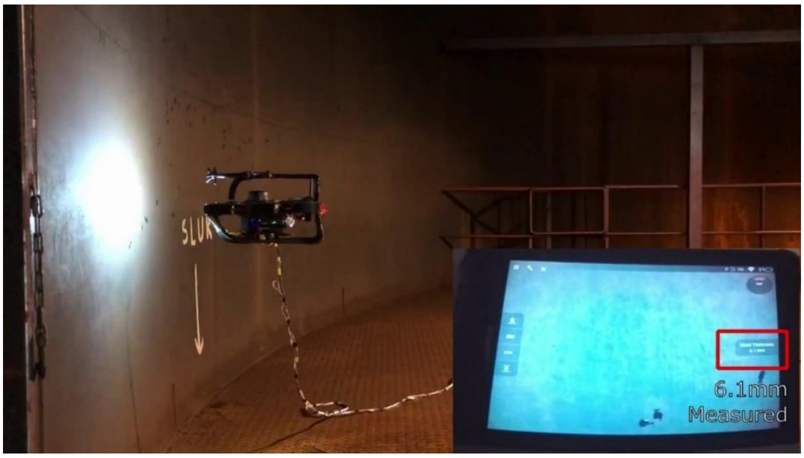


Fig.2: Illustration of the controlled scenario where the drone automatically performed UTM. The drone in the image is a prototype drone from Scout Drone Inspection

2.3 Operational aspects

The drone developed by Scout Drone Inspection is using a tethered power solution instead of a battery. This allows the drone to be used for prolonged durations, without the need to land for a battery-swap. This also allows the drone to be built quite small, less than the size of a 400 mm manhole, while still being able to carry the necessary payload and navigation equipment.

Crucially, the cable also allows data to flow through the same tether, using the same cable-pair as is transmitting power. This enables a secure and reliable connection, even if the drone is not in line-of-sight from a ground station or operator.

When using a tethered drone, some care must be taken to avoid getting the cable tangled or getting stuck to structures in the tank. It remains to assess and test various mitigation strategies.

2.4 Field testing of the drone onboard an FPSO

The drone capabilities for doing close-up visual inspection in a cargo tank were recently demonstrated onboard an FPSO, in a cargo tank. This section reports on the drone navigational tests. Section 4.5 reports on the drone's computer vision tests.

In early March 2020, a demonstration was conducted as a cooperation between an FPSO operator, DNV-GL, and Scout Drone Inspection to inspect one of the cargo tanks on-board an FPSO. The vessel was anchored to shore during the tests.

The goals were twofold: to conduct both General Visual Inspection (GVI) and Close Visual Inspection (CVI) of certain pre-defined hotspots in the tank, and stream the inspection live over the internet to demonstrate remote inspection.

Two different operational scenarios were considered. In the first, the drone operator was inside of the tank, operating the drone through the app which supplied video feed and position data as described above. In the second, the drone was operated while the drone operator was on deck. In both operational scenarios, the system allows the operator to easily judge and sense the current position of the drone, to know that the drone was observing the correct spots. The generated map is in Fig.**3**.

It was shown that due to the automatic navigation system and collision prevention, it seems feasible that the expert inspector (or other on-site available crew) will be able to operate the drone without the need of a specialized drone operator. Furthermore, by using an on-premise 4G router, the inspection was streamed live over the internet to each stakeholder.

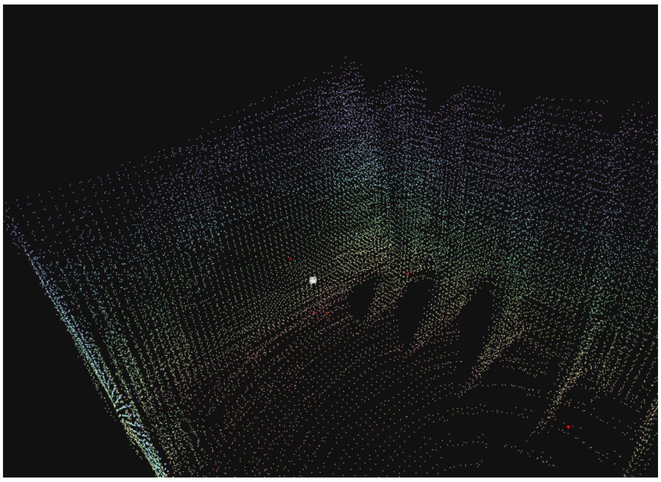


Fig.3: Map point cloud generated by drone

3. Hyperspectral imaging – test trials and results

3.1 Hyperspectral imaging technology

A hyperspectral camera observes many more colors (wavelengths) than an RGB camera. While an RGB camera sees only three colors: red, green, and blue, a hyperspectral (HS) camera can record around one hundred or more spectral bands. Hyperspectral imaging is used to classify different materials based on their characteristic spectra, which emerges from molecular absorption and scattering and is used in several applications across several industries, such as agricultural monitoring and food quality assurance.

In the ADRASSO project we are testing the capability of hyperspectral imaging to detect different types of phenomena that might endanger a ship's structural strength, such as corrosion. In particular, we want to find out whether hyperspectral imaging could reveal faults that human surveyors are not able to see. In the project, we have, together with DNV GL surveyors, identified 13 scenarios to investigate. In this paper, we present (preliminary) results from two different use cases.

3.2 HS use cases

The first use case reported is testing the capability of hyperspectral imaging to distinguish between different rust mechanisms. Rust is often present in ship tanks and on the hull. The rust can have been caused by humidity, humidity mixed with salt, or microbes, microbial induced corrosion (MIC). The latter rust mechanism is usual in FPSO's because the crude oil contains microbes. We want to know the rust mechanism because it impacts the corrosion rate. For example, MIC is more aggressive than the other rust mechanisms and will require other maintenance plans. It is therefore important to know the rust mechanism to better plan the maintenance of the structure. In this study, we report on the ability of HS imaging to distinguish between humidity-rusted steel and salt-rusted steel.

The second use case reported in this study, is on chemical binders in paint (The binder is the film-forming component of paint and is an essential part of a coating's makeup that must be in the mix, so that everything else meshes together.). The coating used in ship tanks and on hulls can have different chemical compositions. In order to coat an industrial structure or substrate, it is essential to know what needs to be mixed into a coating for the specific application. They impart properties such as gloss, durability, flexibility, and toughness. Getting the correct ratio of binders in coatings makes a big difference in its appearance and durability.

There are different binder chemistries that are used in various coatings, to obtain a range of different mechanical properties. Degradation phenomena of these binders are also different. Therefore, we are interested in the capability of HS analyses to identify differences between coatings with similar color/appearance but with different binders and how strong is adherence of paint to the underlying steel.

3.3 HS data collection – rust mechanism use case

We prepared test panels at Jotun's material laboratory. These are steel plates approximately 7,5x15 cm. The laboratory has several chambers in which temperature, humidity and salt contamination can be applied to steel plates in order to corrode them. We also exposed the test panels to the laboratory environment for different time durations.

HS image data were collected with two different HS cameras, one high-end camera (Hyspex from NEO) and one small, cheap, light-weight camera (from NTNU). Both are push-broom cameras.

The Hyspex VNIR-1800 has 182 bands, 1800 pixels, and 0.4-1.0 µm spectral range. It weighs 5.0 kg, consumes 30 W, and the dimensions are 39 x 9.9 x 15 cm^3, (https://www.hyspex.com/hyspex-products/hyspex-classic/hyspex-vnir-1800/). It is operated in the NEO laboratory, which is a controlled environment with constant camera distance, constant scanning speed, and stable and uniform lighting.

The NTNU camera is built from commercial off-the-shelf and 3D-printed parts. It is smaller, weighs 152 g, and requires less power, but only covers a ~ 0.4-0.8 µm spectral range binned into 280 bands during pre-processing, *Sigernes et al. (2018).* The pixel size in this study of the Hyspex and NTNU cameras is ~ 0.3x0.1mm and 2x0.4mm, respectively. The Hyspex camera is the high-end, benchmark camera. We test the small and cheap NTNU camera to see if it is good enough for the different use cases, as it would be easier to mount on a small inspection drone.

3.4 Data – rust mechanism use case

For the analysis presented in this paper, 20 panels from three different categories were used: 2 painted, 9 humidity-rusted, and 9 salt-spray-rusted. The two painted panels have paint thicknesses of 35 and 95 microns, respectively. The 9 humidity-rusted panels have different *rust severities*, rusted in the chamber for 4, 7, and 13 days, respectively. The salt-spray rusted panels also have different rust severities, 1, 4, and 7 days in the chamber, respectively. The NEO and NTNU cameras scanned the same paint panels, while the NEO camera scanned one of each rusted panel per rust-severity level, whereas the NTNU camera scanned two. This helped to compensate for the smaller number of total pixels in each scan from the NTNU camera. The panels were divided into training and test panels.

The training data were taken from one panel of each type. The training paint panel had 35 µm thick paint. The rust training data were selected from the medium rust severity level, which was 7 days for the humidity-rust and 4 days for the salt-spray rust. The medium rust severity level was chosen for the training data, hypothesizing that it would be more representative of the rust mechanism than either of the extreme rust severities.

Table I: Rust mechanism; number of pixels

	NEO train	NEO test	NTNU train	NTNU test
Paint	200.000	300.000	8.000	11.200
Humidity rust	200.000	300.000	6.370	48.200
Salt-spray rust	200.000	300.000	10.000	45.950

The test data were taken from a second painted panel, as well as several panels that were similarly exposed to environmental stresses, but for different durations. The number of data points is provided in Table I. Each pixel is a data point. The NEO camera has higher spatial resolution and therefore captured more data points per panel. Some parts of the panels were obscured by the device used to hold them during imaging by the NTNU camera. The portions obscured by the positioning device were not used for classification.

3.5 HS analysis procedure – rust mechanism use case

The same approach was used for the analysis of the data from both cameras.

- Data preparation and analysis method
  The pixel data were transformed from reflectance to absorbance in order to emphasize molecular absorption relative to scattering and analyzed using Principal Components Analysis (PCA). The mean spectrum of the training data was calculated for each class: humidity rust, salt-spray rust, and paint, respectively. Then, pixels from the test dataset were classified by calculating their distance to the mean of the training data. A distance threshold was set so that test pixels were classified as "unknown" if the distance to any of the three classes was farther away than the threshold.

- Building the classification models
  The models we use are found as a decomposition of each (n x k) training data matrix **X,** into scores, ***T*** *(n x A)*, and loadings, ***P*** *(A x k)*, where n is the number of samples (or spatial pixels in the detector in the HSI case) and k the number variables (number of wavelength channels), and $A = 3$ is the number of principal components used (in our case 3). Each principal component consists of a score for each pixel and the wavelength loadings used to calculate those scores, *Vitale et al. (2017).* The scores can take in any real number, positive or negative. The scores associated with each principal component are then used for classification. The loadings are the same for all the different training samples, but the scores differ. After some trial and error, we used the scores of the first three principle components of the training data set from the NEO hyperspectral camera in the classification.

- Classification
  Classification is obtained by calculating the Mahalanobis distance between a pixel and the mean of the different training classes in the PCA model. The pixel is classified as the nearest class. For distances over 7 standard deviations from any class mean, the pixel is classified as "unknown".

3.6 HS analysis results – rust mechanism use case

Both of the hyperspectral cameras are able to distinguish between paint and rust, as expected. The confusion matrices in Fig.4 (left) and Fig.5 (left) present the classification of the training data pixels, the percent of pixels classified in each category. One observation is that most training pixels, for both cameras, are classified as the correct class. However, not all training pixels are classified correctly. This is expected, since a rusted panel is not rusted completely homogeneously. Some areas of the panel will exhibit more rust than other areas, and some pixels of salt-spray rusted panels may not contain much salt but just humidity.

From Fig. 4 and Fig.5 you can see that more pixels in the NTNU camera are classified "correctly". We suspect that the reason for this is that the NTNU camera is blurrier than the NEO camera and will not be able to spatially and spectrally separate all the variability in the samples that are removed when you use the average spectrum for training the model. Because the camera from NEO has very sharp optics per pixel and very low misregistration, it can detect subpixel variations in the samples. If we wanted to improve the results here, we would need to have more uniform samples and better ground truth to train the models correctly.

In Fig.**4** (right) and Fig.**5** (right), the confusion matrices present the classification of the test data pixels. The simple classification model tested here shows that the NEO camera separates salt and humidity rust with above 80% accuracy in the test data set. The NTNU camera classifies salt-spray rust correctly 53%, but incorrectly classifies salt-spray rust as "unknown" 40%. The results suggest that hyperspectral imaging is a viable tool for distinguishing salt-induced rust from rust caused only by humidity.

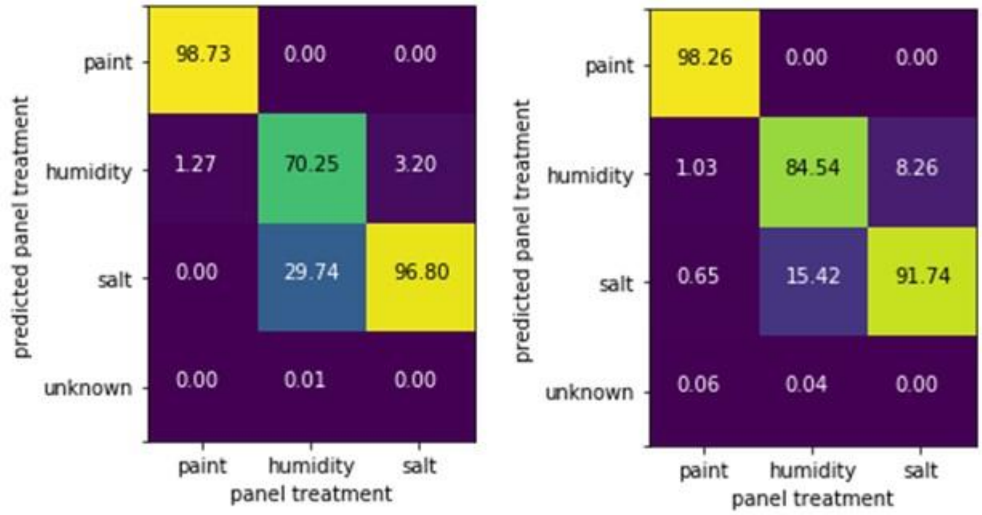


Fig.4: Results from the model classification on the NEO training (left) and test (right) data. The per-pixel percent accuracy is displayed for each category.

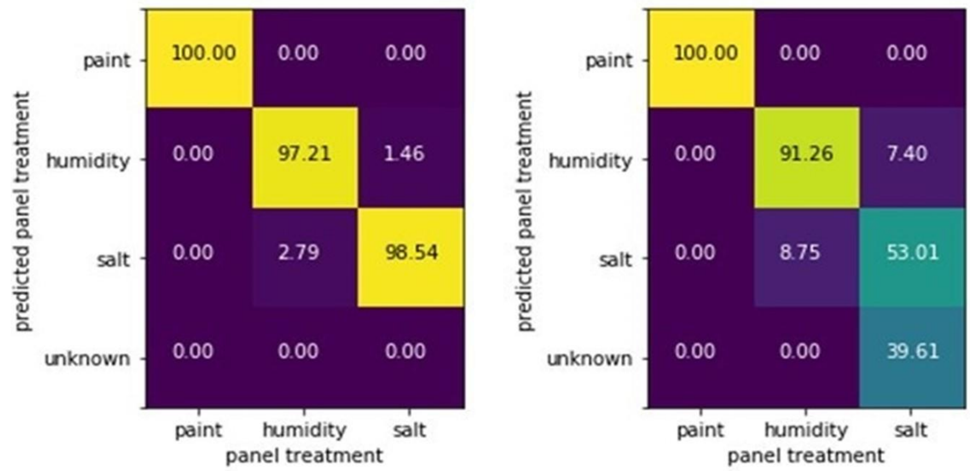


Fig.5: Results from the model classification for the training (left) and test (right) data from the NTNU camera

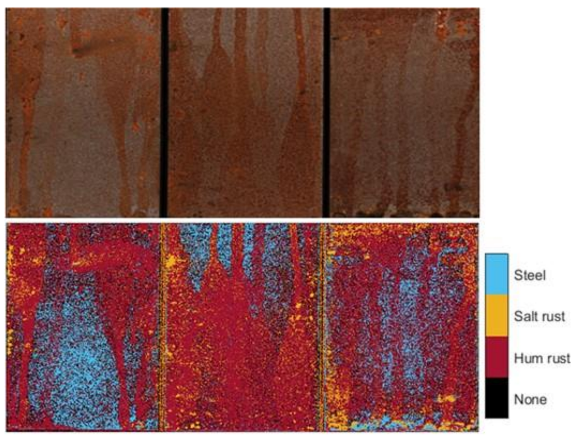


Fig.6: Humidity-rusted panels imaged by NEO camera and classified according to NEO camera derived model determined by the procedure described above. Panels were exposed to humidity for 4, 7, or 13 days (left to right).

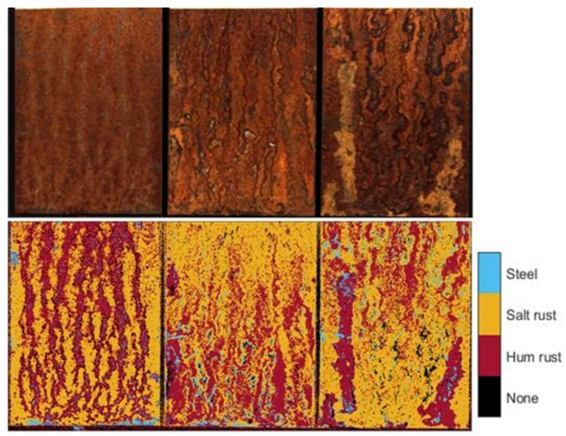


Fig.7: Salt-rusted panels imaged by NEO camera and classified according to NEO camera derived model determined by the procedure described above. Panels were exposed to salt spray for 1, 4, or 7 days (left to right).

Fig.**6** presents an example of the humidity-rusted panels of the different rust severities (RGB image at the top), and the bottom of the figure shows in which classes the pixels that have been classified. This example also includes the class “steel”, i.e. non-rusted steel.

Fig.**7** presents an example of the salt-spray rusted panels of the different rust severities (RGB image at the top), and the bottom of the figure shows in which classes the pixels that have been classified. This example also includes the class “steel”, i.e. non-rusted steel.

3.7 HS discussion and conclusions – rust mechanism

The results suggest that we may distinguish between salt-spray rust and humidity rust. The results are, however, preliminary. The first use case, to distinguish between salt-spray rust and humidity rust, seems to work with the lower spectral and spatial resolution of the NTNU camera. This is encouraging since the aim is to develop a HS camera with better specifications than the NTNU camera but still much smaller and cheaper than the existing NEO camera.

However, it is too early in the research to draw conclusions with high confidence. There are several error sources in the study. The two cameras have very different key quality parameters; while the NEO cameras have very sharp pixels (smaller PSF than the pixel size) for all places in the FOV and for the whole wavelength range, the NTNU camera has very different PSF depending on where you are in the FOV and a huge variation in the PSF as a function of wavelength. The NEO camera has smaller than 10%, of a pixel and bands, spatial and spectral misregistration. While the NTNU camera has spatial and spectral misregistration on the order of a pixel, after calibration and pre-processing *Henriksen et al. (2019)*.

3.8 Chemical binders use case - HS data, data collection, method

For the analysis presented in this paper, two panels were used, one panel with a phenolic epoxy cured by an amine curing agent and the second panel with polydimethylsiloxane (PDMS) containing 3-aminopropyltrimethoxysilane as curing agent. Both panels were cured at room temperature for one week. The panels were produced as reference panels for the mentioned binders in order to check the ability of an HS camera in differentiating between chemically different binders. There are specific chemical compounds (generally known as functional groups) in each cured binder, which result in some peaks in the mean spectra that are well-known from scientific literature.

The binder panels were scanned with HySpex SWIR_384_SN3129 camera at NEO’s lab, which is a push-broom camera with 288 bands, 384 pixels, and 1.0-2.5 µm spectral range. The number of pixels collected is shown in Table II:. Fig.**8** shows the RGB image of two panels which were coated with different binders used commercially in various coatings.

Table II: Chemical binders; number of pixels (N)

Phenolic epoxy	12.600 pixels
PDMS	12.600 pixels

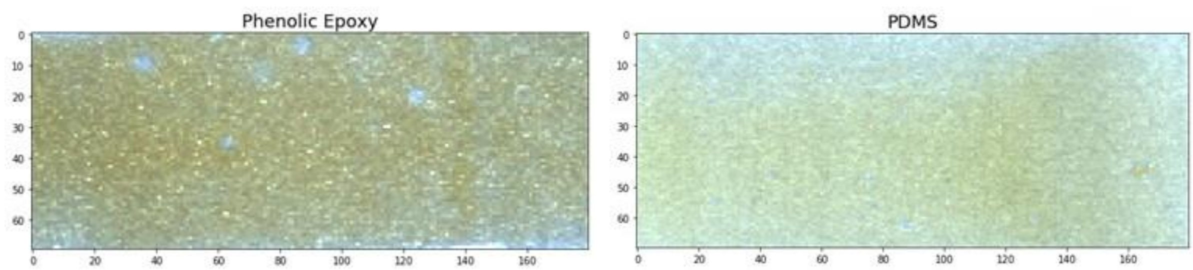


Fig.8: RGB images of panels with drawdowns of different cured binders

The mean spectra of the pixel spectra for each binder were then calculated. Observe that in the study on chemical binders, only the mean spectra of all pixels in each binder were calculated. Unlike in the rust mechanism case, we did not classify individual pixels. The reason for this difference in approaches, is that large coated areas likely will consist of the same paint, the same chemical binder, that is. For rust, we may have small areas in the steel that have been subject to different rust mechanisms. MIC for example, can be very local, covering only a few square centimeters.

3.9 Chemical binders use case – results and conclusions

The mean spectra of Phenolic epoxy and PDMS panels are depicted in Fig.**9** and Fig.**10**, respectively, along with the corresponding peak assignments. Literature references have been used to assign the different peaks observed in the spectra of cured phenolic epoxy and PDMS. *Pramanik et al. (2014)* used near infrared (NIR) spectroscopy to study the curing behavior of different epoxy prepolymers with various amine curing agents, whereas *Cai et al. (2010)* used NIR spectroscopy to characterize different PDMS based silicon rubbers. Both references were used in this work for the assignment of peaks in the mean spectra of phenolic epoxy and PDMS. As observed, the SWIR-HSI camera provides reliable information about the different signature compounds present in cured phenolic epoxy and PDMS.

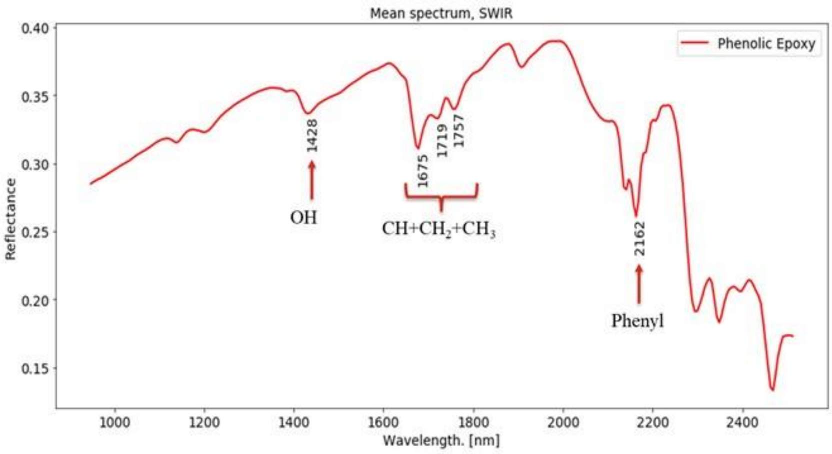


Fig.9: Peak assignments for mean spectrum of Phenolic Epoxy

The conclusion is that the SWIR range of the Hyspex camera can differentiate reliably between the two chemically different coatings, phenolic epoxy and PDMS; respectively. Likely, this result is generalizable to all chemically different coatings and coatings of different age. Existing literature supports this finding in that most of the organic chemical compounds (or functional groups) do show signature absorption bands in the SWIR range of 1 to 2.5 micrometer.

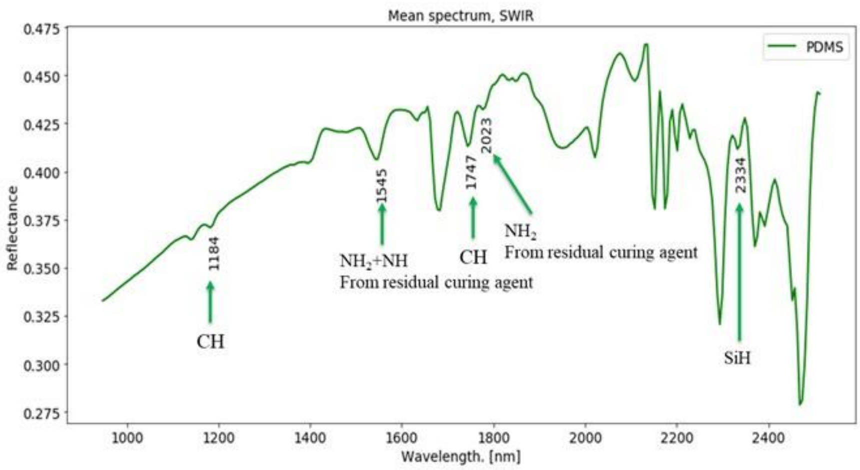


Fig.10: Peak assignments for mean spectrum of PDMS

4. Computer vision for detection of cracks in images

4.1 Introduction

There are three levels of granularity for classifying objects in ordinary images: image-level classification, object detection, and semantic segmentation, respectively, see example in Fig.**11**. In image classification, the whole image is labelled as an object, in this case as "crack" or "no-crack". In object detection, a bounding box is drawn around the object and labelled as "crack". In semantic segmentation, each pixel of the defect in the image is labelled as "crack". Unlabelled pixels are by definition "no-crack" or "background". *Stensrud et al. (2019)* reported results on image classification. This paper reports results on both object detection and semantic segmentation.

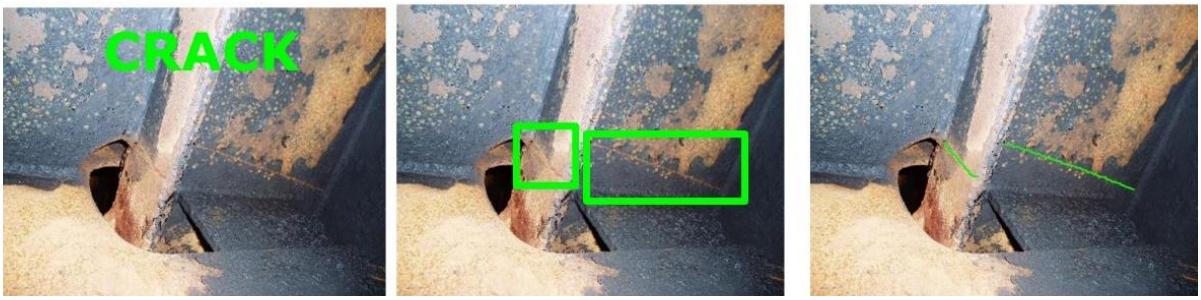


Fig.11: Classification levels of objects in images from left to right: image classification, object detection, semantic segmentation

4.2 Data collection and preparation in general

We collected 1.5 million uncategorised images that have been captured by DNV GL surveyors during inspections to document structural damages on the hull. The damages range from poor coating to cracks, corrosion, indents, and buckling. From this huge dataset of 1.5 million images, 11000 images have currently been identified as containing cracks. A more detailed data collection and sorting procedure is described in *Stensrud et al. (2019)*.

Image-level classification requires the least labelling effort whereas object detection is more labour-intensive, requiring the manual drawing of bounding boxes around the cracks. Semantic segmentation requires the labelling of each pixel that is part of a crack and is therefore the most labour-intensive approach for preparing images. Currently, more than 1000 images of cracks have been labelled on image-level, object-level, and pixel-level, respectively.

4.3 Object detection – Data, Method, Results

In this study, images labelled with bounding boxes were used. The labelling was automated, using the pixel-level labelling as the ground truth, and then fitting ground truth bounding boxes around the pixel-labelled cracks. This approach is illustrated in Fig.**12**. The green pixels have been manually labelled. The blue bounding boxes have then been fitted around continuous areas of green pixels.

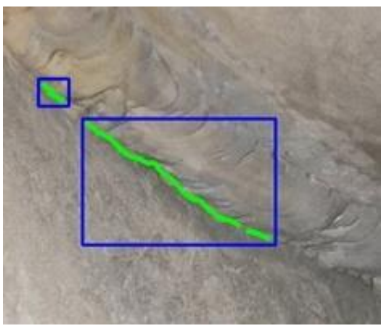


Fig.12: Example of labelling an image with bounding boxes based on the initial pixel-level labelling

The descriptive statistics, object detection, for the labelled images are provided in Table III. The number of true cracks in the test dataset is N=68.

Table III: Object detection – number of images and bounding boxes in training, validation, test data

Description	#images	#bounding boxes
Training data	1025	1876
Validation data	45	86
Test data	49	68

The model used in this study is the Keras, https://keras.io/ implementation of RetinaNet, *Lin et al. (2017)*, a convolutional neural network. The training data were augmented by a factor 15, to increase the size of the training dataset. Several augmentation techniques were used: rotation, flipping, scaling, translation, and shearing.

To evaluate the detection performance, i.e. the number of true and false detections, the basic metrics are True Positives (TP) and False Positives (FP). The decision of whether a detection is a TP or FP depends on the degree of overlap of the ground truth bounding box and the predicted bounding box. The degree of overlap is measured with Intersection over Union (IoU). Different IoU thresholds were evaluated in the study.

Table IV: shows the results of object detection performance. The number of True Positives and False Positives is a function of the IoU threshold. A lower threshold means that we require a smaller overlap between the ground truth bounding box and the predicted bounding box. An IoU threshold of 0.5 is quite common. With this threshold, 44 out of the 68 cracks are detected. With an IoU threshold of 0.2, 60 out of the 68 cracks are detected. As expected, the less overlap is required, the more True Positives are detected. It is also observed that many False Positives are "detected", regardless of the IoU threshold value.

Table IV: Object detection performance

IoU threshold	True Positives	False Positives
0.2	60	1361
0.25	58	1363
0.3	55	1366
0.35	53	1368
0.4	53	1368
0.5	44	1377
0.6	32	1389

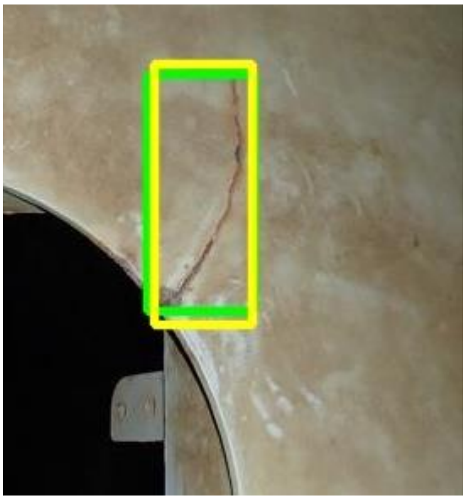


Fig.13: Example of object detection – ground truth (green) vs. predicted (yellow) bounding boxes

4.4 Semantic segmentation – Data, Method, Results

In this study, images labelled at pixel-level were used. Every picture contains at least one crack. The selection of images to label, and the labelling of the cracks, were done by the team that developed the prediction model, which are not subject-matter experts. In addition to labeling the cracks, ambiguous areas were marked to be ignored during training and testing. The same kind of marking was automatically added to the borders of the cracks to make the training easier. The original images vary in resolution, but for training and evaluation they were resized to match the number of pixels in a 640x480 image, while keeping the original image ratio intact. This resolution has seemed to be a good compromise between the level of detail required detection of cracks and the computational cost of training the model. In semantic segmentation of cracks, there is a two-fold class imbalance problem. The first problem is that the crack covers only a small fraction of the image. The second problem is that cracks are rare; in a real inspection setting, most of the time there will be no cracks. In this work, only the first kind of imbalance have been addressed. The training, validation and test sets used are the same images as reported in Table III:. This distribution reflects that the development is still in an early stage where the focus has been on increasing the size of the training data, rather than on the testing. Images used for validation and testing have been handpicked to represent specific challenging scenarios.

The models used for semantic segmentation are different versions of UNET, *Ronneberger et al. (2015)*, implemented in Pytorch. The specific version reported here was a UNET with a ResNeXt50, *Xie et al. (2017)*, decoder-part. During training, various augmentation techniques were randomly applied, such as rotation, scaling, flipping, mirroring, and color adjustments. Due to GPU memory limitations, the model was trained with single images batches. Because of this, the model used instance-normalization, *Ulyanov et al. (2017)*, instead of the more common batch-normalization. Training was done with the Adam optimizer, *Kingma and Ba (2017)*, at a learning rate of 0.003, beta1 = 0.9, beta2 = 0.999 and no weight decay. To evaluate the detection performance, 3 different metrics have been focused on: precision, recall and intersection-over-union (IoU). Precision and recall are good metrics in cases where there is a large class imbalance, and IoU is a common metric for semantic segmentation. These are useful metrics during the development of the model, but we believe that to evaluate the model for a production setting ,it will be advantageous to base performance metrics on reported cracks by the system that will use the model, instead of basing it on pixel-wise performance.

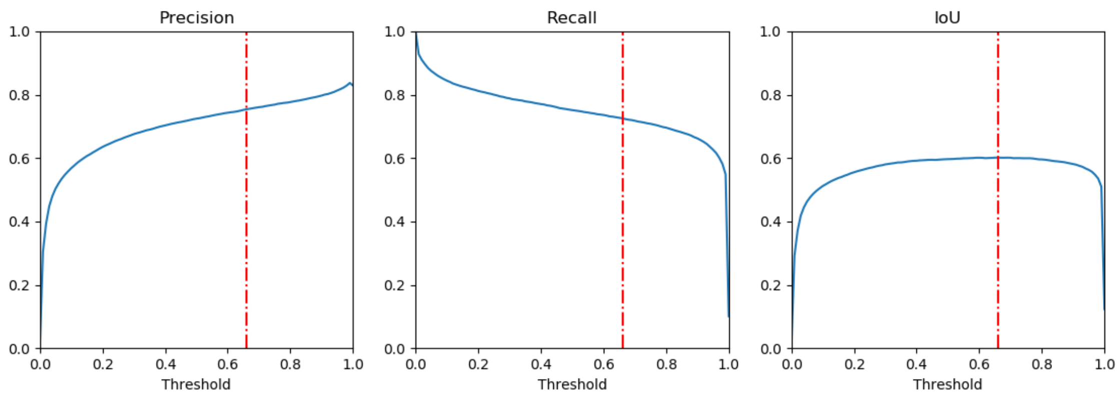


Fig.14: Performance metric of our semantic segmentation model on our validation set. The dashed red line represents the threshold value for crack pixel classification that gives the highest IoU on the validation set. This threshold value is used to evaluate the test set.

Precision, recall, and IoU for the validation set as a function of the threshold value can be seen in Fig.**14**. The threshold value that gave the best results on the validation set, was 0.66. The same metrics at this threshold for both the validation and the test set can be seen in Table V. There is only a small difference between the performance in the validation and test set, so the model appears to have generalized well. When inspecting the errors, we see that the errors come from images that are also hard for untrained humans, and that the model to a large degree has learned to differentiate between cracks and similar structures such as sharp edges and shadows, which were problematic in the start of the project.

Table V: Performance on validation and test set with a threshold of 0.66

	Precision	Recall	IoU
Validation	0.75	0.73	0.60
Test	0.73	0.72	0.55

4.5 Field testing of the crack detector onboard an FPSO

The computer vision, currently a crack detector, was also demonstrated and tested onboard the FPSO. A crack detection algorithm was run in real time on the video stream from the drone, as part of the demo. Cracks in the tank were simulated by bringing paper photos of cracks that were glued to the wall in the cargo tank. The reason for this approach, was that there was no prior information given about any existing cracks in the cargo tank, and there was a need to have some detectable cracks for the demo.

The crack detection algorithm did detect all the cracks, and there were few false positives. Some challenges were observed with regard to specular reflection when the light from the drone was reflected straight back to the camera. This might be a challenge that was particular to this demo as the cracks were on printed paper, and paper is quite reflective.

The current (limited) performance of the algorithm is one challenge. A more important challenge is the business aspect, i.e. how such a crack detector may add value in future, remote inspections. In many cases, it may be difficult to do a detailed inspection in real-time over a remote video link. An alternative, could be to split the process into a data collection step, followed by an off-line step where a more compute-intensive crack detector algorithm is run on the collected video data. In the data collection step, it is then important to ensure that the information gathered is as good as possible for the detailed, off-line evaluation. The value of the real time detector could be to assist in the video data collection process by e.g. guiding drone flight near suspicious areas to collect more data of these areas. The value of the offline crack detector would be to screen the collected video data, and alert the surveyor only when a crack or a suspicious area is detected. This would save the surveyor from doing screening of many hours of video. This approach is contingent on having a detector with very low rate of false negatives (ideally zero), and not too many false positives. In the first case, actual cracks are missed, which is a safety issue. In the second case, a large number of false positives would not save any video screening work for the surveyor.

5. Overall Conclusions

Drone operation: it seems feasible that the expert inspector will be able to operate the drone without the need of a trained drone operator. Furthermore, by using an on-premise 4G router, the inspection can be streamed live over the internet to each stakeholder.

Hyperspectral imaging, rust mechanism: The results suggest that we may be able to distinguish between salt-spray rust and humidity rust. It might therefore be possible detect microbial induced corrosion (MIC) from other rust mechanisms. Furthermore, it may be possible to classify rust mechanisms with a cheaper and smaller camera than the high-end Hyspex camera.

Hyperspectral imaging, chemical binders in coating: The results suggest that hyperspectral cameras in the SWIR spectrum may be able to detect and classify the chemical constituents in paint. However, this preliminary conclusion is based on our observation that the camera can differentiate reliably between the two chemically different coatings, phenolic epoxy and PDMS, respectively.

Computer vision, object detection of cracks: Currently, too many false positives are “detected”. This implies that the detector will not save much work for the surveyor until the number of false detections is significantly reduced. In order to reduce the false positives without increasing true positives, we are investigating how to combine object tracking with object detection to build a more robust and reliable crack detection pipeline.

Computer vision, semantic segmentation for crack detection: the model has to a large degree learned to differentiate between cracks and similar structures such as sharp edges and shadows. However, there is still an issue with false positives. Furthermore, the detector should alert the surveyor only if a crack is detected, and not if a single pixel is detected as a crack. This is a topic of further research.

Acknowledgements

The ADRASSO project is supported by project grant number 282287 from the Norwegian Research Council. The field demonstration and testing were facilitated by Dag Børre Lillestøl and Ulf-Gunnar Mikalsen, both DNV GL. Marie Henriksen, NTNU, collected some of the NTNU camera data. Milica Orlandic and Tor Arne Johansen, both NTNU, coordinated the development of the NTNU camera. João Fortuna interfaced the camera with the DUNE environment. Anders Skilbred, Jotun, assisted in preparing panels. Øyvind Smogeli, DNV GL, reviewed the paper. Trond Løke, NEO, thoroughly reviewed the hyperspectral sections. Altera hosted the field test of the drone onboard the FPSO, motivated by digitalization and remote operation, which may add value to the business going forward.

References

CAI, D.; NEYER, A.; KUCKUK, R.H.; HEISE, M. (2010), *Raman, mid-infrared, near-infrared and ultraviolet–visible spectroscopy of PDMS silicone rubber for characterization of polymer optical waveguide materials*, J. Molecular Structure 976, pp.274-281

HENRIKSEN, M.B.; GARRETT, J.L.; PRENTICE, E.F.; STAHL, A.; JOHANSEN, T.A.; SIGERNES, F. (2019), *Real-time Corrections for a Low-cost Hyperspectral Instrument*, Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS)

KINGMA, D.P.; BA, J. (2017), *Adam: A Method for Stochastic Optimization*, https://arxiv.org/abs/1412.6980

LIN, T.Y.; GOYAL, P.; GIRSHICK, R.; HE, K.; DOLLAR, P. (2017), *Focal Loss for Dense Object Detection*, IEEE Int. Conf. Computer Vision (ICCV), http://openaccess.thecvf.com/content_ICCV_2017/papers/Lin_Focal_Loss_for_ICCV_2017_paper.pdf

PRAMANIK, M.; FOWLER, E.W.; RAWLINS, J.W. (2014), *Cure kinetics of several epoxy-amine systems at ambient and high temperatures*, J. Coat. Technol. Res. 11(2), pp.143–157

RONNEBERGER, O., FISCHER, P., T. BROX (2015) *U-Net: Convolutional Networks for Biomedical Image Segmentation*, https://arxiv.org/abs/1505.04597.

SIGERNES, F.; SYRJÄSUO, M.; STORVOLD, R.; FORTUNA, J.; GRØTTE, M.E.; JOHANSEN, T.A. (2018), *Do it yourself hyperspectral imager for handheld to airborne operations*, Optics Express

STENSRUD, E.; SKRAMSTAD, T.; CABOS, C.; HAMRE, G.; KLAUSEN, K.; RAEISSI, B. (2019), *Automating inspections of cargo and ballast tanks using drones*, COMPIT Conf., Tullamore

ULYANOV, D.; VEDALDI, A.; LEMPITSKY, V. (2017), *Instance Normalization: The Missing Ingredient for Fast Stylization*, https://arxiv.org/abs/1607.08022

VITALE, R.; ZHYROVA, A.; FORTUNA, J.F.; DE NOORD, O.E.; FERRER, A.; MARTENS, H. (2017), *On-The-Fly Processing of continuous high-dimensional data streams*, Chemometrics and Intelligent Laboratory Systems 161, pp.118-129

XIE, S.; GIRSHICK, R.; DOLLÁR, P.; TU, Z.; HE, K. (2017), *Aggregated Residual Transformations for Deep Neural Networks*, https://arxiv.org/abs/1611.05431

Rapid One-of-a-Kind Project Planning Suite

Thomas Koch, Atlantec Enterprise Solutions GmbH, Hamburg/Germany, thomas.koch@atlantec-es.com
Konstantin Kreutzer, Atlantec Enterprise Solutions GmbH, Hamburg/Germany, konstantin.kreutzer@atlantec-es.com
Andreas Roppelt, Atlantec Enterprise Solutions GmbH, Hamburg/Germany, andreas.roppelt@atlantec-es.com

Abstract

To enable access to advanced technologies allowing rapid and sufficiently detailed one-of-a-kind project planning in the early project stage an affordable software suite has been developed. Product- and process-related data are directly derived by reconstruction of 3D models from point-clouds and using requirement engineering techniques based on advanced text analysis. Full-scale project scheduling and cost estimation is supported, including optimisation and 4D-animation targeting configurable goals (e.g. shortest time, best utilization, and minimized cost). An underlying data management system allows integration of generated data to an integrated simulation component to identify bottlenecks and minimise production risk, including what-if analysis.

1. Introduction and Background

In shipbuilding and offshore projects, which are characterised by their one-of-a-kind, made to custom order nature, reliable forecasts of project duration and estimation of costs at the bidding or early project planning stages prove to be a difficult trade-off between precision and effort. Since deadlines are typically very tight, it is not possible to achieve high precision within reasonable time unless a broad and well-maintained empirical data collection can be utilized. At the same time there are very limited possibilities to apply repetitive manufacturing principles for built-to-order products, since historical/empirical data does either not exist or is only of limited value and cannot be applied easily without major adaptation effort.

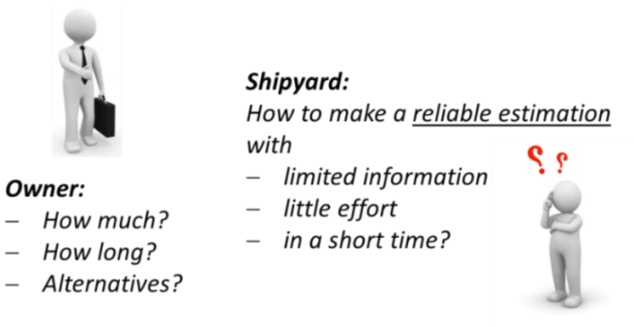


Fig.1: Inquiry scenario

Fig.1 illustrates the typical request-for-quotation situation in the shipbuilding industry, where the owner/customer provides only limited information of widely varying level of detail and scope (such as general arrangement drawings or textual specifications) to a shipyard to inquire the delivery dates and cost to build, retrofit or repair a vessel. For a shipyard or designer to provide a workable proposal is a demanding task. Considering several fundamental alternatives in more detail will result in a substantial effort for analysis, system engineering, estimation and planning. As a consequence, due to the lack of tailored tools to be applied in the maritime industry, shipyards commonly do “quick” estimations using empirically assessed or derived design solutions, schedule and cost, including appropriate empirical buffers during the negotiation and early design stages. In most cases there is insufficient time to investigate alternatives, as ship operators expect a quick response to minimize either time till service availability of the new vessel or disruptions of existing vessel operation in retrofitting or repair scenarios.

In consequence, the current common approach easily results in over- or under-estimated costs and/or durations and unmitigated risks. Since shipyards are nevertheless trying to manage potential risks (e.g. unplanned delays, unforeseen cost and technical or logistical problems) overestimated buffers can therefore result in losing the bid. On the other hand, in case costs and durations are underestimated, the project may be quickly in acute danger of becoming an economical failure due to delays, cost overruns and problems in fulfilling the specifications. Finally, although the common approach to planning provides only rough estimations, the effort for doing this can still be very high and thus costly, since up to 40% of the time may be spent on searching for information for documents and reports to be provided to the owner – which are mainly created manually, *Morais (2019)*. In summary, there is a considerable potential in saving by pushing forward the digital transformation in the bidding and early design and planning stages.

2. Improving Early Production Planning

In this paper an approach is described which has been developed during the SHIPLYS project resulting in a toolset named PPT (Production Planning Tool[kit]), *Bharadwadj (2017)*.
PPT represents a tool suite combining several technologies as shown in Fig.2 allowing **reliable estimation** of cost and schedule within short time by developing and using digital twin data during the early bidding project stage.

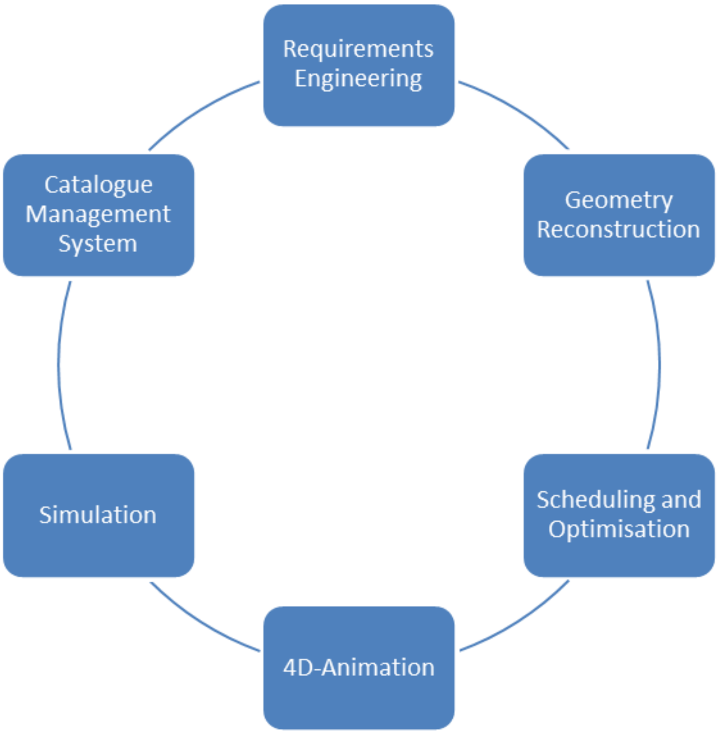


Fig.2: Technology Collection in PPT

Furthermore, cost and project duration can be reduced using the included optimisation functionality and various alternatives can be analysed to assist in preparation of more competitive offers.

The technologies applied have different grades of novelty. For instance, geometry reconstruction based on scan data represents a relatively new technology still undergoing development helping to rapidly create a digital twin in the shape of a 3D product model of the vessel itself. A method such as discrete event simulation (DES, *Banks (2013)*), which essentially provides and utilizes a digital twin of the production area (e.g. simulation model of the whole shipyard), are today used for daily business applications in other industries (e.g. automotive industry) where repetitive manufacturing principles can be applied. In shipbuilding only a few shipyards have started to use production simulation in limited ways to manage complex systems and related risks. However, it is commonly used only periodically to support investment decisions concerning new or refurbished, more advanced machinery or in order to improve the production flow. In particular, the effort of data acquisition and preparation for feeding into simulation model is considered as a major obstacle for everyday use. As a result, effective data acquisition and generation is a key factor in making simulation a feasible tool.

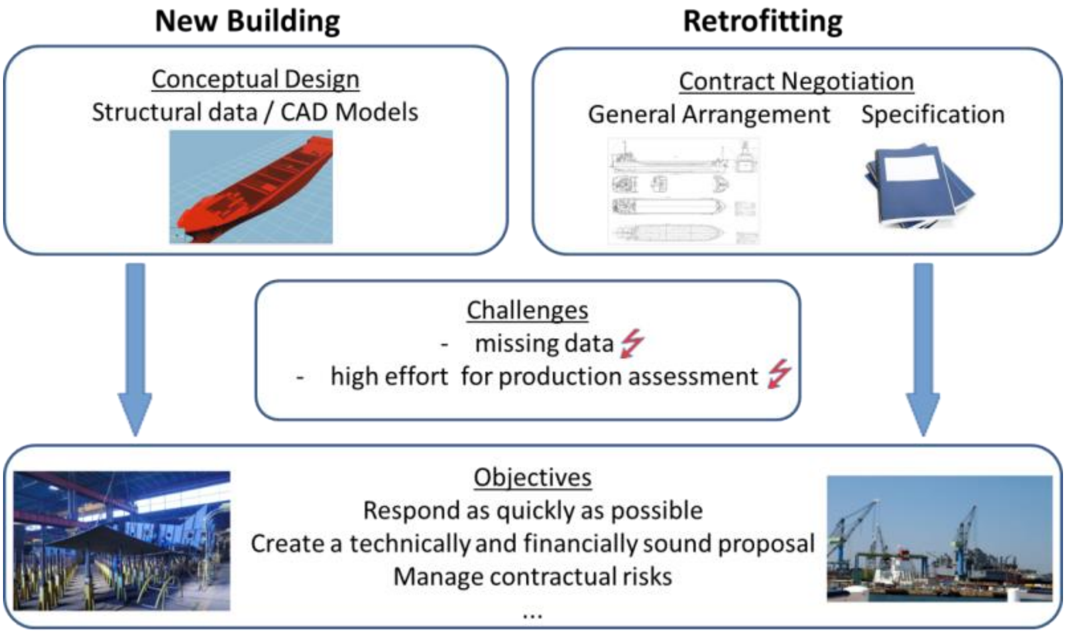


Fig.3: Newbuilding vs. Retrofitting

Since PPT has been built from a collection of specialized components, it has been an essential fundamental design decision to link all the relevant technologies tightly together ensuring fast and effortless data interchange by means of a rich and extendable data model implementation which has been developed and used over many years in shipbuilding industry applications.

The data model supports all stages of project execution and is as part of a commercially available data management platform, *AES (2018)*.

Furthermore, integrated rapid data generation functionality is another key factor allowing overcoming the mentioned hurdle and to perform optimization and to use the applied simulation technology at the very early planning stage. Tests in the industrial environment at shipyards during the SHIPLYS project have shown that the provided solution addresses these key issues as reported by end-users.

While newbuilding projects regularly allow for some more extended planning phase, repair or retrofitting scenarios will often be even more demanding since only a – possibly sketchy – specification containing task or problem descriptions is provided by the owner/operator and therefore shipyards need to estimate overall project duration and costs very rapidly. Fig.3 shows the similarities and differences between newbuilding and repair/retrofitting cases.

3. Technologies provided by PPT

PPT tools follow a unified approach for both new-building and retrofitting applications allowing to rapidly produce more reliable predictions regarding the overall project duration and costs during the early bidding stage in the shipbuilding industry resulting in more competitive offers while mitigating risks at the same time.

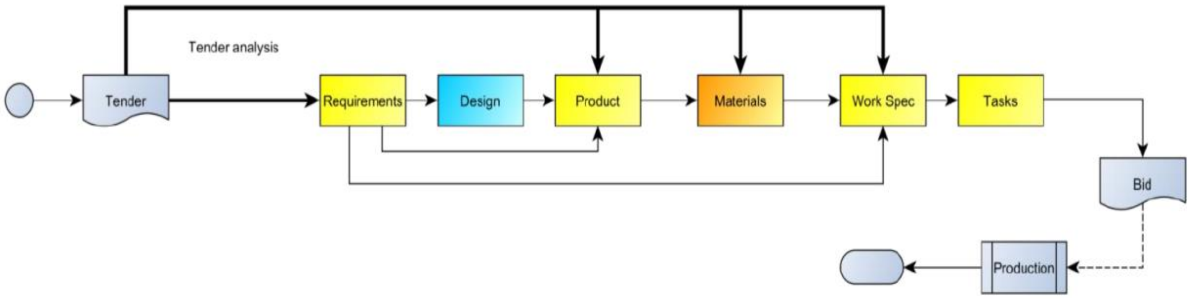


Fig.4: Unified Workflow

Fig.4 shows the general unified workflow beginning with a tender document provided by the owner/operator. Depending on the actual type and details of project (newbuilding, retrofitting or repair), some stages shown may be optional. The different stages are supported by the technologies provided by PPT whose interaction is ensured by appropriate data model presented as next. For development of the functional set the identification of similarities and differences between the different use cases was an important step to define a powerful combination of capabilities.

3.1. Data Management System

Planning and scheduling tools used in one-of-kind, on-demand projects in of today use have often quite limited capabilities to interact with engineering data. Therefore, a considerable effort is required to provide linkage between engineering data and planning. During early project stages such linkage is reduced to general estimates due to the lack of detailed information.

To facilitate advanced data integration and management features for the maritime industry, a rich extendable Business Data Model has been developed for many years covering most aspects of maritime industry requirements. It is based on various ISO standards and other industry standards (such as: ISO 10303-215, -216, -218, -227, ISO 15384, WfMC, X500) but extends beyond their scope and contains hundreds of different entity types considering e.g. product data (ship), production environment (shipyard) as well as tasks, supplies and organizational aspects, Fig.5.

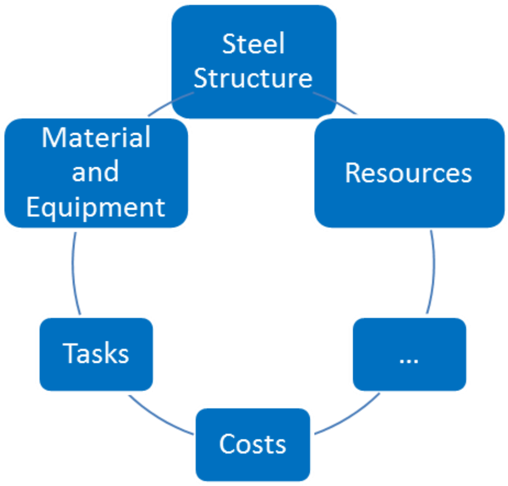


Fig.5: Data Model Scope

Furthermore, design related data can be versioned which helps bridging the gap to commonly used CAD systems lacking this capability while maintaining a history of changes. The related server infrastructure allows managing the information based on this data model and furthermore simplifies connecting to the existing system landscape of a shipyard by means of so-called adapters enabling to rapidly retrieve and receive any amount of data from different applications such as ERP- and/or CAD-systems etc.

In this way all required input data as well as the output data generated by PPT (described below) can be managed and interchanged with the existing systems.

3.2. Requirements Engineering

Requirements engineering is a well-established methodology originally evolving from very large scale projects in space technology, aircraft development and, to some extent, large scale information technology development projects. The benefit of using such systems is to register, monitor and evaluate requirements, thus being able to verify compliance and fulfilment. Unfortunately, due to the complexity and cost of introducing and using this technology, its application in shipbuilding has been quite limited to areas like large-scale offshore projects or in the defence industry. For commercial shipbuilding there exist at least the following obstacles:

- Cost of tools: requirements engineering tools are found in a high priced range both in licensing and operating cost
- Ease of use: while large-scale projects can accommodate the personnel effort to operate these systems, this seems prohibitive for most shipbuilding operations and in particular SMEs. This is related to the aspect that the level of detail and documentation is considered too high for smaller business types.

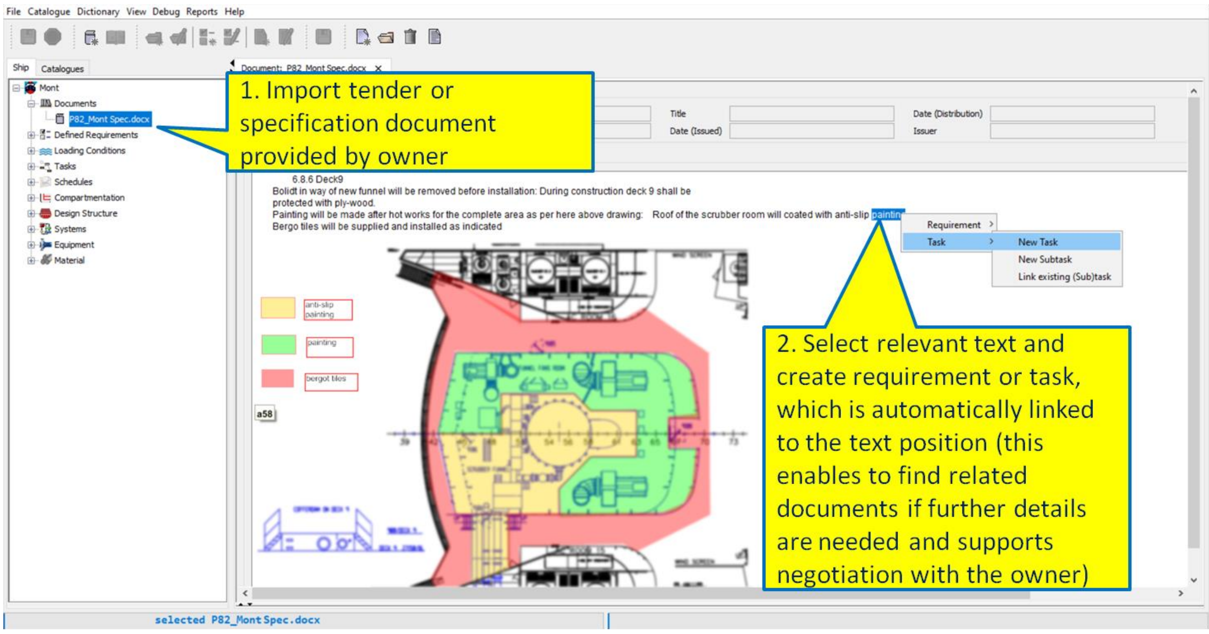


Fig.6: Requirement and Task identification

For PPT, two prototype editors have been developed and tested which are based on the combination of text analysis features with engineering-oriented classification of product components, tasks and design goals, to enable planners and engineers to quickly process technical specification documents while deriving a concrete data representation suitable for engineering analysis, schedule derivation etc. It is deliberately not intended to provide a fully automated process, since many decisions need to be made by creatively acting users during the planning and early design process steps. Nevertheless, this data substantially helps during data generation in the steps to follow, such as schedule derivation and cost estimation as well as during later steps for quality control and project validation. The editor components have been integrated into the PPT tool suite with the specific intention to be used in the early stages of the project execution. Tender or specification documents are used as input since these kinds of documents are typically provided by the vessel owner/operator as shown in Fig.6.

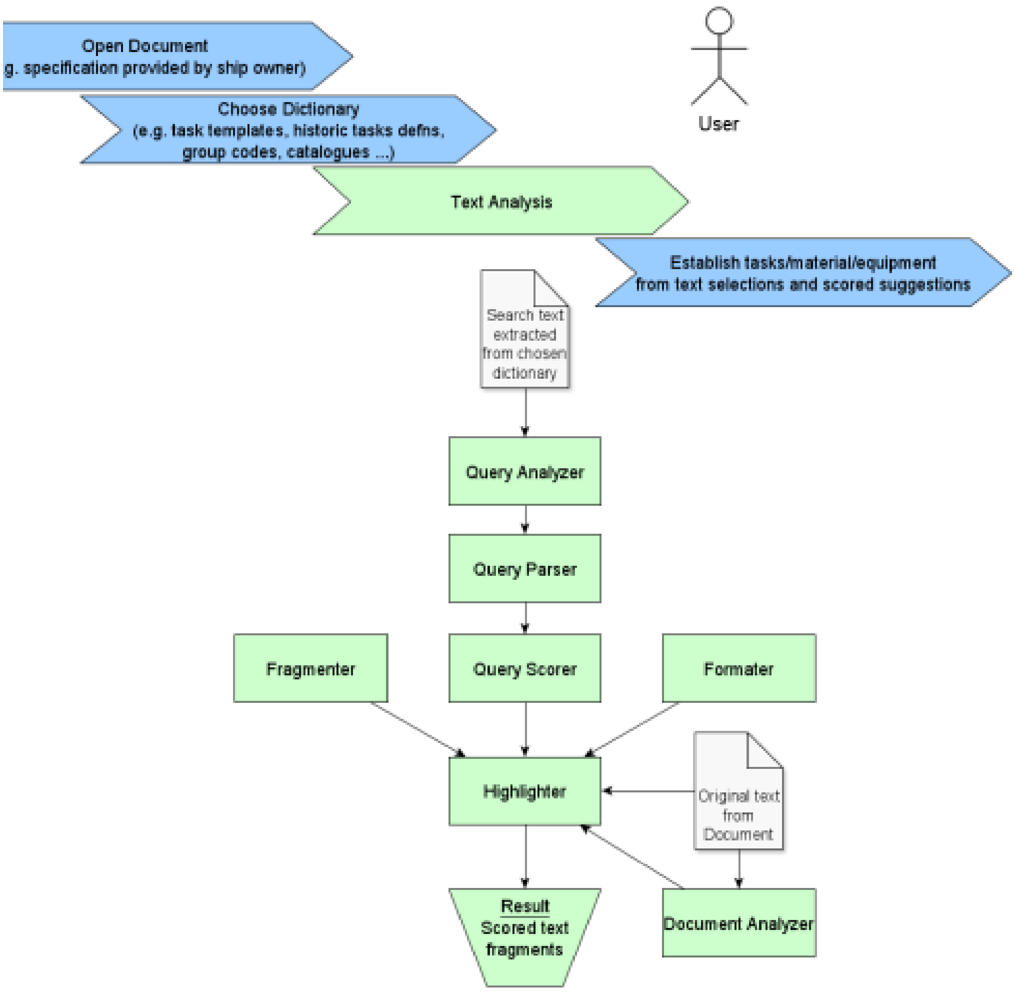


Fig.7: Text Analysis for RE

Documents are imported and requirements defined/identified, automatically linking them to the relevant text positions and submitting them to a text analysis engine, *McCandless (2010)*, with the aim of supporting the user in identifying the relevant statements and definitions in the document. Furthermore, scoring algorithms are applied to match the text portions against various dictionaries. The complete flow of text processing is shown in Fig.7. This concept is not intended to fully automate the process but to assist the users in their activities for extracting the relevant information to create requirement definitions and task descriptions.

Once the requirements are created, they are represented in machine-process-able format and can be used to evaluate the evolving design in respect to fulfilling the requirements. Wherever feasible, tasks descriptions can be directly derived, which is especially useful and relevant for the repair/ retrofitting cases, where tenders are often based on short specifications of work items or issues to be resolved.

To progress efficiently with the definition of requirements and tasks, PPT provides a catalogue management system to store general data such as material and equipment items, standards like SFI grouping codes or task templates, Fig.8. This predefined, but fully configurable information can be easily applied for a specific project and revised when required and constitutes an important feature for enhancing the task and schedule information based on engineering details. It also has the effect of normalizing data in such a way that duplications and deviations from yard standard practice are minimized. Since these catalogues will be configured for individual shipyards based on their organisation and work standards, it will also substantially simplify the validation of requirements and tasks against the individual shipyard standards.

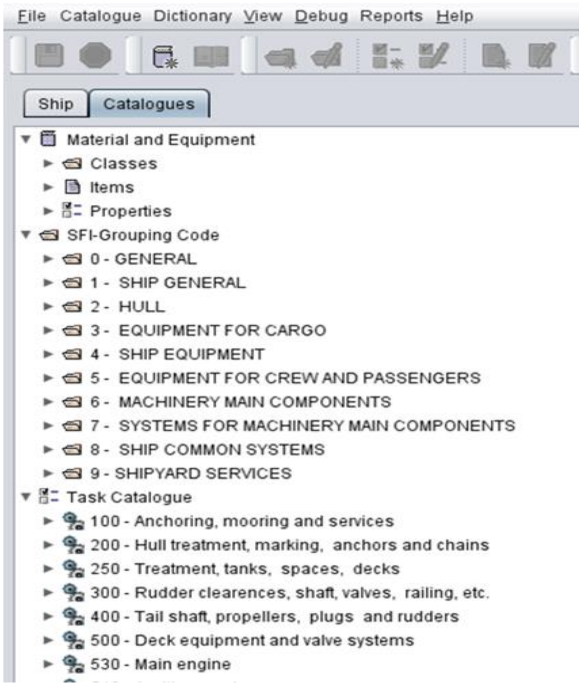


Fig.8: Catalogue

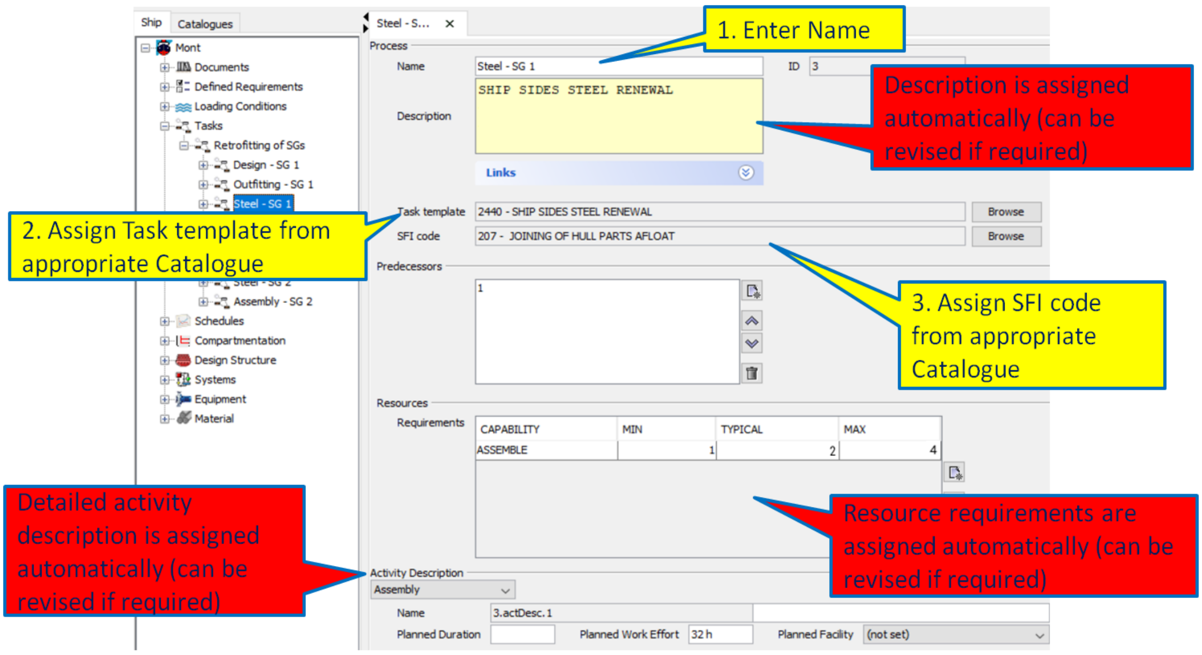


Fig.9: Rapid Task definition

Fig.9 shows how the functions for requirement and task definition can be applied using the definition of a task as example. Furthermore, equipment and material information can be identified or extracted in a similar fashion to define them ad hoc using the mentioned catalogue management system.

3.3. Geometry Reconstruction

This topic is particularly relevant in retrofitting and repair scenarios. Vessels that have been operating for years are no longer in the same shape as when they were originally designed or built. Numerous modifications have usually been applied to the vessel since its initial delivery. Therefore, reliable CAD model information representing the actual on board situation is lacking (often due to IP right restrictions of the original builder or non-existence of such a model).

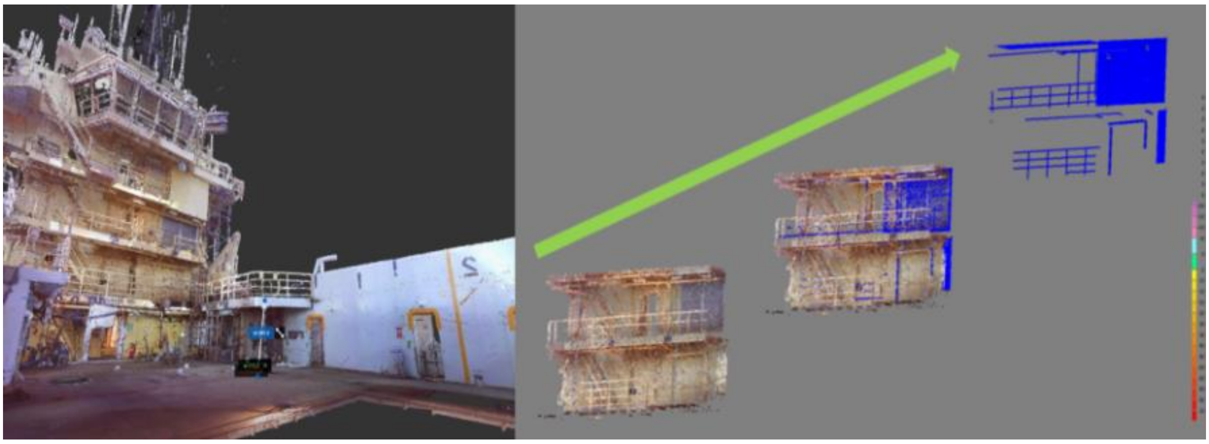


Fig.10: Geometry Reconstruction

This situation forces shipyards regularly to assess the actual situation on board a vessel with a lot of effort. Unfortunately, until today such investigations do not provide an attributed geometry model of the as-is-state of the vessel, but rather a compilation of photographs, measurements, sketches and, to a growing degree, photogrammetry or point-cloud data acquired from laser scanning. If such scanned data is available, attempts are being made to reconstruct geometry models by means of interactively modelled CAD models using the point cloud data as reference. This "almost manual" reconstruction is labour intensive and has proven to be prohibitively expensive for general application so far, so existing state-of-the-art CAD tools cannot be applied efficiently.

Software-assisted semi-automatic geometry reconstruction based on such point cloud data, which can be created by various scanning tools, is an evolving innovative approach addressing the need to generate 3D models faster and, most importantly, with less effort. Tools focusing on specific requirements of the shipbuilding domain are missing not at least because requirements are very specific in this area. In the SHIPLYS project, promising tests were carried out using a prototype of reconstruction software specifically targeting the shipbuilding industry needs, which has been integrated into the PPT suite. This makes it possible to reconstruct "real-world" geometry and to merge it with newly designed partial models, normally required in case of extensive modifications such as lengthening, scrubber installation etc.

Another important benefit is gained from this is the option of establishing an interim product structure as it reflects the base information for production planning and which is essential for defining the work breakdown structure (WBS). A 3D model of the interim product structure (representing the ship or relevant parts thereof) is therefore a key input to determine details like job sequences, task dependencies (e.g. a painting task depends on the surface to be painted), work content, required material, available space, access paths etc. as a prerequisite for accurate production planning.

Fig.10 shows a sample of the reconstruction procedure of a relevant portion of a point cloud. By selecting only few points within the cloud the algorithm automatically identifies points belonging together and forming a part such as a wall, deck, profile, plate, pipe, flange, valve etc. In doing so, time and cost compared to manual creation of a 3D model are substantially reduced.

3.4. Scheduling and Optimisation

Having defined tasks using PPT functionality, a schedule can be automatically derived afterwards considering the predecessor/successor relations and duration of tasks (Fig.11). Most importantly, this tool relates the scheduling information to the underlying engineering and planning data which goes well beyond the capabilities of common scheduling tools.

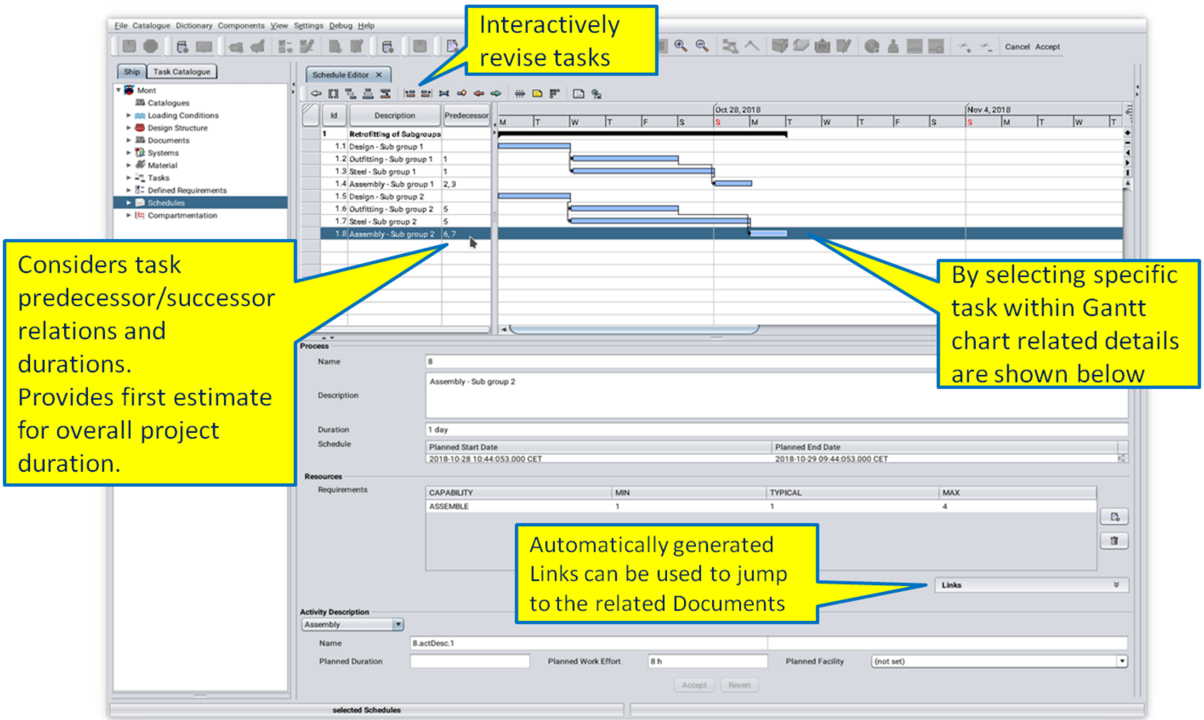


Fig.11: Schedule Editor

Due to the lack of time and the high effort spent in data acquisition the optimisation of project duration, costs, resource utilization etc. is often neglected and therefore considerable potential for improvement is wasted. This is particularly important during the early bid preparation stage where data is often unreliable or missing. Furthermore, the optimisation component has to be fully integrated into the existing IT system in such a way that all required data can be retrieved without much effort for data selection or conversion.

Due to the rapid data generation functionality and the possibility to connect to various data sources, the optimisation component can be easily supplied with the relevant input data such as tasks and resource capacities. It then provides optimized problem solutions in short time, considering a variety and combination of goals as selected by the user (e.g. shortest time, limited or levelled resources, best utilization, lowest cost, etc.), see Fig.12.

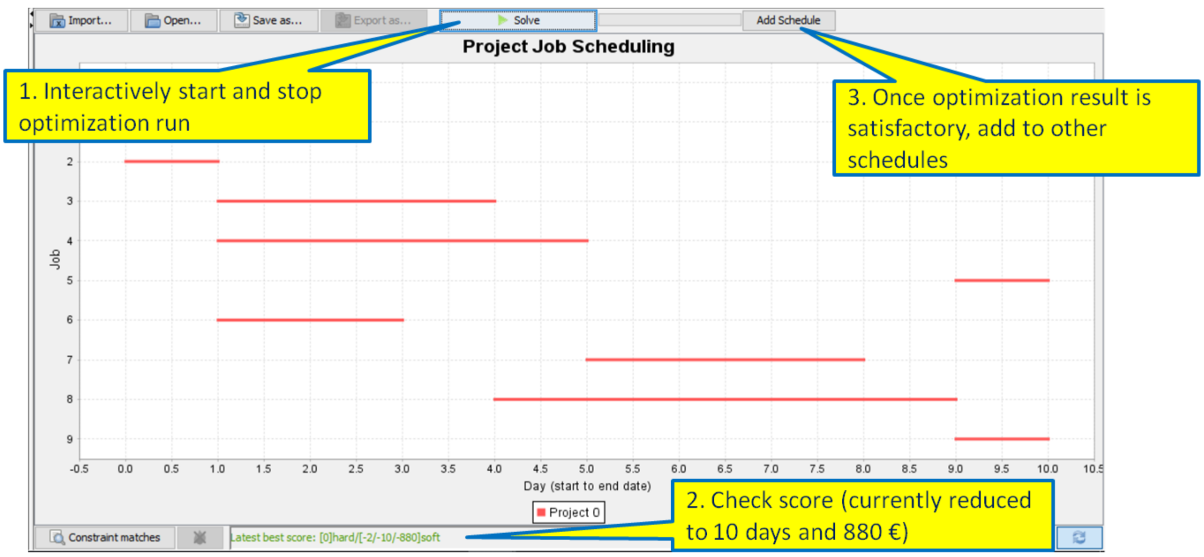


Fig.12: Schedule Optimisation Editor

The related project duration and resource cost can be used to prepare a competitive offer. Thanks to the report generation functionality available in PPT the related bid can be created automatically resulting in further time savings.

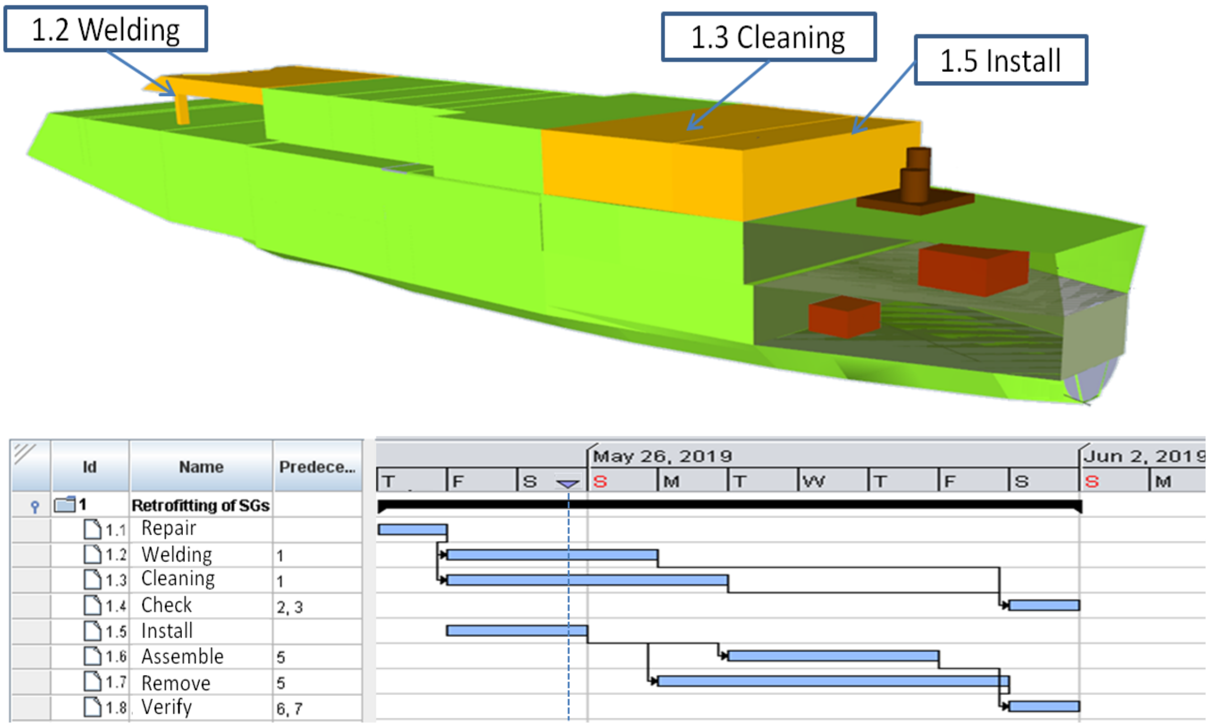


Fig.13: 4D-Animation

3.5. 4D-Animation

As there is a tight interlinkage between schedule, requirements and engineering, it is useful to provide 4D-Animation capabilities allowing navigating through a fully attributed 3D model of the vessel as well as along the project timeline. This can be used to support the internal planning and discussions/ reporting/negotiations with the vessel operator/owner. In the viewer prototype, shown in Fig.13, by shifting a “time” slider within the schedule, the appropriate tasks are visualized in the 3D pointing to the related ship components. 4D-Animation can be applied for both new building and retrofitting projects.

3.6. Production Simulation

Discrete event simulation represents an approach for managing the complexity related to the production planning in shipbuilding and provides more realistic estimations for the common KPIs (such as duration, resource utilization etc.). To use simulation, a considerable amount of data is required to reflect reality at a sufficiently high level of detail as a prerequisite obtain realistic results. Besides production facilities and resources, which in most of cases are defined only once as an initial set-up activity, the data describing the product itself is needed. For this reason, simulation is used in repetitive manufacturing environments since the product does not change too frequently. In a one-of-a-kind industry setting such as the maritime sector each ship is to a large extent unique and product data is not complete especially during the early project stages. Therefore, the application of existing simulation solutions is very labour intensive. Due to lack of sophisticated data-generation functions only few shipyards are applying simulation in limited ways. Their focus is on later stages of the project execution by applying tailored interfaces to the existing current system landscape. Fig.14 shows the effort distribution related to simulation-based analysis where data collection represents the most time-consuming part, *Hübler (2017)*.

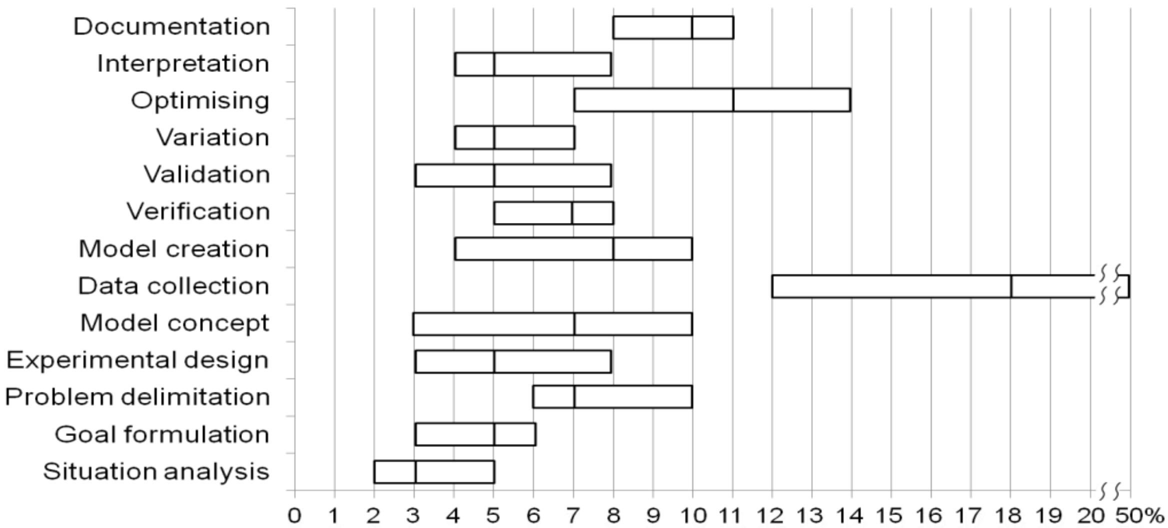


Fig.14: Distribution of effort for simulation projects

The functions included in PPT as described above can help to overcome this hurdle such that rapidly generated product data, task and resource definitions can be used by the simulation component in order to provide a more detailed analysis. This is used to locate bottlenecks, investigate technical aspects of production processes (e.g. evaluation of new production technologies or processes), and to minimize production risks by carrying out a what-if analysis. Furthermore, LCC/LCCA relevant data such as environmental footprint and consumables of production facilities and related cost can be determined in a quite detailed way, as required. In case an optimized schedule has been generated before, the resulting optimal task and resource allocation can also be considered during the simulation run by higher prioritisation of tasks to be carried out first. In Fig.15 the main steps required for simulation runs are shown. The second phase to establish the product model has to be carried out for each project for one-of-kind products and represents the most time-consuming activity when operating without using the rapid definition functions of PPT.

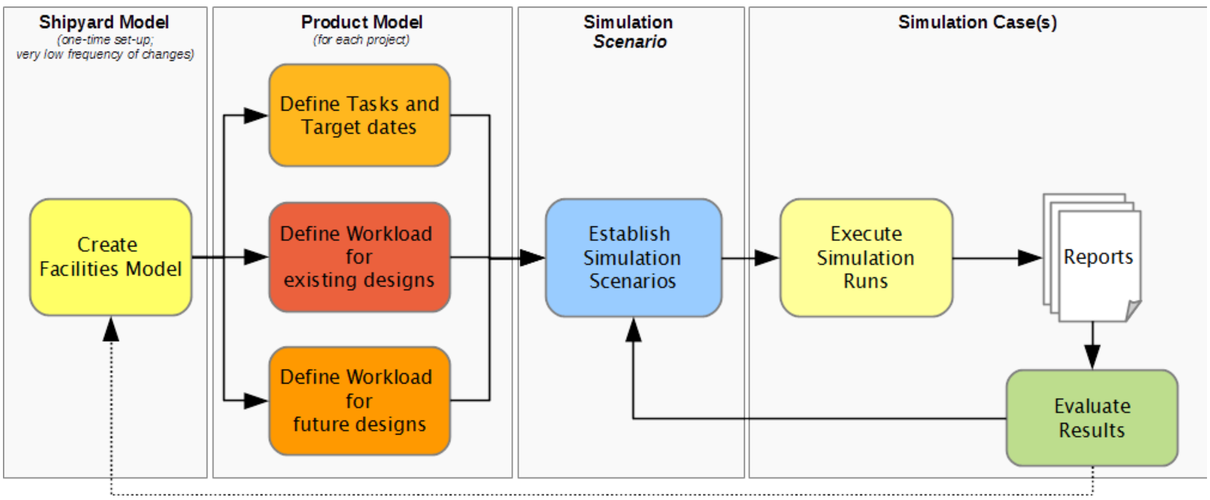


Fig.15: Simulation Procedure

A component for facility modelling has also been integrated into PPT as part of the simulation functions to support the definition of various types of facilities of a production site. This can then be directly used by the simulation component which aids the first phase of setting up a shipyard (asset) model as depicted in Fig.15. A wide range of production equipment types and transportation means commonly found in ship production can be instantiated in such a model.

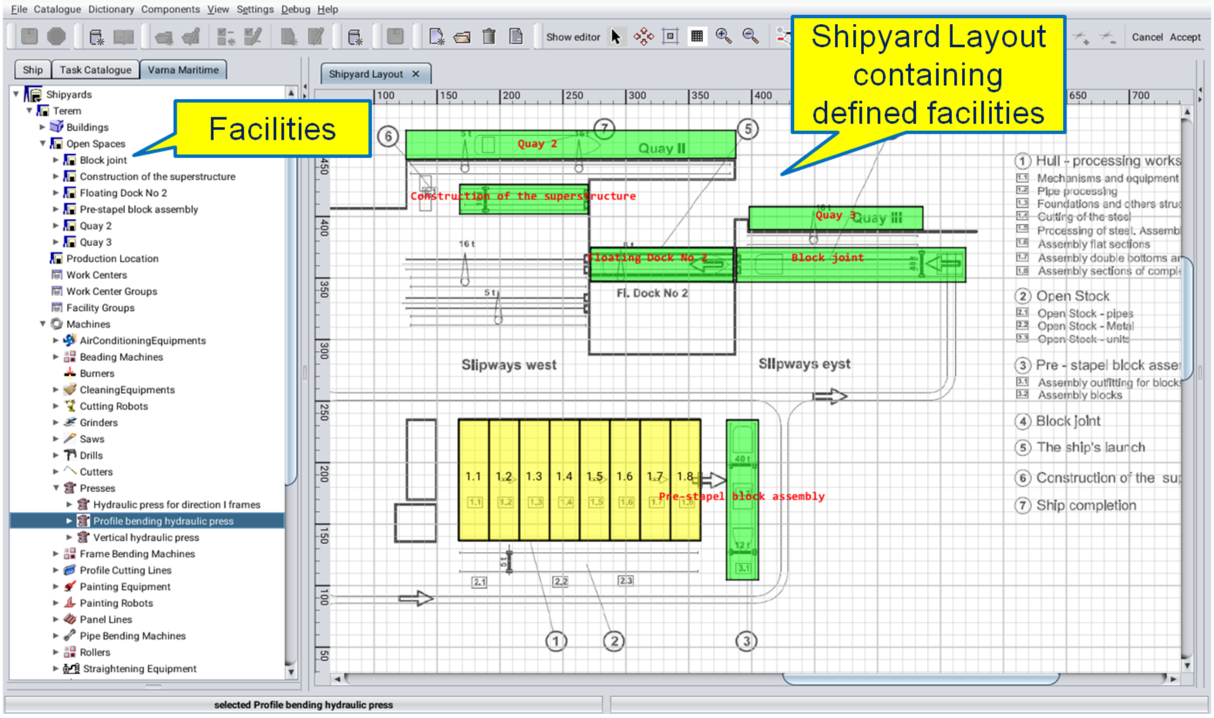


Fig.16: Facility modelling for a shipyard

In Fig.16, an example of a shipyard model is shown allowing further investigations by the use of simulation engine. Since the effort of setting up the shipyard model does not have to be repeated for each project, it is an investment well spent to provide a reasonably detailed digital twin of the shipyard's production environment.

Nevertheless modifications of such a model can be provided at any time, e.g. include new equipment, modifications of production processes etc. Since the activation of such items can be controlled during simulation model configuration and simulation runs, respectively, it is possible to investigate alternative production methods in a straightforward way.

Technical "hot-spot"/risk identification

The simulation component enables the analysis of the various technical hot-spots and risks by investigating what-if scenarios, e.g.:

- Ship design changes
- Use of special facilities
- Essential logistics operations
- Alternatives of production sequences
- External impacts such as weather conditions, business disruptions, equipment failure etc.

After a simulation run is finished, various generated reports can be used to detect possible bottlenecks and underutilization of resources. Consequently, precautionary measures can be taken to minimize production risks and keep deadlines.

4. Conclusions and Outlook

PPT represents a software suite using a unique combination of advanced tools to enable rapid and sufficiently detailed one-of-a-kind production planning applicable to the early project stage based on limited information. The suite can be used for newbuilding as well as repair/retrofitting projects to derive more detailed and reliable data regarding cost and schedule instead of experience-based estimations commonly used nowadays especially in SME shipyards. Therefore, the following benefits for the shipyards seem possible:

- Faster and more reliable evaluation and preparation of bids,
- Winning more contracts by reducing the response time during the bidding stage and by being able to provide more competitive offers by reducing the overall project duration and related costs,
- Creating more reliable production plans will result in improved adherence to delivery dates and therefore increase the customer satisfaction and loyalty,
- Reduction of planning effort will reduce the related costs, whereby the reduction of risks will avoid large cost deviations.

Future development should focus on:

- Further automation of the geometry reconstruction process based on point clouds,
- Extension of functionality for production control allowing reacting in an optimal way with respect to unforeseen changes by connecting optimization and simulation tools to the actual operating data.

Acknowledgements

This paper presents work done in the project "Ship Lifecycle Software Solutions" (SHIPLYS) that has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 690770. It reflects only the author’s views and European Union is not liable for any use that may be made of the information contained therein. The authors also wish to thank the partners of the SHIPLYS Project for their input and support.

References

AES (2018), Topgallant Information Server, http://www.atlantec-es.com/topgallant-product-is.html

BANKS, J.; CARSON, J.S.; NELSON, B.L. (2013), *Discrete-event system simulation*, 5th ed., Pearson Education

BHARADWADJ, U. (2017), *Ship Lifecycle Software Solutions (SHIPLYS) – an overview of the first phase of development and challenges addressed*, Annual Conf. Int. Maritime Association of the Mediterranean (IMAM), Lisbon

HÜBLER, M.; NARAYANAN, D.; MÜLLER, M. (2017), *Efficient retrofitting of vessels by using simulation tools and reverse engineering technologies*, International Shipbuilding Progress 63, pp. 109-136

ISO (2003), *ISO 10303-216: Industrial automation systems and integration - Product data representation and exchange - Part 216: Application protocol: Ship Moulded Forms*, ISO, Geneva

ISO (2004), *ISO 10303-215: Industrial automation systems and integration - Product data representation and exchange - Part 215: Application protocol: Ship Arrangement*, ISO, Geneva

ISO (2004a), *ISO 10303-218: Industrial automation systems and integration - Product data representation and exchange - Part 218: Application protocol: Ship Structures*, ISO, Geneva

ISO (2005), *ISO 10303-227: Industrial automation systems and integration - Product data representation and exchange - Part 227: Application protocol: Plant spatial configuration*, ISO, Geneva

McCANDLESS, M; HATCHER, E; GOSPODNETIĆ, O. (2010): *Lucene in Action*, Manning Publ.

MORAIS, D. (2019), *Is Shipbuilding Ready for a Digital Transformation?*, https://www.ssi-corporate.com/blog-waveform/is-shipbuilding-ready-for-a-digital-transformation/

Augmenting OpenBridge: An Open User Interface Architecture for Augmented Reality Applications on Ship Bridges

Kjetil Nordby, Etienne Gernez, Synne Frydenberg, Jon Olav Eikenes,
The Oslo School of Architecture and Design, Oslo/Norway, kjetil.nordby@aho.no

Abstract

Augmented reality (AR) technologies support navigators by overlaying the perceived world with virtual information collected from the ship bridge systems. However, the variety of operational scenarios, types of ships and bridge equipment from different vendors requires an integration system that enables multiple maritime applications to employ AR as a shared platform. We address the lack of such a system by proposing a user interface (UI) architecture that describes how AR can function as an open, shared platform across different bridge systems by supporting the integration of generic maritime applications in AR.

1. Introduction

Head-mounted augmented reality (AR) technologies can augment the perceived world by overlaying it with digital content. Recent developments in AR technologies, particularly the introduction of Microsoft HoloLens, *Microsoft (2020)*, have increased the number of potential applications in the market. Industrial sectors, such as aerospace, automotive and manufacturing, have already shown extensive use of the technologies, *Capgemini Research Institute (2018)*. In our user-centred exploration of AR technologies and applications, we discovered a variety of potential use scenarios for AR in the maritime domain, *Frydenberg et al. (2018a, 2018b)*. In addition, we reviewed some initial experimental uses of AR, *Nordby et al. (2020)*, including support for ship bridge crew during navigation and operation, *Erlandsson and Jansson (2004), Hareide and Porathe (2019), Procee and Baldauf (2014)*, support for shipyard floor workers, *Friedewald et al. (2015), Matsuo (2016)*, and support for remote inspection and maintenance of ship systems, *Helle et al. (2014), Lee et al. (2010)*.

AR technologies can support navigators in their work by presenting data from the ship bridge systems to augment the navigator's real world. The navigator may interact with and monitor the bridge systems while simultaneously maintaining focus on the primary visual field outside the ship. AR may improve operator performance and situation awareness (SA) by supporting navigation focus, reducing information overload and linking real and digital information, *Rowen et al. (2019)*.

However, to realise these benefits, AR technologies must to be adapted to maritime users' needs, *Nordby et al. (2020)*, and their environmental and technological contexts, *Nordby and Morrison (2016)*. There is a range of issues that can increase the threshold for safely applying AR technology in the maritime sector. For instance, the ship as a reference point is constantly moving; lighting conditions may vary from pitch black to extremely bright light; temperatures may differ within a wide range; users may suffer from fatigue or motion sickness; and users often move around the bridge while working, *Frydenberg et al. (2018a)*.

In this article we focus upon three types of challenges that must be overcome in order to realise the benefits of AR for ship bridge users. First, a ship bridge is assembled by a multitude of systems that are often delivered by multiple actors with no common user interface (UI) design guidelines across these systems, *Nordby et al. (2019a)*. As a result of poor integration between the systems, navigators often exert high levels of effort and awareness to integrate the translations, overlaps and gaps between the various UIs of the ship bridge systems, *Lützhöft and Nyce (2008)*. In this context AR applications need to be able to function as a common UI platform that can mediate between any maritime application on the ship bridge. However, there is no previous research on how AR technologies can be designed and implemented as a part of a consistent ensemble of ship bridge systems, *Nordby et al. (2020)*.

Second, AR applications may use the entire world as a canvas to display digital information, and therefore a definition of how multiple systems may share the world is needed. For example, there is a need for structuring visualisation formats and their placement and managing how multiple information elements may share the world. Currently, no definition exists for how single or multiple AR applications may safely be rendered over the world, *Nordby et al. (2020)*.

Third, the ship environment and the work situation of the bridge users constantly change during operations. As a result, a system is needed for how bridge users can adapt the AR interfaces to changing needs. Currently, we have found no practical or theoretical examples of AR used in the maritime domain that show how AR applications adapt to changing work conditions and tasks, *Nordby et al. (2020)*.

Overall, there is a lack of frameworks that describe how multiple maritime systems can be designed to enable efficient exploitation of AR on a ship bridge. We address this problem in this article by proposing a system designed to enable multivendor ship bridge systems to share AR as a generic platform for safe use in ship bridges. Our design proposals are iteratively built towards assembling what is referred to as a "system architecture" in computer science and software engineering. *Allen (1997, p. iii)* explains that the architecture of a system "provides a model of the system that suppresses implementation in detail, allowing the architect to concentrate on the analyses and decisions that are most crucial to structuring the system to satisfy its requirements." In order to support the work of ship bridge users, our work focuses on the development of a UI architecture of AR applications that provides bridge users with digital information overlaying the real world in a safe and efficient way. The current version of the UI architecture is built upon the OpenBridge Design System, which includes a UI architecture for maritime workplaces that emphasises screen-based, graphical UIs, *Nordby et al. (2019b, 2018)*.

A UI architecture for AR needs two main components: AR applications and an integration system. AR applications are software programmes that make use of the integration system to access information-augmenting functionalities in a way tailored to the user's needs and their work context. Current AR systems like the Windows Mixed Reality framework, *Microsoft (2020)*, have a similar structure, in which users may open multiple applications simultaneously and place them in a shared space. This approach is inspired by, and derived from, the mobile and internet technologies, in which "apps" are built to behave harmoniously across all the different platforms that use a common operating system, such as Android or iOS. However, our experience is that the Windows Mixed Reality framework is not well adapted to maritime needs and is not tailored to support situational awareness in maritime operations.

This work is part of the SEDNA EU research project and some of the concepts we describe here have been discussed in a previous project report, *Nordby (2019)*. SEDNA is developing an innovative and integrated risk-based approach for safe Arctic navigation, ship design and operation. Our approach at the architectural level enables us to focus on specific Arctic navigation use, while also proposing solutions that may apply to other cases. In the next sections, we first present our research approach and methods, then introduce the different components of the UI architecture and give examples of AR applications built for this architecture.

2. Research approach

We employed user-centred design (UCD), an approach that involves users throughout the development process, to make sure that the solutions we created are anchored in the identified needs of actual users, *Giacomin (2014)*. The research activities took place in our lab facilities and during design-driven field research, *Lurås and Nordby (2014)*, onboard an ice breaker vessel and an ice class coast guard vessel engaged in operations and expeditions.

We developed the UI architecture with the explorative approach illustrated in Fig.1. It describes a parallel process containing a practical level and an abstract level that runs in iterations.

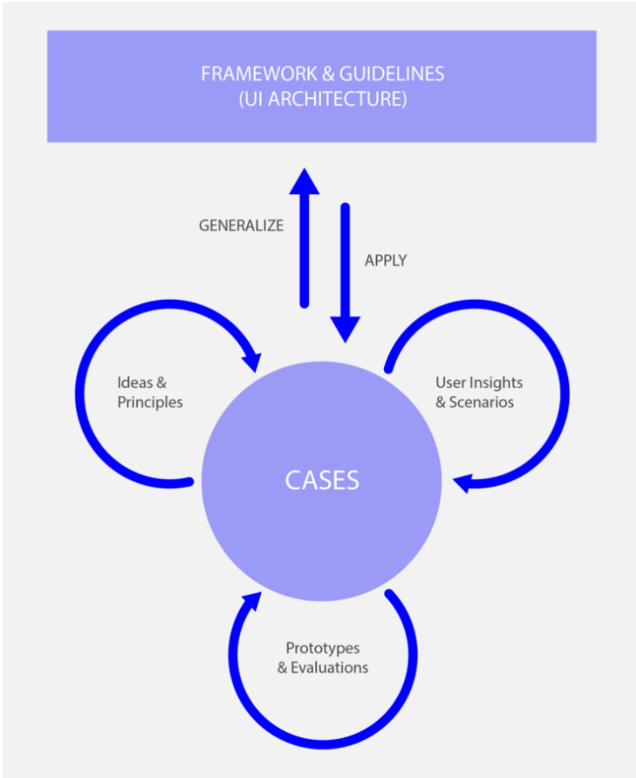


Fig.1: UI architecture process model describing the iterative development of frameworks and guidelines based on case studies

Since our research process has continuous feedback loops, we present the method in parts rather than in a linear fashion.

2.1. Design cases

At the practical level ("Cases" circle in Fig.1), we have a case we need to solve. An example of a case might be to develop a route planning tool for ice going vessels in a convoy. The process of solving a case involves three kinds of activities: 1) creating ideas and principles, 2) collecting user insights and constructing scenarios and 3) developing and evaluating prototypes.

- Creating ideas and principles
  We developed a wide collection of design ideas for potential use in bridge system applications, information display and the definition of AR zones. The ideas are provided as sketches with description. Based on these ideas, we worked further with design principles, schemas and diagrams to describe their use.
- Collecting user insights and constructing scenarios
  Gathering user insights is a continuous process embedded in all forms of activities where users are involved. During our design-driven field research at sea, we employed an explorative and opportunistic approach using methods such as participatory observation, *DeWalt and DeWalt (2011)*, semi-structured interviews, *Kvale (1996)*, eye-tracking, *Hareide and Ostnes (2017)*, and co-design, *Sanders and Stappers (2008)*. The mixed-methods approach for investigating the premises of designing AR systems for navigators resulted in a broad set of both targeted and serendipitous data in form of images, videos, notes, recordings and test protocols, *Frydenberg, Eikenes and Nordby (2019)*. The data were analysed through expert evaluations during collaborative data analysis workshops, *Millen (2000)*. Additionally, the data has formed basis for further idea creation and prototype development. Finally, the data helped to define operational scenarios that we used as a basis for developing ideas and prototypes for each case. We used layered scenario mapping to work with the scenarios, *Lurås (2016)*.

- Developing prototypes and evaluations
  Techniques for prototyping AR applications range from paper sketches to virtual reality (VR) environments in which AR UIs are embedded. VR enables a fast and cost-effective prototyping process with high levels of realism. We built a prototyping platform that combines a cloud-based simulator with a VR environment that uses the Unity game engine, https://unity.com/solutions/architecture-engineering-construction. The VR scene has a realistic ocean environment that can handle several ships at the same time, and allows the daytime, light and weather conditions to be modified. Realistic simulation data (for example, ship speed, heading, engine power load, etc.) can be fed into all the UIs present in the VR scene, including simulated AR UIs. In addition, the simulator data can be used in "real" AR UIs running on an AR headset (a Microsoft HoloLens in the current version of the platform). We tested the prototypes both in the lab and with expert users in the field to evaluate the usability of the AR concepts. In addition to the expert evaluations, data were collected through videos captured from the VR environment using the virtual camera of the VR headset or a secondary virtual camera.

2.2 Frameworks and guidelines

At the abstract level (rectangle in Fig.1), we aim to develop frameworks and guidelines based on our previous cases that can ultimately support solving new cases. The process of generalisation is performed first by analysing the case in order to understand its parts, relationships and premises; second, we identify all the common characteristics we believe are shared among other cases; third, we formulate these characteristics as general concepts that we need to refine by acquiring more knowledge about them, categorising and organising them, putting them into a hierarchy and understanding their interrelations. Finally, these refinements result in guidelines or frameworks, which can function as an architecture.

This parallel process is iterative, involving a new case to solve for every cycle. Since each additional case offers a new set of premises, another generalisation process is required to ensure that new case-specific aspects are incorporated by adjusting the existing frameworks and guidelines. In other words, each iteration increases the reference data at the practical level, which thereby contribute to a gradually wider and more comprehensive architecture at the abstract level.

The UI architecture is gradually developed from a collection of UI components across all the prototypes from the different cases. The exploration of how these components might be systematically and safely placed in the user's context through simulation and in the field has resulted in several frameworks. Finally, we considered how to structure the UIs in the user's environment according to operation and the user's situation.

3. UI architecture

In developing a UI architecture that seeks to enable multivendor UI integration of maritime applications in AR, our emphasis has been on defining a set of rules and building blocks that detail what functionalities AR applications need to include and their appearance. As we did with the OpenBridge UI architecture, *Nordby et al. (2019b)*, we looked at 1) the hardware through which UIs could be accessed, 2) the individual components of generic AR applications and 3) a system to integrate the different applications together. In this case, the UI hardware is defined as a head mounted display (HMD) AR headset such as the Microsoft HoloLens or the Magic Leap. The integration system is divided into:

1. AR application components that handle the kinds of information objects used to show AR information in the world.
2. The information display system that manages how AR components are organised in the world.
3. AR zones that define the way the organisation of AR components adapts to the user's position in the real world.

3.1. AR application components

In theory, AR can display information anywhere and in any way over the real world. As a result, information display needs to be regulated in AR to make it possible for multiple applications to work together and to facilitate an adequate user experience, meaning information is displayed in ways that are predictable for the users and that are well adapted to the specificities of the user´s work environment. We defined a set of basic information objects that show various types of information in the world. These AR application components were designed to facilitate the kinds of information we have seen in our own cases and in published research about AR experiments in the maritime industry, *Nordby et al. (2020)*.

Furthermore, we developed components that are compatible with the OpenBridge UI architecture and design guidelines, *Nordby et al. (2019b, 2018)*. This compatibility is important because we envision AR as an extension of current workplaces and not an independent system, i.e. an application designed for OpenBridge should be accessible through the AR system as well as through traditional screens.

Fig.2: Overview of the five types of application components currently in use in the UI architecture

We defined five main application component types, Fig.2: App display, Widget display, Annotation, Ocean overlay and AR map. The components offer distinct methods of information display and various systems may take advantage of their inherent affordances. Table I shows how the functions of an artic vessel can be mediated to officers using the various formats. Later in this section we show the current versions of a number of the components that we have tested in our lab and onboard ships.

Table I: Examples of functions that can be shown in different display formats of the AR architecture

	App display	Widget display	AR map	Annotation	Ocean overlay
Arctic specific	Convoy app	Ice pressure indicator	Satellite/drone images	Ice sheets	Ice status overlay
	Voyage planning tool		Ice forecast charts	Ice bergs	Ice 3D mesh overlay
	AR map			Snowmobile	
General	ECDIS	Compass	Land	Vessels	Drawing/annotation
	Radar	Heading & course	Depth/safety contours	Waypoints	Safety contour
	Conning	Thrusters	Planned track	People	Depth
	Tasks / checklist	Speed	Cross track distance	Animals	Planned track
	Echosounder	Machine power	Waypoints	Shallow water	Cross track distance
	DP	Wind	Vessels	Helicopter / drone	Other vessel tracks

3.1.1 App display

Display of an application, in its entirety. For example, ECDIS or radar. In our review of cases that use AR, *Nordby et al. (2020)*, we found two examples of the use of an app display: one was used for manoeuvring functionalities associated with conning applications, with some indication of heading, rudder angle, speed and power or load for different engines, *Hugues et al. (2010*); the other was used for navigation functionalities associated with ship traffic surrounding the vessel, with some indication of the position, name and heading of other ships in the area, *Walther et al. (2019)*.

In the current version of our system, the app display is based on the OpenBridge UI libraries. The current resolutions in the HMDs are limited; therefore, there might be a need to repurpose the applications for a lower resolution than on normal screens, which could be achieved through a scaling technique such as responsive design, which allows UIs to scale to different formats. We used the OpenBridge design system that supports free scaling of the UI through responsive design principles.

3.1.2 Widget display

Display of a smaller component from a full application. For example: compass, speed indicator. In our review of AR use cases, *Nordby et al. (2020)*, we frequently found this type of visualisation, which appeared in 15 out of 19 reviewed references. As in the case for the App display, manoeuvring and navigation are the functionalities most often displayed using widgets. The widget format helps to combine small flexible components that can be assembled in different ways throughout the UI. We designed widgets based on the OpenBridge UI libraries and used responsive design for scaling and stacking. Fig.3 shows examples of widgets that we tested on a field study using our mixed reality platform in 2019. The first example is a combination of information usually found in a conning interface, displayed on the bow of the ship. The second example is wind information, displayed on the surface between two windows in the aft deck.

Fig.3: Examples of widgets and a container with several widgets digitally attached to various surfaces Above, a combination of information usually found in a conning interface, displayed on the bow of the ship. Below, wind information, displayed on the surface between two windows in the aft deck.

3.1.3 AR map

Display of location-based information on a 2D or 3D map within a frame, by default placed vertically above the horizon. Out of 19 references analysed in our AR use case review, we found three references using AR maps of the following types: a map used for navigation in a 3D isometric view, *Hugues et al. (2010),* a map showing the positions of other vessels in the neighbouring area of the considered ship, https://www.youtube.com/watch?v=Ioepw3am-KY, and a "velocity obstacles diagram", *Procee et al. (2018)*, which combines the position, heading and speed of surrounding vessels.

We found in our field studies that there is a general need to show data on a map. Because of this, we defined a simplified map information type as a basic component. We envision this area showing any map-related data and having integrated functions such as the ability to link content with real world points of interest (POIs; Fig.4). It should also be able to support various orientations such as north up or following the user´s gaze, Fig.5.

Fig.4: AR map concept linking a point in the map with a position in the world

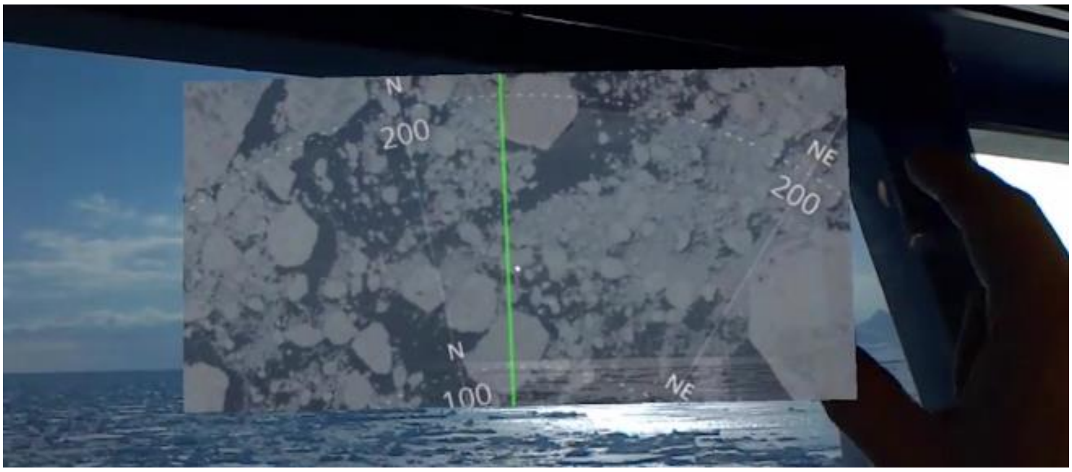


Fig.5: AR map concept in which the map is locked to the user's head gaze direction and will rotate with user's head movements

We proposed a number of concepts for AR applications that may help navigate in ice. The AR map can display the location of an object present in the user's field of view with a line connecting the real object and its location on the map, Fig.4. The map can also follow the user´s gaze, so that the contents of the map are updated depending on where the user is looking, Fig.5. We tested a concept where the AR map can be positioned in containers or be freely positioned in the space, Fig.6.

Fig.6: AR map concept of a map as a free object that can be placed anywhere, for example on a table used to plan a route through ice

3.1.4 Annotation

The annotation format displays a small piece of information connected to a physical object in the world (a POI); for example, it might present information about a vessel present viewable from the bridge. This is a typical part of AR interfaces and we found this type of visualisation in 15 of the 19 AR publications we reviewed, *Nordby et al. (2020)*. Annotations are often used in combination with widget displays, most commonly to display information associated with manoeuvring and navigation functionalities. They are not defined in the current version of OpenBridge. Fig.7 shows an early experiment with annotations from a field study.

Fig.7: Examples of annotations simulated in the mixed reality platform: a lighthouse (top), iceberg (bottom left) and neighbouring vessel (bottom right)

3.1.5 Ocean overlay

An ocean overlay displays information directly on the ocean surface so that the AR graphics match the real location of the point or area it refers to. We found this type of visualisation in 13 of the 19 references reviewed in our AR use case analysis, *Nordby et al. (2020)*. The examples we encountered most often related to navigational functionalities, for instance, the charted route and zones to avoid displayed on the ocean surface.

3.2 Information areas

There is a vast amount of potential information that can be placed in the world. However, there is a finite space around the users, and each type of available space affords different possibilities for information placement, *Norman (1999, 2004)*. Certain areas such as the water surface should have a limited overlay because it might occlude important objects like small boats. Other areas might be less critical for operation, such as the walls or the sky. A structure is needed for how AR may make use of the various affordances of the spaces around the users in a ship bridge. To define this structure, we need information areas that can be used to specify how the AR components may be distributed in the world. We have not seen any explicit description of a generic system for maritime use in previous work, *Nordby et al. (2020)*.

In our proposal for information areas, we focus on the outside region, since this is arguably where the potential of AR is highest, yet also where the risk of obstructing the view is high. We distinguish three generic areas where AR information may be displayed, and we propose to use each area for specific application components, Fig.8. In addition, we allow the free placement of apps and widgets or app/widget containers, Fig.9. Although the proposed structure is usable in most of the cases we investigated, likely there are instances where it must be adapted to individual workplaces. This is because there might be important objects that cross the bands we have defined or other unforeseen operational requirements for the AR interface. In such situations non-occlusion areas should be defined that can be cropped out of the generic display areas.

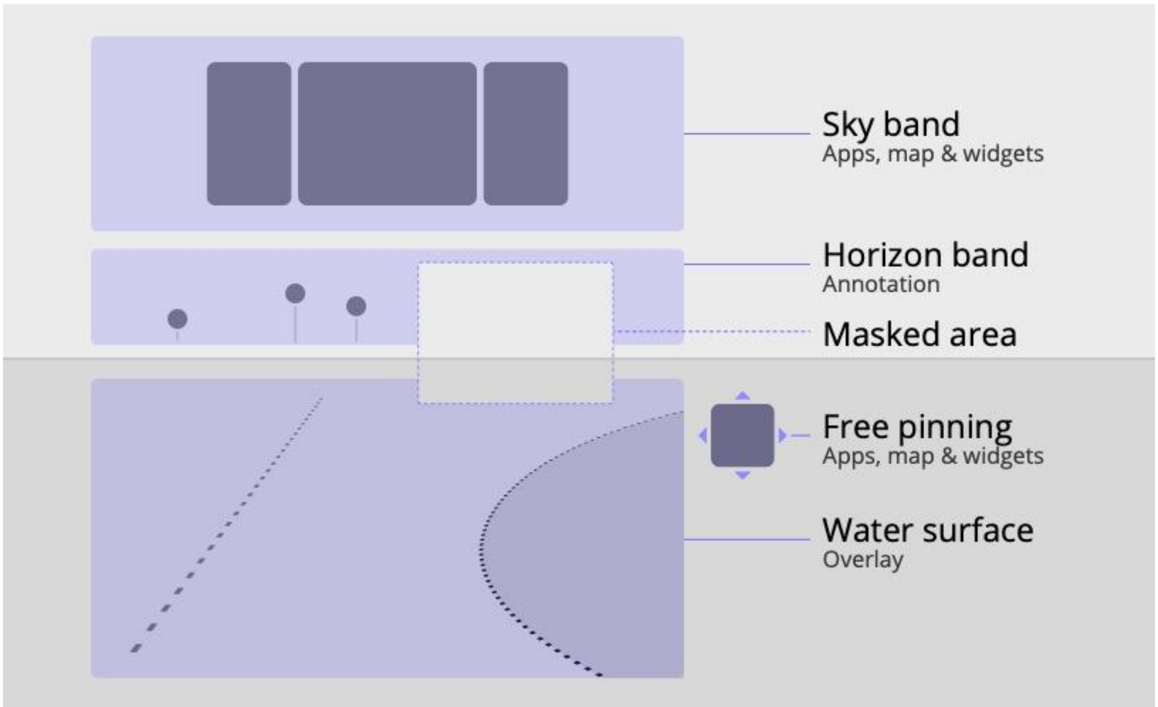


Fig.8: Suggested information areas and the types of application components they may contain

The following presents each of the information areas and describes how they may affect the application components.

- Sky band: The area located above the horizon should be reserved for full apps, an AR map or widgets. We propose this area since it does not occlude the ocean or the central equipment of the workplace. We envision this area being much closer to the water's surface than traditional monitors placed above the windows. As such, neck strain will most likely be less of a factor than any strain associated with screens mounted above widows. It should be possible to automatically organise the components shown in the sky according to importance and available space. The sky band can be fixed at the horizon relative to the ship or repositioned according to the direction the user is facing.
- Horizon band: We propose that the area located near the horizon should only contain annotation components. We want to keep the annotations close to the objects on the water while not overlapping them. When several objects are close to each other, a system is needed to separate the various annotations so that they do not overlap even if the objects on the horizons do.
- Masked area: We introduce areas where no information may be displayed based on certain rules or user input. For example, we believe that masked areas will be useful if a ship or another important object is close to the bridge and it would be unwanted to display information on top of it.
- Water surface: The area located under the horizon band should contain only ocean overlay components. This is a critical area where there is a high probability that the graphics may occlude important objects in the water. Because of this we emphasise the need to reduce information on the water and we suggest the necessity of designing efficient mechanisms that allows users to regulate information density, including control over the information layer and the ability to swiftly take away any overlay. In addition, the UI needs to be able to visually show connections between annotations and objects in the water, and between elements in the sky band and objects on the water.
- Free pinning: Free pinning differs significantly from the information bands since it allows for the free positioning of AR apps, maps and widgets anywhere in the world. This includes the ability to establish container areas that function much like the horizon band and enable the structured display of AR apps, widgets and maps, Fig.9. Free pinning should be used with care to avoid information overload. Should the information displayed become too dense, the user needs to be given the control to reduce the quantity of information. To do this, each object should be easily turned on and off, and whole layers of information could be turned on and off.

Fig.9: Free pinning of a container object with widgets attached to the back of the ship under the horizon band

Fig.10 offers an example of the implementation of information placement on the proposed areas in the mixed reality simulator. Notice that on this image it appears that the information container is placed inside the bridge. However, when wearing the VR headset, it is spatially positioned far away from the ship.

Fig.10: Screenshot from the mixed reality platform with prototypes of AR applications that follow the placement rules of the information areas: the sky band with apps, widgets and an AR map (top), the horizon band with annotations (middle) and the water's surface with an ocean overlay (right)

We are currently detailing this implementation and preparing it for formal user tests. We plan to implement the information display system in HoloLens 2, *Microsoft (2020)*, in the future before we specify further the proposed information areas.

3.3 AR zones

We observed that users need to access different types of information depending on where they are working in the bridge and how they move. In earlier research, for instance, we noticed that users often move from a workstation closer to the windows to be able to observe situations better, *Nordby and Lurås (2015)*. In order to accommodate users' considerable need to move freely around the bridge, we propose to divide the bridge in three types of zones, Fig.11. We suggest that the structure of the information areas as well as their informational content change as users enter and leave these zones.

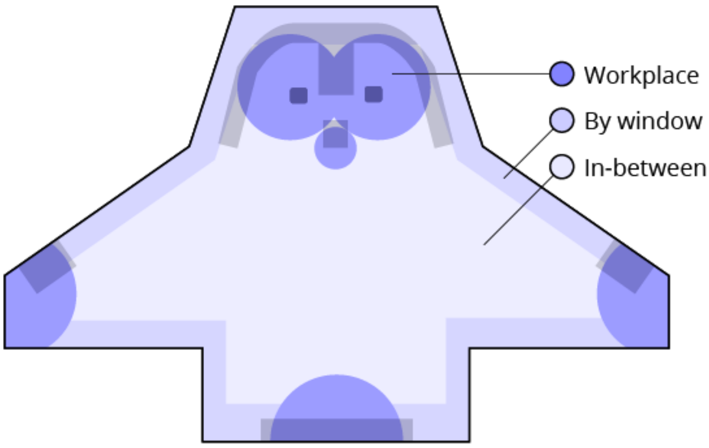


Fig.11: Dividing the bridge into three types of zones that affect the behaviour of the AR application components

The workplace zone is centred on traditional workplaces in the bridge, such as the bridge wings or navigation station. The by window zone is any zone close to the windows. The in-between zone is any other area inside the bridge. Informational content and areas change as users move between these zones, Fig.12. The workplace zone has information adapted to each individual workplace. In addition, we suggest that the sky band shown in Fig.8 becomes fixed to the horizon, right in front of each workplace. Apps and widgets that have been pinned by the user to support an operation or task carried out at this location appear when the user enters the zone.

We propose that the by window and in-between zones should be less rigid than the workplace zone since the physical, collaborative and operational conditions in these areas are hard to predict. Therefore, we suggest the sky band should move with the users' gaze instead of being fixed in any specific direction. In addition, we suggest that there should be a limited number of application components in use when a user is in the in-between zone since they may interfere with the work inside the bridge. Finally, we suggest no use of a water overlay and a limited use of annotations in the in-between zone, unless the user specifically asks for this information to be presented or there is a relevant alert, because the bridge itself and its content occlude the outside world significantly when the user is in the in-between zone.

Fig.12: AR Zones: when the user goes from the "in-between" zone (top) to the "workplace zone" (bottom), the content and placement of the AR application components is rearranged.

4. Discussion

We have presented the need to develop the following frameworks to enable the safe and efficient use of AR applications on ship bridges:

1. A framework for managing the integration of AR applications into existing bridge systems
2. A framework for how multiple AR applications may share the world
3. A framework for adapting AR interfaces to changing needs

We then presented a proposal for a UI architecture with different components and rules for how the components should behave. This proposal constitutes a first step towards the development of the abovementioned frameworks. With regards to the first framework, our approach is that AR applications are an extension of existing bridge systems. In order to offer a seamless user experience, it is important that AR applications follow the same general UI design guidelines as the other screen-based UIs in the bridge. Our experiments show that an AR interface will have some components that are very similar to traditional screen-based applications and some components that are new. Thus far in our work, we see a big potential for harmonising AR design guidelines with the current OpenBridge design guidelines.

The integration of AR applications with existing bridge systems requires a re-examination of the notion of integrated ship bridge environments, because AR applications will not exist in a vacuum. We can envision, for instance, that applications on screens within the AR could be dragged using gestures or that there will be mechanisms to ensure that AR information does not overlap with screens. Further research is needed for the practical integration of AR applications with existing screen-based systems.

With regards to the second framework, our proposal offers concrete solutions for how AR applications may share the world. We proposed a standardised list of five components that follow a set of rules for what information may be displayed with each component and how the components should interact with each other. We see a need to continue this work and to focus specifically on managing information density. The current framework deals with information placement, but does not explain how to manage the density of that information based on automation or user input. The goal is to make sure that information density may be managed effortlessly by users to avoid information overload. This is important when considering the cognitive cost of AR-enhanced operations compared to the status quo, *Baumeister et al. (2017)*.

As for the third framework of a UI architecture that adapts to changing user needs, researchers need to experiment, document and analyse the different ways in which users manage and interact with AR applications. With the implementation of AR zones, our current proposal suggests that the location of the user should automatically impact what and how information is displayed in AR. To go beyond a location-based approach, an operation- or task-oriented framework should be developed as a foundation for managing information displayed in the ship bridge. This framework could build on existing research on cognitive work analysis, *Procee et al. (2017)*, or frequency of use, *Vu et al. (2019)*. Another question is related to how users interact with the information. The current UI architecture proposal only addresses interaction indirectly by using the users' position for changing the interface. Other input mechanisms could include voice commands, a connection to existing generic interaction devices, eye tracking and more. Since the types of data input will vary with work conditions, *Nordby and Lurås (2015)*, it is necessary to develop a UI architecture that caters to different forms of input.

Finally, our current architecture only deals with visual information. Previously there has been a demonstrated use for 3D audio in ship bridges. A future framework could include 3D audio mediated through AR, making it possible to connect radio transmissions to points of interest or separating alarm channels.

5. Conclusion

We presented ongoing work towards a UI architecture for maritime AR applications. We suggest AR should be seen as an extension of current ship bridge systems' interfaces. The UI architecture should be adapted to maritime use and should allow multiple systems to be accessed through AR simultaneously. We argue that the development of AR applications should be tightly connected to design frameworks that are built for an entire maritime workplace. In the case of the ship bridge, we propose to work with frameworks that 1) manage the integration of AR applications into existing bridge systems, 2) guide how multiple AR applications may share the world and 3) adapt to changing situations and user needs.

The preliminary UI architecture for AR applications presented in this paper consists of application components, information areas and AR zones that together offer an initial framework for integrating AR applications into existing bridge systems and sharing the AR space. We argue that this current proposal is a useful starting point for the further development of maritime AR applications. The UI architecture is developed in conjunction with the OpenBridge design system, which currently supports maritime screen-based applications. We suggest that the UI architecture can be an important delivery method to bring AR into the OpenBridge design system.

In future work, we will develop the UI architecture further to focus on designing for situated user views, information management and an OpenBridge UI design guideline harmonised across screen-based and AR-based applications.

Acknowledgements

Thank you to Jon Fauske and Philip Stevju Meidell for the work on the AR prototypes and the VR platform; to all the partners of the SEDNA and OpenBridge projects; and to the crew of the KV Svalbard and Atle vessels who kindly welcomed us for several field studies. This research is part of the EU "Safe maritime operations under extreme conditions: The Arctic case" (SEDNA) project. The project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 723526.

References

ALLEN, R.J. (1997), *A formal approach to software architecture*, PhD Thesis, Carnegie-Mellon Univ Pittsburgh PA School of Computer Science

BAUMEISTER, J.; SSIN, S.Y.; ELSAYED, N.A.M.; DORRIAN, J.; WEBB, D.P.; WALSH, J.A.; SIMON, T.M.; IRLITTI, A.; SMITH, R.T.; KOHLER, M.; THOMAS, B.H. (2017), *Cognitive cost of using augmented reality displays*, IEEE Trans. Vis. Comput. Graph. 23, pp.2378-2388

CAPGEMINI RESEARCH INSTITUTE (2018), *Augmented and virtual reality in operations: a guide for investment*, No. Report MACS CS AP 20180907

DEWALT, K.M.; DEWALT, B.R. (2011), *Participant observation: a guide for fieldworkers*, Altamira Press

ERLANDSSON, M.; JANSSON, A. (2004), *Augmented reality as a navigation aid for the manoeuvring of high-speed crafts*, 8th Int. Design Conf., Croatia, pp.741-748

FRIEDEWALD, A.; LÖDDING, H.; TITOV, F. (2015), *Augmented reality for the retrofit of ships*, 14th COMPIT Conf., Ulrichshusen, pp.236-246

FRYDENBERG, S.; EIKENES, J.O.; NORDBY, K., (2019), *Serendipity in the Field. Facilitating serendipity in design-driven field studies on ship bridges*, The Design Journal 22, pp.1899–1912

FRYDENBERG, S.; NORDBY, K.; EIKENES, J.O. (2018a), *Exploring designs of augmented reality systems for ship bridges in arctic waters*, RINA Human Factors Conf., London

FRYDENBERG, S.; NORDBY, K.; HAREIDE, O.S. (2018b), *Feltstudier for design av utvidet virkelighets-teknologi i navigasjon* (English: Field studies for the design of augmented reality technology in navigation), Necesse 3, pp.67-70

GIACOMIN, J. (2014), *What is human centred design?*, The Design Journal 17, pp. 606-623

HAREIDE, O.S.; OSTNES, R. (2017), *Maritime usability study by analysing eye tracking data*, J. Navig. 70, pp.927-943

HAREIDE, O.S.; PORATHE, T. (2019), *Maritime augmented reality*, Coordinates Magazine, February, pp.31-35

HELLE, S.; KORHONEN, S.; EURANTO, A.; KAUSTINEN, M.; LAHDENOJA, O.; LEHTONEN, T. (2014), *Benefits achieved by applying augmented reality technology in marine industry*, 13th COMPIT Conf., Redworth, pp.86-97

HUGUES, O.; CIEUTAT, J.-M.; GUITTON, P. (2010), *An Experimental Augmented Reality Platform for Assisted Maritime Navigation*, 1st Augmented Human Int. Conf., New York, pp.12:1-12:6

KVALE, S. (1996), *Interviews: An introduction to qualitative research interviewing*, Sage Publ.

LEE, J.; LEE, K.; KIM, K.; KIM, D.; KIM, J. (2010), *AR-based ship design information supporting system for pipe maintenance*, 11th PRADS Symp., Rio de Janeiro, pp.607-612

LURÅS, S. (2016), *Layered scenario mapping: A multidimensional mapping technique for collaborative design*, CoDesign 12, pp.133-150

LURÅS, S., NORDBY, K., (2014). *Field studies informing ship's bridge design*, Int. Conf. Human Factors in Ship Design and Operation, London

LÜTZHÖFT, M.; NYCE, J.M. (2008), *Integration work on the ship's bridge*, J. Marit. Res. 5, pp.59-74

MATSUO, K. (2016), *Augmented reality assistance for outfitting works in shipbuilding*, 15th COMPIT Conf., Lecce, pp.234-239

MICROSOFT (2020), *Microsoft HoloLens 2 | Mixed Reality Technology for Business*, https://www.microsoft.com/en-us/hololens

MILLEN, D.R. (2000), *Rapid ethnography: time deepening strategies for HCI field research*, 3rd Conf. Designing Interactive Systems: Processes, Practices, Methods, and Techniques, New York, pp.280-286

NORDBY, K. (2019), *Arctic ship bridge demonstrator for SEDNA project*, Report SEDNA D2.4 (unpublished), AHO

NORDBY, K.; FRYDENBERG, S.; FAUSKE, J. (2018), *Demonstrating a maritime design system for realising consistent design of multi-vendor ship's bridges*, Human Factors Conf., London

NORDBY, K.; GERNEZ, E.; EIKENES, J.O.; HAREIDE, O.S. (2020), *A review of augmented reality applications for ship bridges*, Submitted to Necesse Journal

NORDBY, K.; GERNEZ, E.; MALLAM, S. (2019a), *OpenBridge: designing for consistency across user interfaces in multi-vendor ship bridges*, Ergoship 2019 conference, Haugesund

NORDBY, K.; LURÅS, S. (2015), *Multimodal interaction for marine workplaces used as strategy to limit effect of situational impairment in demanding maritime operations*, Int. Marine Design Conf., London

NORDBY, K.; MALLAM, S.C.; LÜTZHÖFT, M. (2019b), *Open user interface architecture for digital multivendor ship bridge systems*, WMU J. Maritime Affairs. 18/2, pp.1-22

NORDBY, K.; MORRISON, A.D. (2016), *Designing calm technology and peripheral interaction for offshore service vessels*, Pers. Ubiquitous Comput. 20, pp.601-613

NORMAN, D. (2004), *Affordances and design*, https://jnd.org/affordances_and_design/

NORMAN, D. (1999). *Affordance, conventions, and design*, Interactions 6, pp.38-43

PROCEE, S.; BALDAUF, M. (2014), *Augmented reality as part of a man-machine interface in e-navigation*, Int. Symp. Information on Ships, Hamburg

PROCEE, S.; BORST, C.; VAN PAASSEN, M.M.; MULDER, M. (2018), *Using augmented reality to improve collision avoidance and resolution*, 17th COMPIT Conf., Pavone, pp.239-249

PROCEE, S.; BORST, C.; VAN PAASSEN, M.M.; MULDER, M. (2017), *Toward functional augmented reality in marine navigation: a cognitive work analysis*, 16th COMPIT Conf., Cardiff, pp.298-312

ROWEN, A.; GRABOWSKI, M.; RANCY, J.-P.; CRANE, A. (2019), *Impacts of wearable augmented reality displays on operator performance, situation awareness, and communication in safety-critical systems*, Appl. Ergon. 80, pp.17-27

SANDERS, E.B.-N.; STAPPERS, P.J. (2008), *Co-creation and the new landscapes of design*, Co-Des. 4, pp.5-18

VU, V.D.; LÜTZHÖFT, M.; EMAD, G.R. (2019), *Frequency of use – the first step toward human-centred interfaces for marine navigation systems*, J. Navig. 72, pp.1089-1107

WALTHER, L.; SCHULTE, B.; JAHN, C. (2019), *Shore-side assistance for remote-controlled tugs*, 18th COMPIT Conf., Tullamore, pp.274-285

Practical Shape Optimization Using CFD: State-Of-The-Art in Industry and Selected Trends

Stefan Harries, FRIENDSHIP SYSTEMS, Potsdam/Germany, harries@friendship-systems.com

Abstract

The paper presents an overview of shape optimization using Computational Fluid Dynamics. The majority of practical applications found in naval architecture, turbomachinery and engine design builds on coupling parametric modeling to viscous flow simulations and, based on this, running design studies (via Design-of-Experiments) and formal optimizations. Typical methods will be discussed and references for further reading will be given. Interesting trends that can be seen will also be addressed. Example projects from industrial applications will be shown for illustration.

1. Introduction

Optimization is a human trait. People want to find a good, possibly the best solution under the given circumstances. Getting the nicest vacation within a certain budget is essentially not so different to designing a pump with smallest pressure loss, a ship with lowest energy consumption or an offshore platform with the least risk of damage in rough seas: Mathematically speaking it is minimizing (or maximizing) one or several objectives within a set of constraints.

The term optimization is used ambiguously though. Traditionally, an engineer would rightfully say that something was optimized once a handful of feasible options has been considered. However, since quite a few years, modeling tools, formal optimization algorithms, simulation codes and adequate computer power have become available that allow the generation and assessment of hundreds and even thousands of virtual prototypes. Today, resources for high-performance computing (HPC) are not only accessible to a few dedicated experts in large companies but are increasingly utilized by designers even in small companies. This is all the more true as economic models for Cloud-based solutions have been introduced, *Albert et al. (2016)*.

The techniques involved are manifold and, to some extent, domain dependent. The most important watershed is if optimizations are performed with respect to fluid dynamics or structures. In this paper, focus is put on shape optimization for aero- and hydrodynamic performance. Prominent examples are the internal thermo-fluid dynamics of turbines and engines, the external aerodynamics of cars, the hydrodynamics of ship hulls and their propulsion systems as well as the fluid dynamics of mixers, valves, ducts, turbochargers and pumps.

This paper aims at giving an overview so that the reader would come to know common techniques and typical applications, illustrated along projects realized within the Process Integration and Design Optimization (PIDO) system CAESES®, www.caeses.com. An encompassing treatment would call for a book rather than a paper. Consequently, many interesting topics and details are omitted here for the sake of an informative, yet reasonably light read. The paper addresses engineers facing optimization projects, managers wishing to get a general understanding, and students developing an interest in the field of shape optimization using CFD. For details and further reading quite a few references are given.

2. Benefits

Shape optimization offers several benefits: It increases a company's competitiveness via

- Better understanding of the design task (and the design space),
- Creating products with superior performance (and better trade-offs),
- Allowing shorter time-to-market (and faster response to market changes),
- Reducing risk (and building confidence),

- Saving costs (and avoiding expensive late changes).

Shape optimization is conducted both for investigating new ideas and possibilities at the initial design stage and for fine-tuning of a given product at a later stage when only small changes are still acceptable, Fig.1.

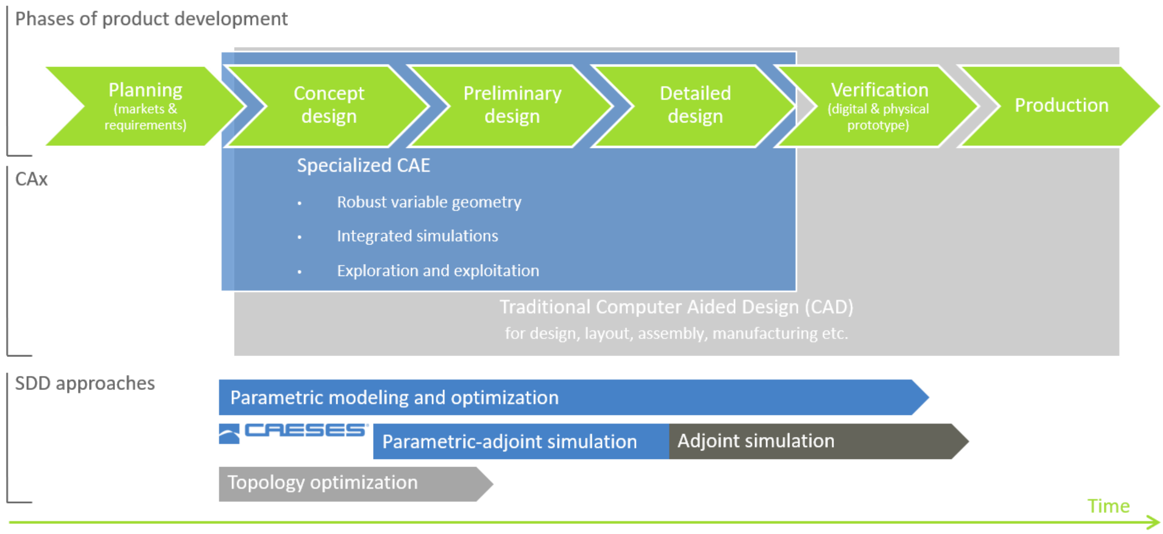


Fig.1: Phases of product development

Naturally, the potential for major gains is highest the earlier an optimization is undertaken. Even experienced design teams can find interesting new ideas during an initial design phase when running design studies and optimizations. Furthermore, if little experience is available, say a brand-new product shall be developed, optimization helps people gain insight.

With flow-exposed surfaces that feature compound curvature, small yet concerted shape adjustments often bring about tangible improvements. For example, automated optimization of ship hulls typically yield a three to five percent gain in energy efficiency even when starting from a mature baseline, *Harries (2008), Harries and Abt (2019b).*

Often optimization seems to be purely regarded as an approach of finding products of superior performance. It is of equal importance, however, that optimization is also about providing options from which to choose, complementing the knowledge base for high-quality decisions. With the increasing popularity of Artificial Intelligence (AI), there are two additional aspects that have come into play: the generation of input data for response surfaces (surrogates) and their subsequent utilization in the context of multi-disciplinary optimization, *Harries and Abt (2019a).*

3. Optimization approaches

There are two major approaches: Parameter-based and parameter-free shape optimization. Parameter-based optimization, as the name suggests, is built on parametric models. Parameter-free optimization comprises topology optimization and adjoint simulation, *Harries and Abt (2019b).*

3.1. Parameter-Based Optimization

Parametric modeling defines a product (or system) by means of important descriptors. Today, most optimization projects are parameter-based. The main reasons are that, firstly, design teams can easily interpret the meaning and impact of design parameters and, secondly, that multi-objective and multi-disciplinary optimizations can be carried out without any conceptual hurdle.

Many parametric models are developed within traditional CAD systems during the detailed design phase, Fig.1. These models are mainly intended for production and often contain details that complicate simulation (e.g. need for defeaturing). Also, they are typically not meant for automated variation, which often results in a high rate of failure to regenerate the geometry when changing parameter values. Thus, for shape optimization using CFD, special parametric models are needed, so-called engineering models, which describe the product with as few significant parameters as possible, sometimes deliberately leaving out characteristics of lesser importance. These models address the concept and the preliminary design phases, focusing on simulation-ready CAD and robust variation, and are realized within dedicated CAD systems, Fig.1. Two major traits of upfront CAD are distinguished: Fully parametric modeling (FPM) and partially parametric modeling (PPM), *Harries et al. (2018), Abt and Harries (2007), Harries et al. (2004).*

3.1.1. Fully parametric modeling

In fully parametric modeling the entire shape is defined by means of parameters. A hierarchical model is created in which parameters describe all features of the envisioned product. Some parameters may be at a high level like the length, width and height of an object. Other parameters may determine details like an entrance angle at a particular location. Typically, many parameters are set relative to or as combinations of other parameters, while some parameters may even be determined from additional analysis (e.g. formulas, background computations, nested optimization to reach a target value). A parametric model can be looked at as a system that takes parameters as input and produces a shape as an output. Any shape is realized from scratch and variants are brought about by simply changing one input at a time or several inputs concurrently. For optimization, fully parametric modeling is very powerful since it enables both large changes in the early design phase and small adjustments when fine-tuning at a later point in time.

While many traditional high-end CAD systems support or even advocate fully parametric modeling, very few systems were actually developed for the parametrics of flow-exposed shapes such as turbine blades, ship hulls and pump volutes. These shapes often feature one distinct path in which design information changes rather slowly while the building strategy (or building pattern) orthogonal to that path stays pretty much the same, Fig.2. E.g., the blades of a propeller do not change significantly from hub to tip. Rather, the (cylindrical) profiles are nicely defined by the same parameter set with different values at each radius, comprising chord length, pitch angle, maximum thickness and camber etc. Similarly, the sections of a ship hull change only gradually when sweeping from stern to stem, smoothly transferring between a handful of topologically different regions. Here, the parameter sets in longitudinal direction contain information about the deck, design waterline, sectional area curve, lateral profile etc., *Harries (2010)*. Furthermore, the cross section of a volute does not really look that different from one angle to the next but rather slowly evolves in circumferential direction with gradually changing input such as the local ratio of cross-sectional area to centroid radius. Finally, the piston bowl profile of a diesel engine might change circumferentially, too, even though there may be some periodicity in the building pattern.

This is illustrated in Fig.2, taking models created in CAESES. The system offers upfront CAD functionality dedicated to variable geometry as needed for optimization. High-level geometric constraints can be readily incorporated while processing the shapes for maximum fairness. The blade of a turbine or a propeller may have to comply with a prescribed area distribution (e.g. for load considerations) while the hull of a ship or an offshore platform must meet a given displacement (e.g. to balance weight estimates).

Rather than treating these characteristics as output from the modeling process and subsequently adjusting geometry manually until all constraints are satisfied, as would be done in most traditional CAD tools, many constraints can be directly incorporated into the model. Fig.3 illustrates this for three variants of a semi-submersible (i.e., the lower part of an offshore platform). The parametric model is engineered such that a specified target volume is intrinsically satisfied for all modifications. Similarly, the volume of a piston bowl needs to be kept constant during an optimization campaign to maintain a fixed compression ratio, see Fig.2 and Fig.13. An inner optimization is undertaken within the model to adjust selected parameters such that the equality constraint is readily satisfied before launching any simulation, constituting a process of nested optimization.

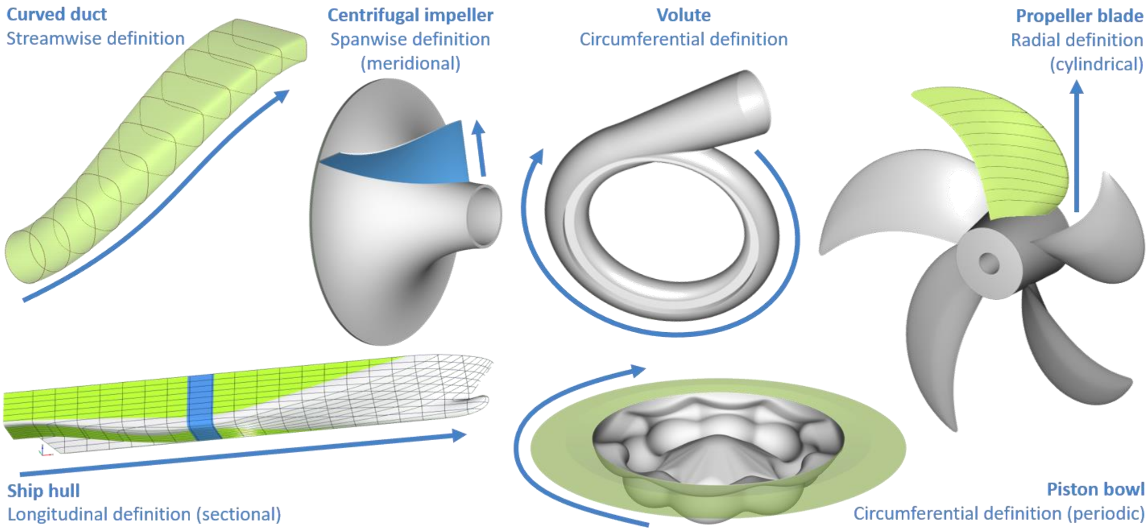


Fig.2: Examples of fully-parametric models with distinct directions of information
[See also www.caeses.com/support/videos]

Fig.3: Variations of a semi-submersible via a fully parametric model with an integrated constraint on volume at fixed draft (realized internally as a nested optimization)

3.1.2. Partially parametric modeling

In partially parametric modeling only the changes to an existing shape are defined by parameters, while the baseline is taken as input. The baseline may stem from any previous modeling process, for instance from a traditional CAD tool. Prominent representatives of partially parametric modeling are morphing, free-form deformation (FFD) and shift transformation (e.g. shifts in coordinate direction, radial shifts).

In morphing two or more objects are combined that are geometrically different but topologically identical. A well-known example is that of a cat and a dog which both feature two ears, two eyes and a snout (same topology) but, naturally, look quite different (different geometry). Mixing them with weights between zero and one creates anything from a pure cat to a pure dog with stages of transition from more-cat-than-dog to more-dog-than-cat. (In small-talk, often all partially parametric modeling techniques are subsumed as morphing even though that is not quite right from a mathematical point of view.)

In free-form deformation, also known as box deformation, the geometry to be modified is enclosed by a regular grid of vertices (i.e., rows, columns and layers defining a B-spline volume). For all parts of the initial shape which lie within the box (or lattice) a coordinate triple can be determined. By moving any of the vertices the box changes its shape and, along with it, the baseline is transformed, *Sederberg and Parry (1986)*. Here, the free coordinates of the box vertices serve as parameters. The technique is applicable to both surfaces and volumes, allowing FFD to be exercised on a CFD mesh, too.

Shift transformations typically change any point in space by adding a certain displacement depending on the point's initial position. It can be applied to both continuous data (e.g. surface patches) and discrete data (e.g. points, offsets, tri-meshes as used for data exchange via STL). Fig.4 gives an example realized in CAESES, showing a vertical shift of a container ship's bulbous bow.

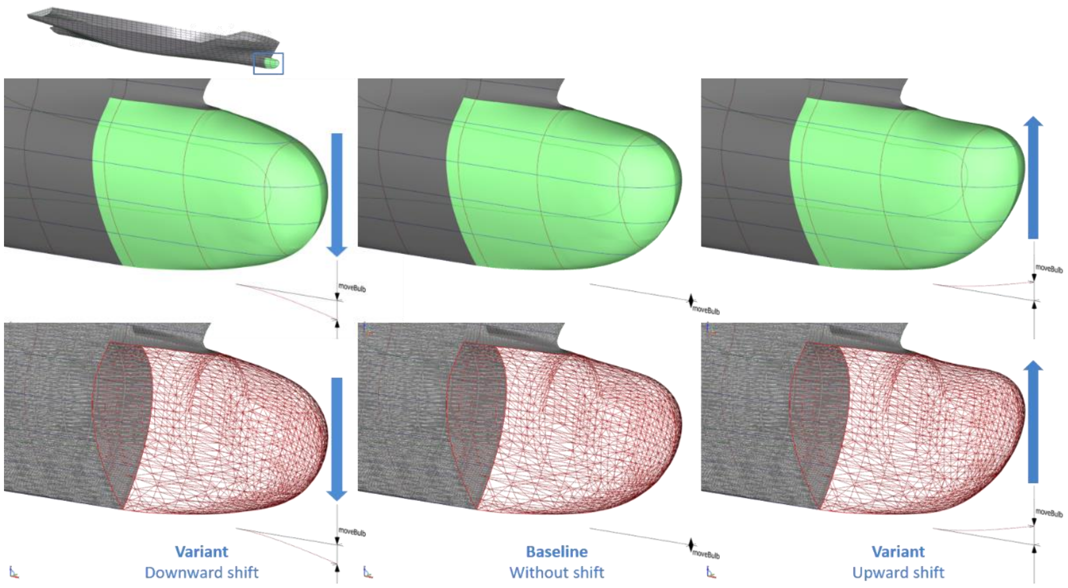


Fig.4: Partially-parametric model for a downward vertical shift (left column), the baseline and an upward vertical shift (right column) applied to both B-spline surface patches and tri-meshes [the entire ship hull along with zoomed-in part (rectangle) is shown in the upper left corner]

Partially parametric models are usually quick and fairly easy to set up. When compared to fully parametric models they typically contain less knowledge (intelligence) about the product. In general, it is more difficult to excite large (game changing) modifications. After all, the new shapes are derived from the baseline and, thus, cannot look totally different. Still, they are well suited for fine-tuning without much overhead.

Parameter-based optimizations usually follow a simple line of action: All parameters that are believed to be important and that are under control of the design team are changed by an optimization algorithm, creating variants for assessment. For each variant one or several simulations are undertaken, returning objectives and constraints. This is repeated until either a certain number of variants has been studied (e.g. via a Design-of-Experiment) or a meaningful improvement was found (or, possibly, time and budget are consumed), see Section 4 for details.

An elaborate discussion of modeling approaches, including hybrid models that combine fully and partially parametric modeling, is given in *Harries et al. (2018)*. Further examples of fully and partially parametric models in naval architecture in the context of holistic design can be found in *Harries and Abt (2019)*.

3.2 Parameter-free optimization

Topology optimization is mostly applied for solving internal flow problems early in the design process (conceptual level), Fig.1. The available domain along with its inflow and outflow boundaries is prescribed by the designer. The most favorable flow path is then automatically established for objectives such as pressure drop, homogeneity, lift and drag. The available space silts up (sanding) during the flow simulation, iteratively establishing the best organic shape, *Stephan et al. (2008)*. Next, the shape has to be manually approximated in a CAD model, taking production constraints into account. At the end, CFD simulations are performed for the final shape to validate the outcome.

Adjoint simulations are utilized to fine-tune well advanced designs rather late in the process, Fig.1. The method is built on a CFD simulation for the baseline (primal solution), mostly solving the RANS equations, plus an additional simulation of similar effort that solves the so-called adjoint equations for the objective of interest, see e.g. *Stück et al. (2011), Kröger and Rung (2015)*. A CAD model is not necessarily involved at this stage since both the primal and the adjoint solution are run on the same computational grid. The adjoint simulation provides sensitivities on the domain boundaries, showing where to push geometry inwards and where to pull it outwards for improvements, Fig.5. Standard objectives are drag, homogeneity of the flow field and pressure drop. A transfer needs to be made from the sensitivities towards the new shape. At CFD level, this is done by shifting selected grid points on those domain boundaries that are free to change. To do so, the grid points are moved by a small amount proportional to the adjoint sensitivities in the direction normal to their corresponding surfaces, displacing adjacent grid points in the discretized domain. Alternatively, a designer modifies the original CAD shape, proposing a slightly better design inspired by the sensitivities. Mathematically speaking, the adjoint simulation provides the gradient of the chosen objective at the current baseline. Hence, the search direction in which to find an improvement is known while the step size for the variation is lacking. Heuristically, small moves are made and the process is repeated a few times, *Christakopoulos and Müller (2012).* At the end, CFD simulations are again performed to check the achievements.

There are hybrid solutions, too, like parametric-adjoint techniques, e.g. *Robinson et al. (2012)*. The adjoint sensitivities and the design velocities (i.e., the changes of the shape when slightly perturbing any parameter at a time) are combined to yield improved shapes without deteriorating the quality of the surface. Fig.5 shows an application within CAESES. Here a fully parametric model of the sports car's rear wing was combined with the viscous flow code iconCFD by ICON, www.iconcfd.com, *Brenner et al. (2014)*, to improve downforce without increasing drag while maintaining the wing's styling.

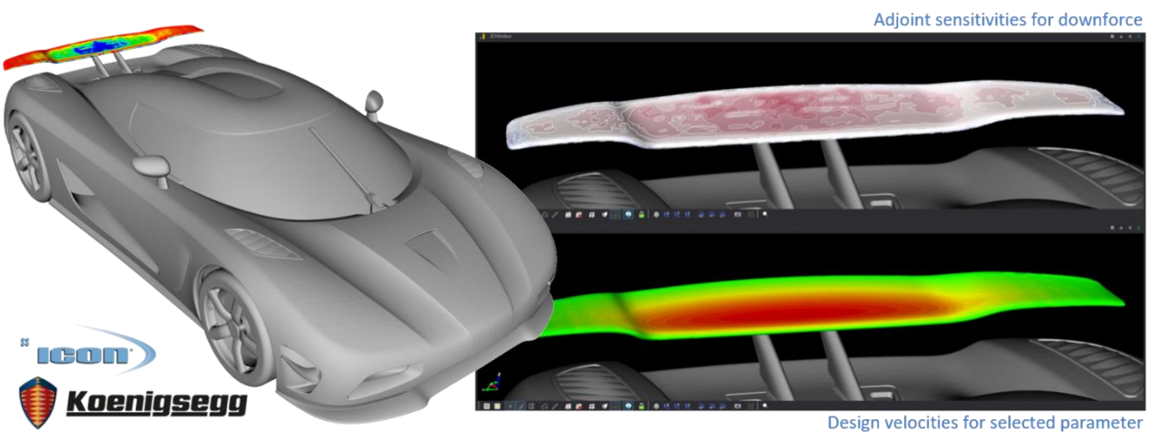


Fig.5: Adjoint sensitivities for increase of downforce and design velocity from parametric model [By courtesy of Koenigsegg Automotive AB and ICON]

3.3 Comparison of approaches

The various parameter-based and parameter-free optimization approaches have their individual advantages – and weaker points. When selecting an approach, naturally, a design team's experience and the available engineering environment play important roles. Suggestions which approach to favor for which task are given in Table I.

Table I: Overview of shape optimization approaches

Approach		**Strong points**	**Weak points**	**Design stage**	**State**
Parameter-free optimization	**Topology optimization**	Innovative and unconventional designs Fast (one extended solver run)	Extra work in remodeling results in CAD tool Sometimes difficult to correlate with constraints, e.g. for production Limited to pre-defined objectives (e.g. pressure drop, homo-geneity) Only for internal flows Difficult for multi-objective and multi-disciplinary scenarios	Concept design	Successful projects in selected industries Ongoing R&D
	Adjoint simulation	Fast (one additional solver run comparable to original simulation effort)	Confined to small changes (unless utilized repetitively) Limited to implemented objectives (e.g. drag, homogeneity) Difficult to apply in multi-disciplinary scenarios	Fine-tuning	Scientific applications Ongoing R&D, e.g. hybrid approach (parametric-adjoint)
Parameter-based optimization	**Fully parametric modeling**	CAD geometry of high quality Allows global & local changes Directly incorporating many constraints Applicable to multi-objective and multi-disciplinary optimization Creates overview and insight	Many solver runs Investment into parametric model needed By definition confined to design space of parametric model Problems with regenerating shapes for modified parameter values when utilizing traditional CAD tools	Initial design to fine-tuning	Mature Successful projects in many industries Ongoing R&D, e.g. robust design optimization, HPC/Cloud
	Partially parametric modeling	Quick and easy to set up and conduct Focus on local changes Applicable to multi-objective (and multi-disciplinary) optimization	Many solver runs By definition limited by design space of parametric model and baseline Incorporating constraints, e.g. for manufacturing, can be challenging	Initial design to fine-tuning	Mature Used in many industries Ongoing R&D, e.g. usability

Since parameter-based optimization is quite mature and very popular the remainder of the paper shall be mainly devoted to this approach.

The quality of a parametric model is decisive for the success of an optimization. This is because understanding an *n*-dimensional design space spanned by *n* free variables of the model – namely the parameters that shall be consciously changed – is anything but trivial. As a rule-of-thumb, a design team needs to study about *n* times *n* variants to gain a reasonable appreciation of system behavior, *Bergmann et al. (2018)*. (A statistically sound estimate involves many more factors, *Siebertz et al. (2010)*.) If there are only two free variables, four variants would give a first insight. From a mathematical point of view it would allow a bilinear approximation of system behavior. If 10 free variables are involved, 100 variants ought to be evaluated. Working with many parameters or with parameters that are not really decisive quickly scales up the optimization task beyond all practical resources. Consequently, industrial projects rarely involve more than 30 to 50 free variables but rather show 10 to 20, possibly after the successive removal of less important parameters. Fig.6 illustrates qualitatively how many simulations are needed depending on the approach taken and, hence, the degrees-of-freedom (DoF) of the system. See *Harries and Abt (2019)* for a detailed discussion.

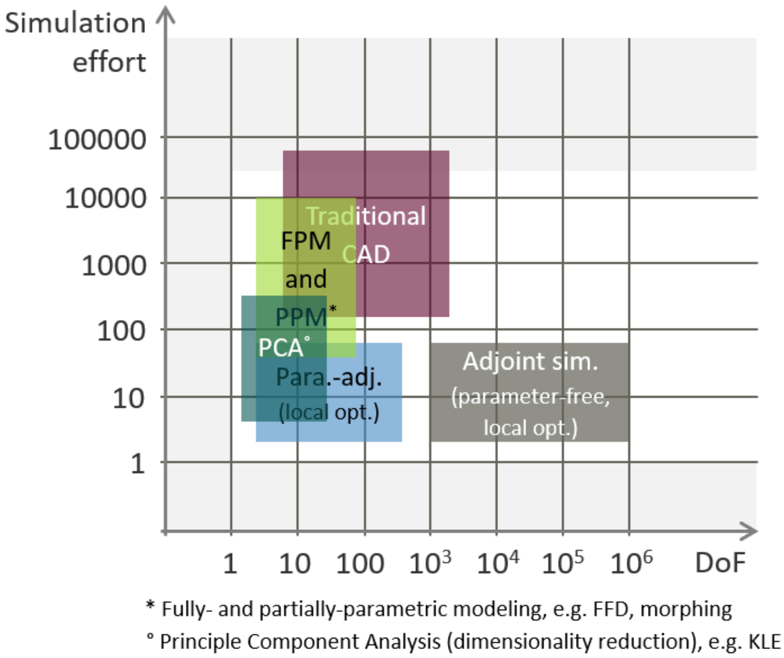


Fig.6: Simulation effort vs. degrees-of-freedom, *Harries and Abt (2019)*

In order to keep the number of free variables (DoF) as low as possible from the start, a parametric model needs to be developed that suits the design and optimization task. An ideal parametric model (engineering model) for shape optimization is characterized as follows:

- All parameters are (rather) independent from each other,
- All potential variants are intrinsically fair (i.e., free of any unwanted shape characteristics),
- Many (geometric) constraints are readily incorporated,
- All variants are geometrically fit for simulation (e.g. free of gaps, folds, overlaps etc.),
- Shapes can be produced beyond the current engineering practice while avoiding unacceptable artifacts (well-balanced model).

The last requirement is the most challenging to satisfy: By definition all parametric models confine the potential outcome. In other words: Design freedom is deliberately reduced. Nevertheless, the model must allow for new shapes and should even contain some element of surprise. Otherwise nothing new can be found.

Finally, it should be noted that the first requirements is also not trivial to achieve. For complex shapes several parameters typically influence certain regions simultaneously, even if in different ways. The (in)dependency of parameters for complex shapes can be tested by means of a Principle Component Analysis (PCA), see also Fig.6, which also supports a deliberate reduction of the dimensionality of the design space (as will be further discussed in the Outlook below).

4. Optimization Process

Optimization is sometimes perceived as a black box that, magically, pops out the best design. Unfortunately, this is not quite right. Rather, apart from the automated number crunching, design studies and formal optimization are interactive processes with adjustments and reconsiderations, iteratively leading to design improvements and innovation on the basis of (and that is good news for engineers) task and product specific expertise.

4.1 Typical Practice

Parameter-based optimization processes – though they might differ in scope and conduct depending on the teams, tools and tasks at hand – have a common set of elements and a typical course of action: At the very start, the team needs to discuss and agree on objectives, free variables and constraints. This is a critical part since all stakeholders need to express their expectations, bring in their knowledge about system boundaries and commit to the content of the project. Most projects then continue with a preparation phase during which a reasonable simulation set-up is established. Ideally, a grid dependency study is undertaken to identify a resolution fine enough for acceptable accuracy, at least with regard to the correct ranking of variants, yet coarse enough for short turn-around time. Two distinct phases often follow: Exploration and exploitation.

During the exploration phase the design space is scanned with the aim of identifying promising regions and understanding sensitivities, distinguishing parameters of higher importance from those with less influence. Sometimes the exploration results are used to build meta-models, i.e., response surfaces or surrogates, to replace the costly CFD in a subsequent step of exploitation.

Next, an exploitation phase follows to squeeze out the best possible results. Commonly this is done within reduced regions, searching for local optima, or within a subspace in which some of the less important free variables are frozen. Finally, selected variants are analyzed and compared to the baseline, possibly at higher grid resolution and for conditions (and in disciplines) not considered during the optimization. From those the most favored variant is selected, concluding the project.

4.2 Format

Optimization problems are cast into a standard format so that available mathematical techniques can be put to use without need for individual adaptations. Five elements have to be written down:

- Objective(s): What shall be improved (i.e., minimized or maximized)?
- Free variables: What can be changed (and is under control of the team)?
- Constraints: What needs to be observed (making a design feasible or infeasible)?
- Fixed parameters: What influences the system but is kept (or assumed) constant?
- Noise: What influences the system but is beyond control (e.g. scatter in material property)?

In shape optimization using CFD, objectives and constraints are non-linear functions of the free variables. Free variables are mostly real numbers with some integers that then often represent topological information, e.g. number of blades. Free variables usually have lower and upper bounds which need to be chosen with care. Tight bounds give a small design space while loose bounds offer more room for improvements. If unsure one may commence with tighter bounds that are subsequently loosened or shifted as the project matures.

Constraints are subdivided into inequality and equality constraints. Inequality constraints describe limits up to which a design is still acceptable. E.g. the shortest distance between the shape to optimize and another object must be larger than a given value (a hard-point constraint), Fig.7. Inequality constraints are often considered via penalty functions (or barriers). The idea is that objectives are artificially worsened by adding an extra term (i.e., the penalty) as soon as a variant is found to be infeasible. The penalty gets larger with the distance from the feasible domain, *Gill et al. (1982)*. Equality constraints describe characteristics that need to be met exactly, e.g. the volume enclosed by a component has to equal a given value as for a hull, Fig.3, or a piston bowl https://www.caeses.com/blog/2016/piston-bowl-design/. An elegant solution of handling an equality constraint is to incorporate it directly into the parametric model. This not only guarantees feasibility with regard to this constraint but may even help to reduce the number of free variables. As an alternative, equality constraints are relaxed to resemble inequality constraints, e.g. the volume has to be equal to or larger than the desired value. Occasionally, an equality constraint is combined with the objective via a so-called Lagrange multiplier, leading towards an extended objective at the cost of an additional unknown, *Birk and Harries (2003).*

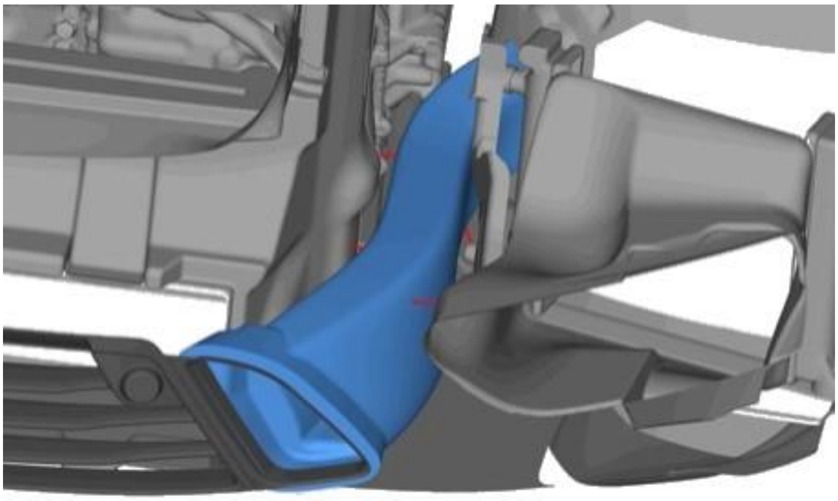


Fig.7: Clearance between component to be optimized and surrounding objects

Sometimes the objective (e.g. resistance, pressure drop) and constraints (e.g. distance to a hard point) are obvious. However, there are design tasks for which several performance measures can serve either as objectives or as constraints. If in doubt, a final decision can be made on the basis of an exploration.

Taking the time to write down objectives, free variables and constraints explicitly is more important than it may initially seem. It is a team-building exercise to reflect what is really important and what may be omitted after all. Furthermore, it helps to avoid the rather unpleasant situation that further requirements come into play at a later point of time. E.g., if a new inequality constraint is introduced well into the project, many good designs may become infeasible all of a sudden (not a nice thing after days of number crunching).

4.3 Components

Typically, a parameter-based optimization project involves the following components:

1. Variable geometry: A parametric model is developed and a shape variant is created as an instance of the chosen parameter values.
2. Pre-processing: The variant is pre-processed (e.g. generation of a watertight triangulation of the shape) to enable the simulation(s).

3. Simulation: For variants of interest one or several simulations are undertaken, by
    - Discretizing the fluid domain (e.g. by generating a volume mesh) and
    - Solving the governing flow equations.
4. Post-processing: Variants and their flow data are post-processed (e.g. visualizing flow fields for comparison) and, finally,
5. Optimization & Assessment: Variants are produced and assessed in accordance to the selected optimization strategy, repeating the sequence from variable geometry to post-processing again and again.

These five components constitute a synthesis model, tightly coupling CAD and CFD, see also Fig.C-1 and Fig.C-2 in Appendix C. The actual process of creating a specific shape, discretizing this shape and, subsequently, the fluid domain, doing the number crunching, collecting the data and, finally, assessing the current design for objectives and constraints is repeated many times. The chosen algorithms, as will be explained below, decide how variants are brought about and how many are considered. Today, if simulations take a couple of hours per case several hundred variants are studied. Frequently, this is done over the course of a long weekend and, possibly, by distributing the heavy workload of simulation to an internal cluster or to an external High Performance Computing (HPC) facility, *Albert et al. (2016)*. If simulations can be completed quickly enough there are projects that even cover as many as several thousands of variants.

4.4 Exploration

In multi-dimensional design spaces, a team has to trade resources against insight. To do so with high efficacy, Design-of-Experiments (DoE) have been developed. These are mathematical algorithms that create as much understanding as possible with as little cost as needed, *Siebertz et al. (2010)*. In general, an exploration – also called a design study – helps to

- Understand the design space and identify regions of interest,
- Find favorable variants, hopefully giving some performance improvements already,
- Evaluate sensitivities, possibly leading to a reduction of the number of free variables,
- Get an appreciation of trends and the potential for further optimization,
- Elucidate (if needed) what should be treated as an objective and what as a constraint,
- Identify good starting points for subsequent local exploitation,
- Provide the input data for building a meta-model, if wanted.

The blunt approach for exploration is that of an exhaustive search in which a prescribed number of variants are created by equidistantly changing one parameter at a time. If you have two free variables and want to afford just three guesses in each direction you end up with a total of nine variants (i.e., a grid of three by three). For three variables you already need 27 variants (three by three by three). It is easy to see that this scales up too quickly to be successful for expensive simulations. Consequently, other strategies were developed that deliberately leave out points in the design space that may not be absolutely necessary. Two popular algorithms are the Latin Hypercube sampling (a factorial method) and the Sobol sequence. For an appreciation of the Sobol, see Appendix A.

4.5 Exploitation

As soon as regions of interest in the design space are known from the exploration phase, an exploitation phase – also called formal optimization – is started. Its purpose is to find further designs that outperform all designs known so far. Usually, several optimization runs are conducted. The team selects a handful of the more promising designs from the exploration phase (or simply considers a few manually created starting points). Optimization strategies are then put to use to systematically change the free variables in order to iteratively advance towards (at least local) optima.

Ideally, the exploitation would yield a true (and even a global) optimum. However, resources rarely permit that optimality conditions are strictly met. For mathematical proof first and second derivatives of the objective(s) with respect to all free variables would have to be computed. This is very expensive as it calls for numerical approximations, requiring a lot of additional CFD analyses. As a consequence, a humbler approach is taken in most industry projects. The best variant is simply chosen from several improved designs. Any improvement of the objectives is welcome and happily lived with ever after. Pragmatically, even though nobody actually knows if some superior design was still out there, yet to be discovered, the final shape is at least better than its baseline.

Many different formal optimization strategies are available, e.g. *Birk and Harries (2003)*. At a high level they are categorized into local (mostly deterministic) and global (primarily stochastic) methods. Local methods are often further subdivided into gradient-free and gradient-based methods such as the Simplex algorithm, *Nelder and Mead (1965)*, and the conjugate-gradient method, *Hestenes and Stiefel (1952)*, respectively. So as to understand the principle of formal optimization, the Simplex algorithm, a very popular strategy, is explained in Appendix B.

Global methods subsume many quite diverse strategies of which particle swarm optimization, simulated annealing and genetic algorithms, *Goldberg (1989)*, are very popular. In a way, global methods combine the two phases of exploration and exploitation. As they sweep over the design space, they generate both an overview and design improvements, albeit at the cost of very high numbers of functional evaluations. Due to their need for many hundreds of CFD runs, even for problems with small sets of free variables, global methods are only utilized by teams with exceptional resources.

4.6 Comparison of algorithms

There is no optimal optimization method. Rather, the choice depends on the design task at hand as well as the time and computer power available. Furthermore, the comfort of having worked successfully with one method might just lead to applying it again for the next project. Table II gives an overview of representative algorithms, helping to make a selection. Generic optimization environments, so-called PIDOs, provide a range of tools to choose from. Popular algorithms for practical shape optimization using CFD are the Sobol sequence for exploration and the Tangent Search Method (T-Search) for exploitation as offered within CAESES. The T-Search was originally proposed by *Hilleary (1966)*. It combines smaller steps and larger moves through the design space (a pattern search) and directly handles inequality constraints, *Birk and Harries (2003)*. Mathematically speaking it is a gradient-free method but it comes up with probing moves not dissimilar to gradient directions. The examples shown below are primarily based on employing the Sobol and the T-Search.

5. Examples

Several examples are presented to illustrate practical shape optimization using CFD. The examples stem from both R&D and industry projects realized by coupling CAESES to the various state-of-the art RANS solvers utilized by the respective design teams on a daily basis. Further examples along with animations can be found at www.caeses.com/industries/case-studies.

5.1 Diffusor Design

For a diesel engine, the hot exhaust gas had to be diffused and, in addition, turned by 90° due to constraints in space. The component is relatively simple but occurs repeatedly and, hence, influences the overall performance of the system. STAR-CCM+ by SIEMENS (formerly CD-adapco) was applied to compute pressure loss and flow homogeneity. Both were treated as objectives and considered at the diffusor's outlet plane and further downstream where the hot gas reached the next sub-system.

Table II: Overview of optimization algorithms

Phase	Purpose	Prominent algorithms	Strong points	Weak points
Exploration (Design study)	Understand design space Find good starting points for (local) exploitation Get relationships between objective(s) and free variables (sensitivities) Get appreciation of optimization potential Help decide on objectives and constraints (if unclear) Provide data for meta-models	Sobol sequence Latin Hypercube sampling Taguchi method Exhaustive search (not a true DoE)	Variants are independent of each other and can be evaluated in parallel (robust execution) Derivative free Easy to comprehend (good feeling)	Intrinsically resource intensive for large design spaces with many free variables
Exploitation (Formal optimization)	Find local optima	Simplex algorithm T-Search	Gradient-free (robust execution)	May overshoot and end up in different local region than started from Each design has to be successfully evaluated (sequential execution)
		Conjugate-gradient method	Fast for well-behaved problems (e.g. nearly quadratic behavior)	Gradients need to be approximated numerically, e.g. by means of forward differencing (with challenge to identify appropriate step size) unless realized within a parametric-adjoint simulation Needs line search (or similar) to advance towards local optimum
Combined exploration and exploitation	Understand design space and identify optima Determine Pareto frontier	MOGA (multi-objective genetic algorithm) MOSA (multi-objective simulated annealing) Particle swarm optimization	Variants (of one generation or time step) are independent of each other and can be evaluated in parallel Robust execution (failure of one or several designs does not lead to termination) Generates good insight Easy to apply to multi-objective and multi-disciplinary design tasks	Very expensive to conduct

N.B. Many excellent optimization algorithms have been proposed and the table does not aim at completeness. References for further reading are given in Section 6.

A DoE was conducted on the basis of a fully parametric model that captured all constraints and allowed modifying the volume distribution along the connecting path between the diffusor's inlet and outlet. Fig.8 depicts the baseline, an intermediate design and the improved design along with flow visualizations at the outflow plane for two of the better shapes. The baseline with its simple sweep of circular sections yielded a pressure loss of 47 mbar (flow rate of 1 kg/s at 500°C) while axial flow homogeneity was 0.51 (1 being the ideal). In comparison, the final design with its "Cobra-style" shape, Fig.8, resulted in a substantially reduced pressure loss of only 14 mbar and an increase to 0.87 in homogeneity.

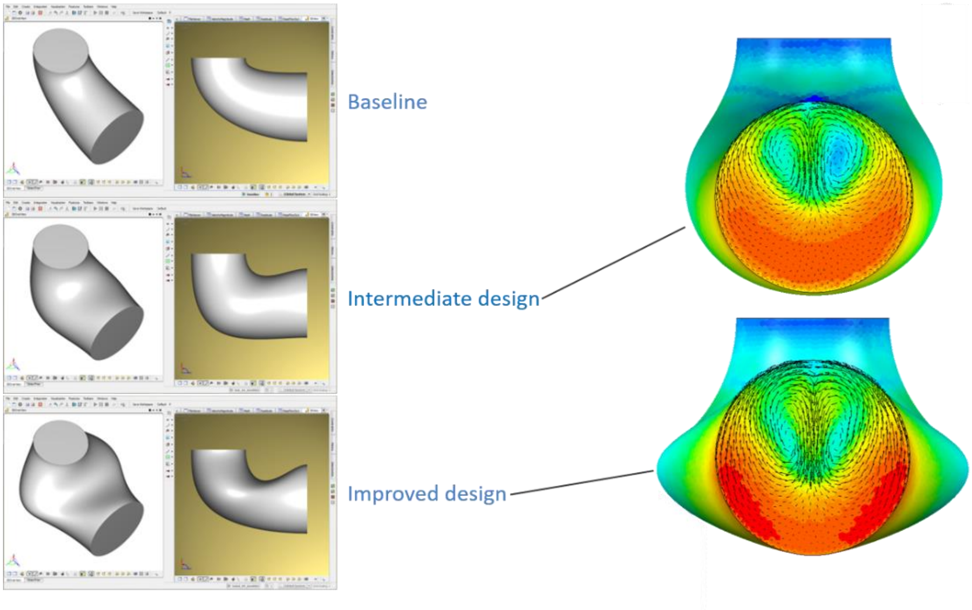


Fig.8: Improvement of a 90° diffusor for a diesel engine

5.2 Funnel Design

For a mega-yacht, the exhaust funnel was studied with regard to gas contamination on the upper deck while at anchor in calm weather, representing the so-called "party condition." The fluid domain was discretized using ICEM while the simulations of the external thermo-fluid dynamic field were performed with ANSYS CFX (www.ansys.com). For details see *Harries and Vesting (2010).*

Fig.9 shows an impression of the yacht along with its plume as emitted from the auxiliary engines, close-ups of the grid and two representative designs. The parametric model within CAESES allowed varying the length and angle of the exhaust pipes along with the size and shape of various deflectors. As an objective for minimization the volume fraction of exhaust gas integrated over a plane downstream of the funnel was considered. An overnight Sobol of 50 variants was run by CAESES on a small cluster so as to identify unfavorable and favorable designs as needed for styling decisions. Longer and steeper funnel pipes in connection with extended and more strongly curved deflectors yield tangible reductions of exhaust gas (the best design being about 40% better than the worst), Fi.9.

5.3 Valve Design

For a project at ARCA Valves an optimization needed to be carried out for a DN700 control valve for liquid and gaseous media, optimized for use as an anti-surge valve on turbocompressors. The target of the optimization was a high flow coefficient Kvs, to exceed the delivery capacity of the turbo-compressor and allow the gas trapped inside the pipelines and heat exchangers to be relaxed in a very short time span upon safety shutdown of the compressor.

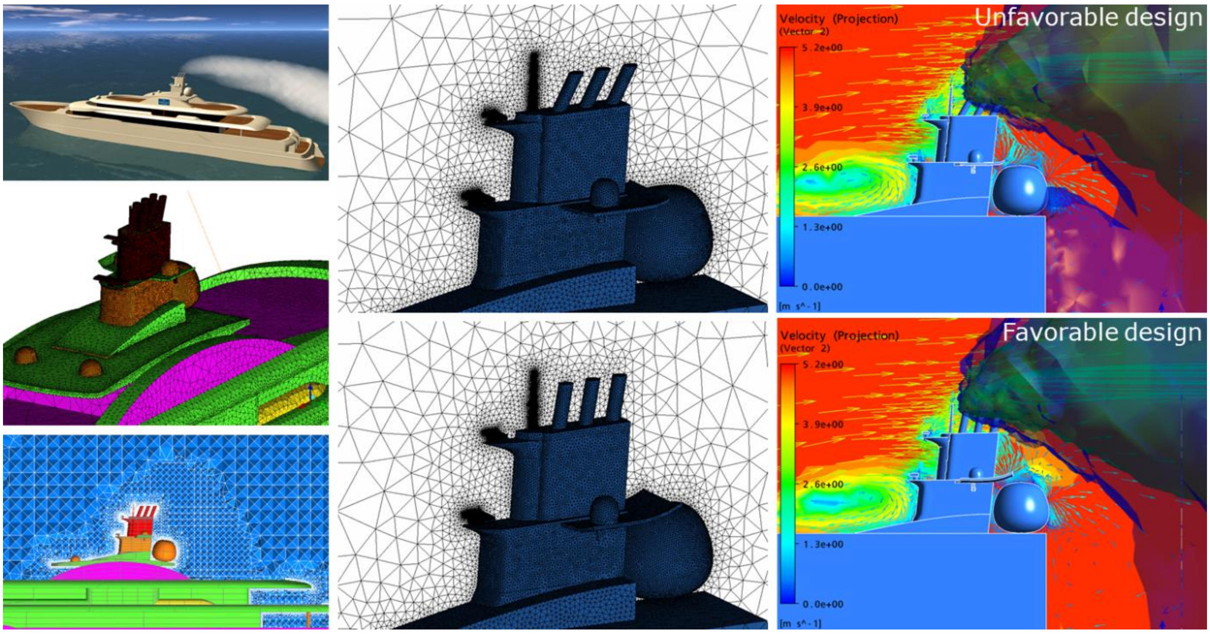


Fig.9: Funnel design study with unfavorable and favorable designs from DoE, *Harries and Vesting (2010)*

Autodesk CFD (www.autodesk.com) was used for the CFD part and a flexible parametric geometry model, including a specified set of constraints (wall thicknesses, flange definitions) was created within CAESES. Inlet and outlet channels of the valve were controlled by parameterized upper and lower contours, resulting in a sweep path, along which an elliptical cross-section was generated. One ellipse axis of the cross-section was defined by the distance of the channel contours while the other one could be varied freely. The flange positions, plug seat ring and cage geometry were fixed, but the seat could be moved vertically.

The optimization process proceeded from an initial DoE over a series of small exploitation runs. This led to the identification of certain distinct trends in the evolution of the shape to better performing designs (most notably, the overhanging outlet cavity as shown in Fig.10), which prompted the addition of further parameters and freedom to the geometry model. This revised model was then put through a further brief optimization phase until a final design was selected.

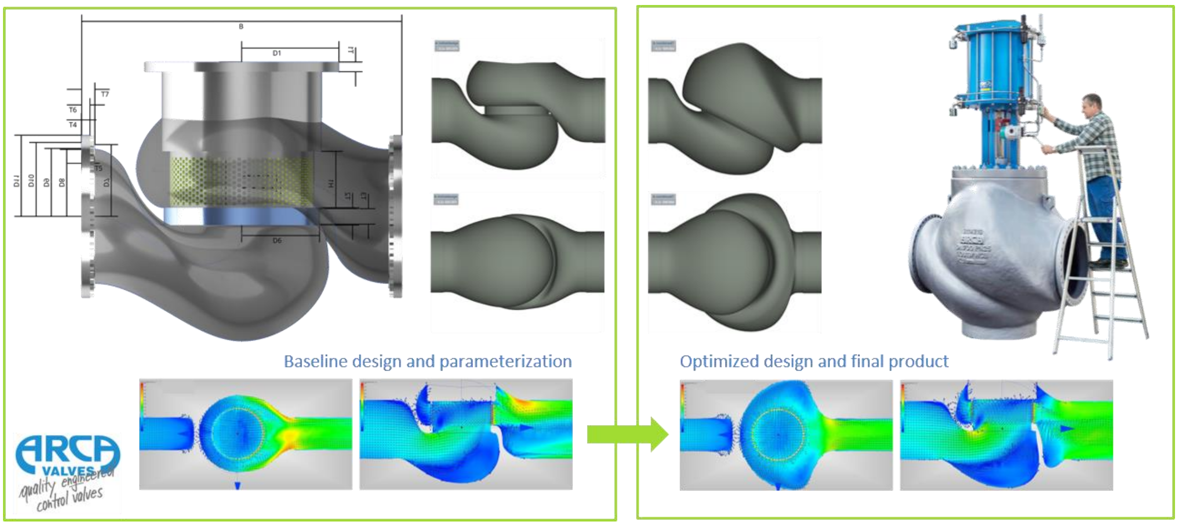


Fig.10: Baseline (left) and optimized (right) design for a large control valve, www.arca.de; see also https://www.caeses.com/industries/case-studies/control-valve-optimization/

The approach allowed for the exploration of unconventional concepts, reaching an improvement of the flow coefficient Kvs by approximately 25% compared to the initial rather conventional design. The target Kvs was surpassed by approximately 3%. Simultaneously, the development time could be greatly shortened (about two weeks for setting up and running the optimization) while greatly reducing the need for physical testing.

5.4 Radial Compressor Design

The volute and diffusor of a radial compressor were optimized for isentropic efficiency, *Ratz (2014)*. CAESES was coupled to software systems by Numeca Int. (www.numeca.com), specifically, Autogrid5 and Hexpress for compound grid generation and FINE/Open for flow simulation. The geometry was described in CAESES by means of a fully parametric model. The volute's cross section was defined by several parameters (such as the radius as a function of peripheral angle) which themselves were subjected to change in circumferential direction, cp. Fig.2. This enabled shape adaptations for given mass flow and allowed controlling the volute's diffusion characteristics. The parametric model of the diffusor comprised the stagger angle, the blades' twist and their pitch and trailing edge positions. These parameters were changed circumferentially, too, allowing the diffusor to be non-periodic (according to a user-specified "normal" distribution). The impeller was left unchanged but accounted for in the simulation.

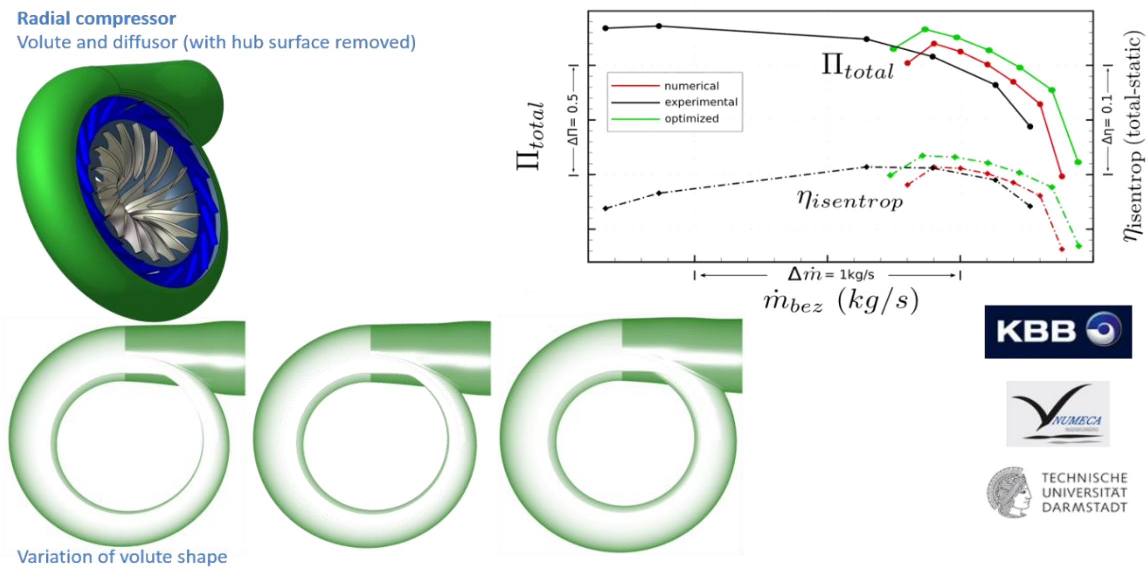


Fig.11: Optimization of a radial compressor with parametric variations of volute (green) and diffusor (blue); courtesy of www.tu-darmstadt.de, www.kbb-turbo.de

During the optimization CAESES was run in batch-mode to provide the variable geometry while Numeca's Design3D was chosen to control the process, using a genetic algorithm in combination with an artificial neural network as a meta-model. Fig.11 illustrates the project. Comparing the baseline to the best design, a small rotational shift (by 4.2°) was observed between the original diffusor blades and the newly identified optimal blade configuration. Further to this, the stagger angle was increased (by 1.4°) and the position of the trailing edge was moved in radial direction (by 1.5 mm). The volute of the optimal design featured a slightly larger area distribution. Isentropic (peak) efficiency was increased by 1% (at 5.9 kg/s) with performance improvements over the entire operating range.

5.5 Motorboat Optimization

Fig.12 shows the results from an optimization campaign in which a fully parametric model of a classic motorboat, inspired by the 1960s Riva Junior, was systematically varied and improved for calm water resistance at 18 kn (i.e., standard waterskiing speed). A fully parametric model was built in CAESES, allowing the change of rocker, spray rails, the hollowness of both the transom stern and central parts of the underwater hull, the beam etc. The displacement volume was maintained during the variations. The viscous free-surface simulations in calm water were undertaken with FINETM/Marine by Numeca Int., https://www.numeca.com/. By running a Design-of-Experiment, building a surrogate model and doing various deterministic searches, an improvement of total resistance of about 7% could be realized. The best design performs better than the baseline from around 11 kn to above 19 kn, Fig.12. Details are given in *Albert et al. (2016)*.

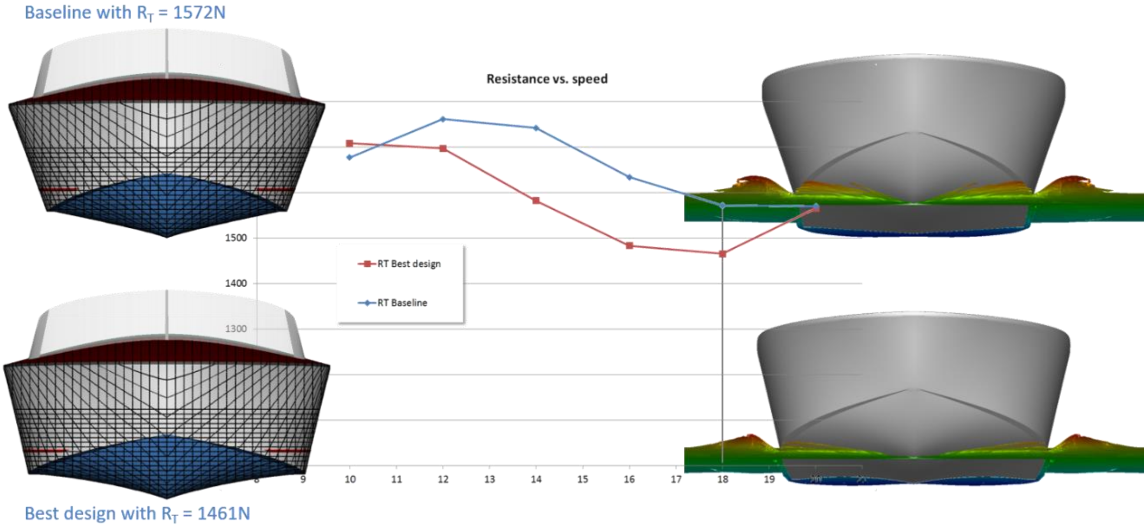


Fig.12: Baseline vs. best design of a motorboat at 18 kn

5.6 Piston Bowl optimization

A piston bowl is a recess in the piston crown used in (direct injection) Diesel engines. The shape of the piston bowl determines the combustion chamber and influences the air-fuel mixture during the compression stroke. A good mixture leads to an efficient combustion, resulting in more power and/or better fuel economy. Additionally, a well-designed piston bowl leads to a reduction of the in-cylinder emissions, in particular NOx and soot, as well as the costs for after-treatment.

The flow and combustion process were computed with CONVERGE, https://convergecfd.com/. The piston bowl was modeled within CAESES® and parameterized such that the compression ratio could be automatically adjusted for each variant. This ensures that during an optimization campaign every variant would feature the same compression ratio, avoiding the generation of infeasible shapes. To do so an internal optimization was set up in which additional parameters were modified to yield the correct compression ratio before any of the simulations were started.
Fig.13 summarizes a representative optimization campaign for a generic piston bowl (without any periodicity, different to the “wavy” piston bowl shown in Fig.2). Four free variables were used, namely the lower radius of the bowl, the size of the lip, the bowl’s inner diameter and the spray angle. While the first three of these variables govern the shape (with additional parameters for the adjustment of the compression ratio as outlined above), the last variable serves as an input to the CFD simulation, controlling the direction in which the fuel is injected. Two objectives were considered, namely the NOx emissions and the soot. The two objectives being of opposing character, three selected designs from the Pareto frontier are shown in Fig.13, i.e., the design with lowest NOx, the design with lowest soot and a trade-off.

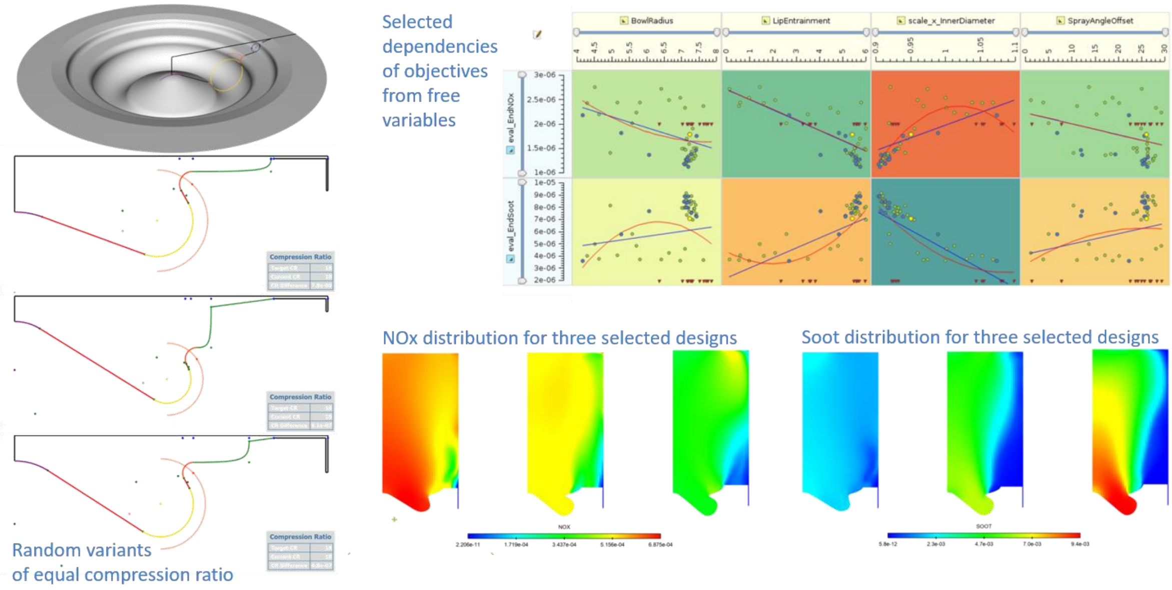


Fig.13: Optimization of a piston bowl for opposing objectives of NOx and soot; https://www.caeses.com/blog/2017/piston-bowl-design-and-optimization/

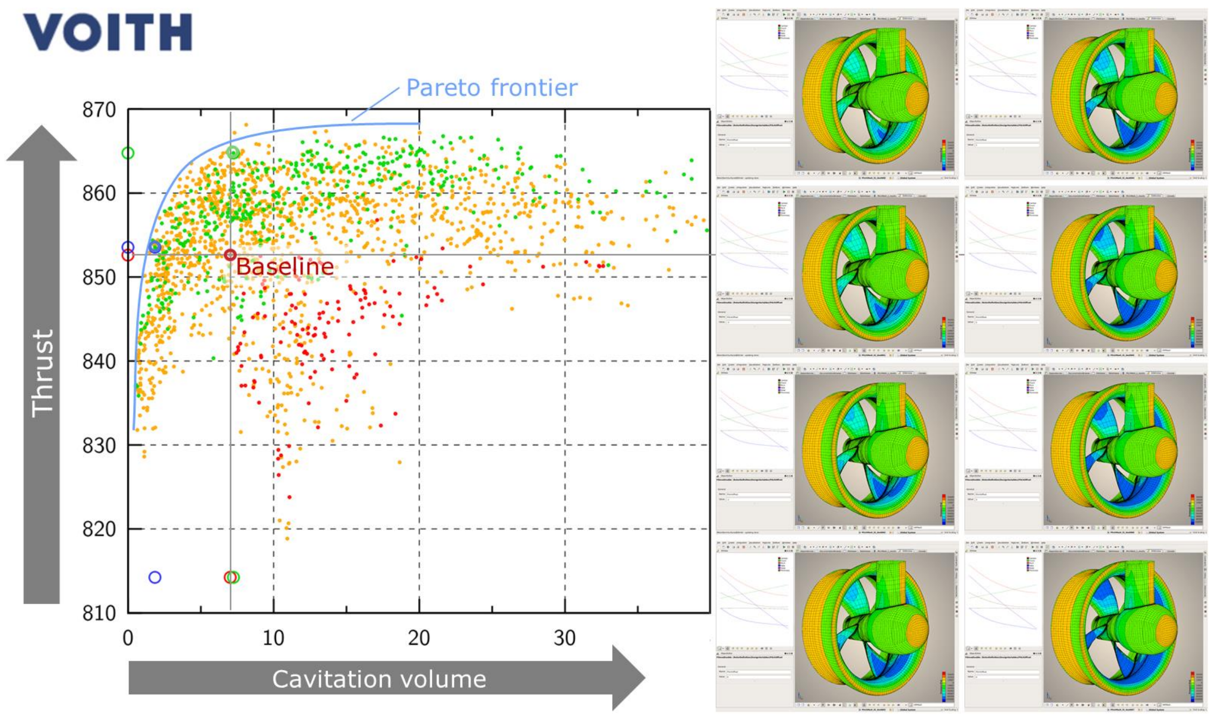


Fig.14: Performance for competing objectives for hundreds of variants along with representative instances of the propulsion system; courtesy of Voith Turbo (www.voith.com)

5.7 Propulsion System Design

Voith developed a propulsion system for dynamic positioning of large offshore structures using CAESES. The system comprises a ducted propeller, a nozzle and a strut, Fig.14. Starting from a manually designed baseline (red circle in Fig.14), many variants were investigated at a time, finally resulting in several thousand variants analyzed with RANS simulations. The parametric model for the propeller blade contained descriptors for camber, chord length, pitch, rake, skew and thickness, while the duct was defined via proprietary profiles. The shape and inclination of the nozzle were parameterized and varied, too. As is common in maritime propulsion, thrust and cavitation volume are two important objectives, bringing about a multi-objective problem (as further explained in Section 6). Designs were found that yielded higher thrust for the same cavitation volume (e.g. green circle). Vice versa, designs were found that gave the same thrust for lower cavitation volumes (e.g. blue circle). The quadrant left and above of the baseline offers designs better with respect to both objectives (creating a Pareto frontier as depicted in light blue).

6. Further issues

6.1 Multi-disciplinary optimization

To start with, the world is multi-objective and so is design. Sometimes it is possible to express all objectives in the same unit (mostly in economic terms) or to normalize and mix objectives that are of different nature. Very frequently, however, objectives cannot be made directly comparable. To solve this predicament, Pareto frontiers can be studied, Fig.14. If a feasible design cannot be further improved for one objective without deteriorating any other objective, the solution belongs to the Pareto set. All these so-called non-dominated designs are respectable candidates for selection and it depends on the preference of the decision makers which one is considered to be best for the design task at hand. Utility functions may serve to rank variants.

Multi-objective optimization often refers to looking at several objectives that belong to the same engineering discipline, say hydrodynamics. In naval architecture a typical multi-objective problem is to consider resistance, propulsion and seakeeping at the same time. A multi-disciplinary optimization is established as soon as several disciplines are involved, say hydrodynamics, structures and economics. An engineering example of multi-disciplinary optimization is presented in *Harries et al. (2011)*. A tanker was optimized for objectives such as fuel consumption, strength, oil outflow in case of accidents (OOI as a measure for safety), required freight rate (RFR as a combined economic target) and energy efficiency design index (EEDI as a regulatory requirement). The design synthesis model called for the combination of quite a few tools, each covering a specific field of engineering and being controlled by CAESES. Fig.15 gives the results aggregated from a DoE with RFR vs. deadweight (i.e. a ship's carrying capacity). There is no single answer but that different designs perform best for different preferences.

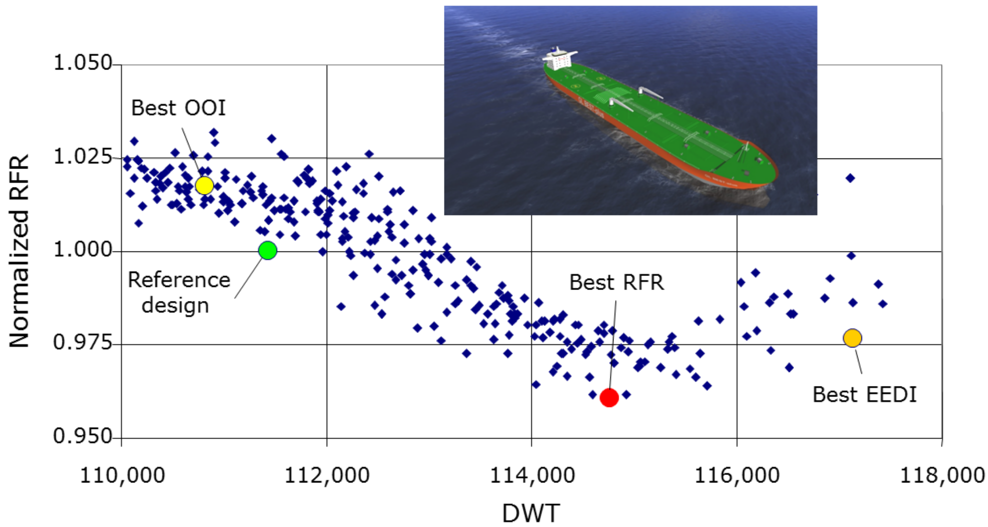
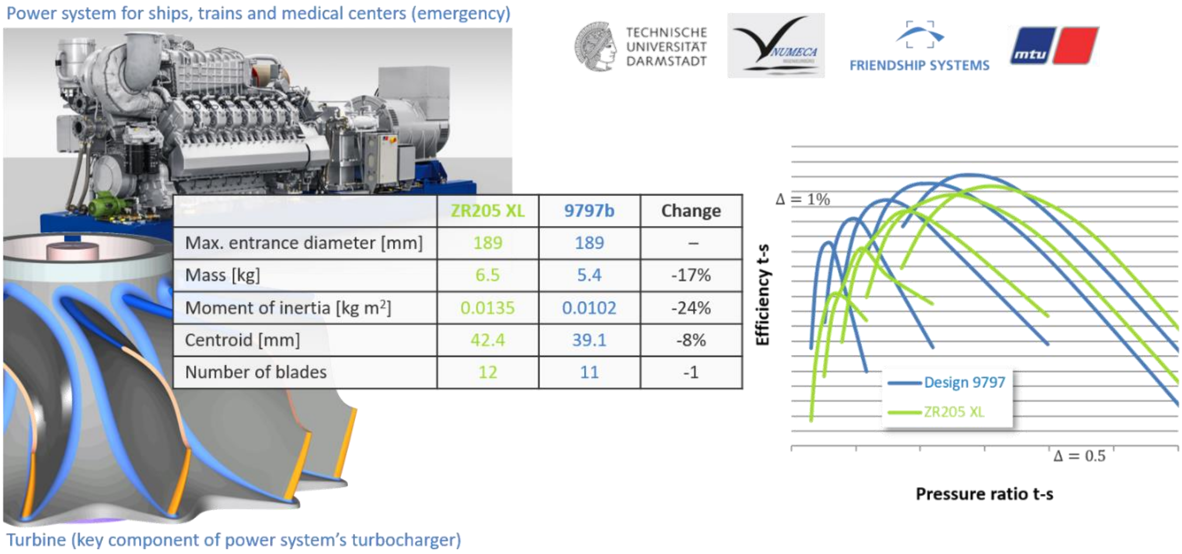


Fig.15: Results from a multi-disciplinary optimization of an Aframax tanker, *Harries et al. (2011)*

Further examples of design and optimization of ships are reported in *Harries et al. (2017,2019)* for a RoPAX ferry and other ship types. The applications stem from a comprehensive R&D project (www.HOLISHIP.eu) in which the idea of multi-disciplinary optimization is taken one step further by considering all relevant aspects of maritime assets in a holistic approach, *Harries and Abt (2019)*.

Finally, Fig.16 shows the results from a multi-disciplinary optimization of a turbocharger. They stem from a comprehensive R&D project led by MTU Friedrichshafen, www.mtu-solutions.com, bringing together the expertise of MTU, TU Darmstadt, www.glr.tu-darmstadt.de, Numeca Ingenieurbüro, www.numeca.de and FRIENDSHIP SYSTEMS, www.friendship-systems.com. Both the compressor and the turbine of a turbocharger for a power system by MTU were optimized, *Fröhlig et al. (2019).* For the turbine as a key component all aspects of thermo-fluid dynamics and structural performance were simultaneously taken into account. Starting from a very good baseline, i.e., ZR205 XL (shown in green), major improvements could be realized, Fig.16. Design 9797b (shown in blue) not only increases efficiency and the performance map but also features lower mass (-17%), decreased moment of inertia (-24%), a reduced centroid (-8%) and a smaller number of blades.

While previous single-objective optimizations had shown that any improvement in efficiency were frequently lost when manually adjusting the most favorable designs for structural integrity, the comprehensive multi-disciplinary optimizations would safeguard that no constraints were violated, leading to highly competitive designs that could not be realized manually.



	ZR205 XL	9797b	Change
Max. entrance diameter [mm]	189	189	–
Mass [kg]	6.5	5.4	-17%
Moment of inertia [kg m²]	0.0135	0.0102	-24%
Centroid [mm]	42.4	39.1	-8%
Number of blades	12	11	-1

Fig.16: Results from a multi-disciplinary optimization of a turbocharger's turbine impeller

6.2 Robustness

An additional issue of any optimization is the sensitivity of a design with regard to smaller (e.g. noise) and larger differences (e.g. operational profile) between the idealized situation and the real world. There is a natural contradiction between optimality and robustness. A solution may be exceptionally good for one objective in one given situation while performance quickly drops if conditions change just a bit. Clearly, this is something to avoid. Robust optimization is a field of continuous research and development, *Beyer and Sendhoff (2007).* An ad-hoc solution is to formulate a range of conditions for which the design shall perform really well, essentially creating a multi-objective problem.

Sometimes, this is again cast into a single-objective problem in which several objectives are normalized (e.g. with baseline performance), weighted (e.g. according to frequency distribution) and summed up. For example, a ship typically sails at different speeds and drafts. A handful of loading conditions are combined so as to yield the objective for a robust optimal solution. Similarly, the blades of a turbine have to be really good at design point but also need to deliver a large operating range. Again, so as to avoid overbreeding a few representative conditions can be taken into account simultaneously.

Sensitivities also need to be considered in the context of noise and deterioration. The former is typically associated with fluctuating conditions such as spread in material property and deviations in production accuracy. The latter may simply stem from the wear and tear encountered after years of work. Therefore, it needs to be judged if the ranking of variants will actually remain true even if conditions slightly differ or change. Let us take again an example from the maritime industry. In practical hull form optimization naval architects neglect hull fouling. It is assumed that hull resistance will increase by about the same amount for all variants. As a consequence, the hull shape that yields best performance when hydrodynamically smooth should still be the best hull after surface roughness has come into play since all variants are equally affected.

More issues and questions, e.g. low-fidelity vs. high-fidelity simulations, are posed in *Harries (2008)*. There are countless papers and many books on optimization, e.g. *Eschenauer et al. (1990), Birk and Harries (2003), Onwubolu and Babu (2010)*.

6.3 Speed-up

A common predicament in design and optimization is that simulations often require substantial resources (expertise, computer hardware and software licenses to mention just a few). As illustrated in Fig.6, simulation efforts quickly scale up with the degrees-of-freedom. If the analysis of a single variant takes a few hours, even if employing an HPC, it is obvious that several hundred of variants take a lot of time to investigate. Consequently, excellent parametric models are needed that reduce a system's DoF as much as possible. In addition, there are some new and promising developments to introduce further speed-up when undertaking an optimization campaign. One approach is to combine parametric modeling with adjoint simulation, see Fig.5 and Tab.I. Another approach is to reduce the dimensionality of the design space without forsaking too much of the system's variability.

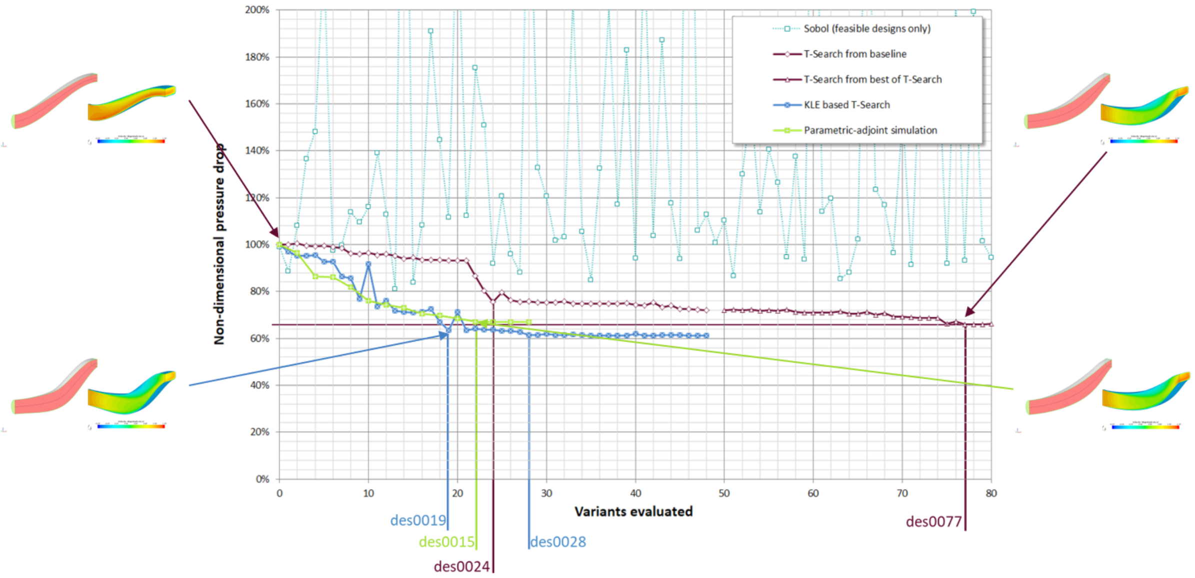


Fig.17: Speed-up of optimization campaigns by parametric-adjoint simulation as well as KLE based parameter space reduction, *Harries and Abt (2019)*

Both approaches are thoroughly discussed in *Harries and Abt (2019)* for the example of a duct optimized for pressure drop, combining CAESES with STAR-CCM+. Fig.17 reproduces the optimization history of the different approaches, giving a clear indication how much faster a local search by parametric-adjoint simulation or a global search within a design space of reduced dimensionality, here on the basis of a Karhunen-Loève Expansion (KLE), can be when compared to a state-of-the-art parametric approach. The final designs selected perform similarly with regard to the objective and the shapes do not differ a lot, Fig.17. However, the associated simulation effort is considerably higher for the standard approach of parametric shape optimization. (Naturally, the associated mathematics for the parametric-adjoint simulation and for the KLE are non-trivial. Nevertheless, once implemented this is not so critical for the practicing engineer.)

The approach of design space dimensionality reduction goes back to *Diez et al. (2015)* with further improvements introduced in *Diez et al. (2016)* and *D'Agostino et al. (2017)*. An example for a SWATH (small waterplane area twin hull) was given by *Pellegrini et al. (2018)*, while *Böttcher (2020)* shows the potential of dimensionality reduction for a turbocharger.

7. Summary

Different approaches of shape optimization using Computational Fluid Dynamics have been presented. Parameter-based optimization is the most popular approach and is applied during both initial and detailed design. Increasingly, shape optimization using CFD is done early on (i.e., upfront) as the insight and the potential for performance gains are highest. In order to be able to undertake optimization projects special parametric models are required, focusing on engineering rather than production. They encompass partially parametric (e.g. free-form deformation) and fully parametric models. The former use existing shapes which are modified parametrically. The latter allow the creation and variation of shapes from scratch. Dedicated CAD software for variable geometry complements the traditional CAD tools, offering parametric modeling as needed for shape optimization.

Parameter-based optimizations frequently comprise exploration (design study) and exploitation (formal optimization). During the exploration phase, a reasonable number of shape variants is created and analyzed with the main aim of understanding the design space, identifying favorable improvements, evaluating sensitivities and, possibly, building of response surfaces. Design-of-Experiments are mathematical strategies to undertake explorations systematically and economically. During the exploitation phase, the shape is modified methodically with the aim of further improving one or several objectives (i.e., performance measures) that are considered to be representative of the quality of the product. Quite a range of strategies are available. Which strategy best suits a particular design task depends on many things, e.g. the simulation time needed, the available resources and the conditioning of objectives and constraints.

Several examples have been given to illustrate applications from different fields of engineering. Shape optimization using CFD is prominent in the automotive, aerospace and maritime industries as well as in turbomachinery and mechanical engineering. This is due to the fact that concerted shape changes can bring about decisive improvements for typical objectives such as resistance and lift, flow homogeneity and pressure drop. In general, energy efficiency is a major driving force for shape optimization, in particular to reduce our environmental footprint.

Acknowledgements

I would like to thank the team at FRIENDSHIP SYSTEMS, in particular current and former key developers and application engineers of CAESES®: Arne Bergmann, Erik Bergmann, Hedi Böttcher, Mattia Brenner, Ceyhan Erdem, Carsten Fütterer, Konrad Lorentz, Jörg Palluch, Park Daehwan, Mike Saroch, Tim Smit, Dr. Sebastian Uharek, Heinrich von Zadow, Stefan Wunderlich, and Benjamin Zeitz. Furthermore, I would like to acknowledge the contributions by my long-term colleague Claus Abt, not only in writing this whitepaper but also in pushing shape optimization using CFD. Finally, I would like to thank the companies and their representatives that provided the examples picked for illustration.

References

ABT, C.; HARRIES, S. (2007), *A New Approach to Integration of CAD and CFD for Naval Architects*, 6th COMPIT Conf., Cortona

ALBERT, S.; HARRIES, S.; HILDEBRANDT, T.; REYER, M. (2016), *Hydrodynamic Optimization of a Power Boat in the Cloud*, HIPER Conf., Cortona

BERGMANN, E.; FÜTTERER, C.; HARRIES, S.; PALLUCH, J. (2018), *Massive Parameter Reduction for faster Fluid-dynamic Optimization of Shapes*, Int. CAE Conf., Vicenza

BEYER, H.G.; SENDHOFF, B. (2007), *Robust optimization – A comprehensive study*, Computer Methods in Applied Mechanics and Engineering 196, pp.3190-3218

BIRK, L.; HARRIES, S. (Ed.) (2003), *OPTIMISTIC – Optimization in Marine Design*, Mensch & Buch Verlag, ISBN 3-89820-514-2, Berlin

BÖTTCHER, H. (2020), *Simulation-driven Design of a Compressor based on Dimensionality Reduction*, Master Thesis, TU Berlin, https://www.friendship-systems.com/contact-us/

BRENNER, M.; ZIMMER, A.; WEICKGENANNT, S. (2014), *Effiziente Optimierung von strömungsbeeinflussten Geometrien durch Übertragung von adjungierten Sensitivitäten auf die CAD Modellparameter*, NAFEMS Konferenz: Berechnung und Simulation – Anwendungen, Entwicklungen, Trends, Bamberg

CHRISTAKOPOULOS, F.; MÜLLER, J.-D. (2012), *Discrete adjoint CFD solver for first and second derivatives via Automatic Differentiation*, Conf. Industrial Design Optimisation for Fluid Flow, Munich

D'AGOSTINO, D.; SERANI, A.; CAMPANA, E.F.; DIEZ, M. (2017), *Nonlinear methods for design-space dimensionality reduction in shape optimization*, 3rd Int. Workshop on Machine Learning, Optimization, and Big Data, Volterra, pp.121-132

DIEZ, M.; SERANI, A.; STERN, F.; CAMPANA, E.F. (2016), *Combined Geometry and Physics Based Method for Design-Space Dimensionality Reduction in Hydrodynamic Shape Optimization*, 31st Symp. Naval Hydrodynamics, Monterey

DIEZ, M.; CAMPAGNA, E.F.; STERN, F. (2015), *Design-space dimensionality reduction in shape optimization by Karhunen-Loève expansion*, Comput. Methods Appl. Mech. Eng. 283, pp.1525-1544

ESCHENAUER, H.; KOSKI, J.; OSYCZKA, A. (Ed.) (1990), *Multicriteria Design Optimization – Procedures and Applications*, Springer

FRÖHLIG, F.; LACHENMAIER, N.; HARRIES, S.; FÜTTERER, C.; RATZ, J.; WENDL, K.; HILDEBRANDT, T. (2019), *Optimizing Turbochargers for High-performance Engines*, EnginSoft Newsletter 16/3

GILL, P.E.; MURRAY, W.; WRIGHT, M.H. (1982), *Practical Optimization*, Academic Press

GOLDBERG, D.E. (1989), *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison Wesley

HARRIES, S. (2008), *Serious Play in Ship Design*, Coll. Tradition and Future of Ship Design in Berlin, TU Berlin

HARRIES, S. (2010), *Investigating Multi-dimensional Design Spaces Using First Principle Methods*, 7th HIPER Conf., Melbourne

HARRIES, S.; ABT, C. (2019a), *CAESES – The HOLISHIP Platform for Process Integration and Design Optimization*, A Holistic Approach to Ship Design – Vol. 1: Optimisation of Ship Design and Operation for Life Cycle, Springer

HARRIES, S; ABT, C. (2019b), *Faster turn-around times for the design and optimization of functional surfaces*, Ocean Engineering 193

HARRIES, S.; ABT, C.; BRENNER, M. (2015), *Upfront CAD – Parametric modelling techniques for shape optimization*, Int. Conf. Evolutionary and Deterministic Methods for Design, Optimization and Control with Applications to Industrial and Societal Problems (EUROGEN), Glasgow

HARRIES, S.; ABT, C.; HOCHKIRCH, K. (2004), *Modeling meets Simulation – Process Integration to improve Design*, Honorary Colloquium for Prof. Hagen, Prof. Schlüter and Prof. Thiel, Duisburg

HARRIES, S.; CAU, C.; MARZI, J.; KRAUS, A.; PAPANIKOLAOU, A.; ZARAPHONITIS, G. (2017), *Software Platform for the Holistic Design and Optimisation of Ships*, Jahresbuch Schiffbautechnische Gesellschaft, Bd.111, Springer

HARRIES, S.; DAFERMOS, G.; KANELLOPOULOU, A.; FLOREAN, M.; GATCHELL, S.; KAHVA, E.; MACEDO, P. (2019), *Approach to Holistic Ship Design – Methods and Examples*, 18th COMPIT Conf., Tullamore

HARRIES, S.; TILLIG, F.; WILKEN, M.; ZARAPHONITIS, G. (2011), *An Integrated Approach for Simulation in the Early Ship Design of a Tanker*, 10th COMPIT Conf., Berlin

HARRIES, S.; VESTING, F. (2010), *Aerodynamic Optimization of Superstructures and Components*, 9th COMPIT Conf., Gubbio

HESTENES, M.R.; STIEFEL, E. (1952), *Methods of Conjugate Gradients for Solving Linear Systems*, J. Research of the National Bureau of Standards 49/6

HILLEARY, R. (1966), *The Tangent Search Method of constrained minimization*, Tech.Rep./Res. Paper 59, United States Naval Postgraduate School, Monterey

KRÖGER, J.; RUNG, T. (2015), *CAD-free hydrodynamic optimisation using consistent kernel-based sensitivity filtering*, Ship Technology Research 62, pp.111-130

NELDER, J.A.; MEAD, R. (1965), *A Simplex method for function minimization*, Computer J. 7, pp.308-313

ONWUBOLU, G.C.; BABU, B.V. (2010), *New Optimization Techniques in Engineering*, Springer

PELLEGRINI, R.; SERANI, A.; BROGLIA, R.; DIEZ, M.; HARRIES, S. (2018), *Resistance and Payload Optimization of a Sea Vehicle by Adaptive Multi-Fidelity Metamodeling*, AIAA/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conf.

PRESS, W.H.; TEUKOLSKY, S.A.; VETTERLING, W.T.; FLANNERY, B.P. (2007), *Numerical Recipes: The Art of Scientific Computing*, Cambridge Univ. Press

ROBINSON, T.T.; ARMSTRONG, C.G.; CHUA, H.S.; OTHMER, C.; GRAHS, T. (2012), *Optimizing Parameterized CAD Geometries Using Sensitivities Based on Adjoint Functions*, Computer-Aided Design & Applications 9/3, pp.253-268

RATZ, J. (2014), *Simultaneous optimization of diffuser with volute for applications in radial compressors*, Master Thesis (in German), TU Darmstadt

SEDERBERG, T.W.; PARRY, S.R. (1986), *Free-form deformation of solid geometric models*, SIGGRAPH Computer Graphics (ACM) 20/4, pp.151-160

SIEBERTZ, K.; VAN BEBBER, D.; HOCHKIRCHEN, T. (2010), *Statistische Versuchsplanung, Design of Experiments*, Springer

STEPHAN, M.; HÄUßLER, P.; SCHÜTZLE, T.; JOHANNSEN, D. (2008), *An Efficient Approach for CFD Topology Optimization*, NAFEMS, Wiesbaden

STÜCK, A.; KRÖGER, J.; RUNG, T. (2011), *Adjoint-based Hull Design for Wake Optimisation*, Ship Technology Research 58/1, pp.34-44

Appendix A: Sobol sequence for design studies (exploration)

The Sobol is a so-called quasi-random low-discrepancy sequence, which means that it produces a pattern in design space that looks random even though it actually is deterministic. The Sobol sequence fills up the design space (namely a hyperspace of *n* dimensions) such that the next variant is always placed in the region that is least populated so far, *Press et al. (2007).* Its behavior somewhat mimics people at a beach. Consider the beach to be a two-dimensional search space. The first person to arrive will probably place his or her towel somewhere close to the center. The next person will choose a spot far away from the first one (unless there is some non-mathematical attraction that we shall disregard). The third person will do likewise, now trying to lie down at a comfortable distance to both the first and the second person. This carries on until the beach is somewhat covered, Fig.A-1. (Similar observations can be made in elevators where strangers naturally find an unbiased distance towards each other, even dynamically by reshuffling every time someone comes in or goes out.)

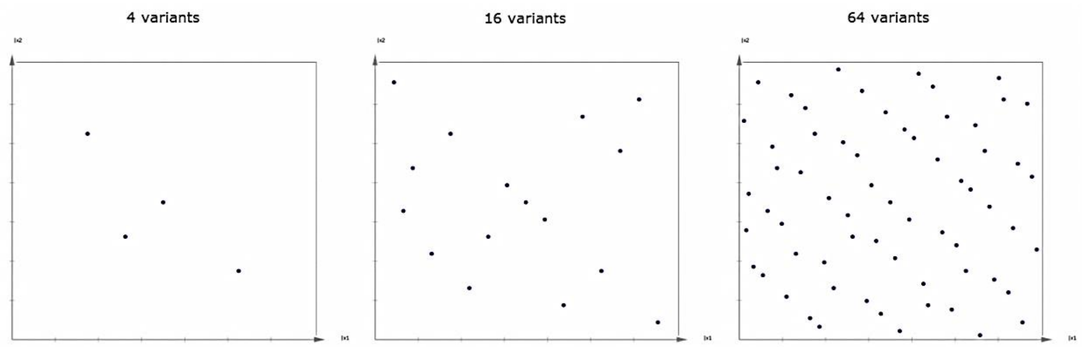


Fig.A-1: Different number of variants shown in two-dimensional space as generated by a Sobol Design-of-Experiment (Note that in three- and higher-dimensional spaces the same sequence appears in the two-dimensional projection)

However, unlike people at the beach the Sobol is a mathematical strategy that comes without any true randomness. This is very convenient as it easily permits repeating a specific sequence. E.g., after running an expensive study you realize that some outputs of interest are missing. It is then straightforward to repeat the Sobol, getting exactly the same parameter values as before, only for the additional analyses. Another strong suit is that large investigations can be subdivided into more manageable portions. E.g., you run the first 50 variants (0 to 49) on one workstation and the next 50 variants (50 to 99) on another computer. Both patterns combined behave as one large search, meaning exactly the same as if you had done all variants in one single sweep (0 to 99). Naturally, this behavior can also be utilized for a successive extension of the exploration phase. If you feel you need more information after 100 variants have been completed you can just carry on with the next 100 without any overlap. Moreover, since every variant is independent of any other the entire DoE can be executed in parallel.

Another attractive side effect of this independence is that certain constraints can be checked for each variant before actually triggering any expensive CFD runs (e.g. threshold on grid quality), saving the resources for simulations where they are most valuable. In addition, it does not really hurt so much if the evaluation of one or several variants fails (e.g. due to a meshing problem or a diverging flow solution) as this would not terminate the process.

The algorithm tends to stay away from the lower and upper bounds of each free variable and only if many variants are generated does it come close to the bounds, cp. Fig.A-1 with four variants (left diagram) and with 64 variants (right diagram). A reasonable way to cope with this potential shortcoming is to set the bounds slightly larger than initially required. Finally, since a Sobol is typically used to explore an *n*-dimensional space, with *n* much higher than 2, one should not be too picky about the apparent unevenness in the distribution of variants when projected into two dimensions; the other *n-2* dimensions are simply not shown.

Appendix B: Simplex algorithm for formal optimization (exploitation)

The Simplex algorithm is an easy yet representative deterministic strategy. In *n*-dimensional space a Simplex is an object of *n+1* corner points (a triangle in 2d, a tetrahedron in 3d etc.). Let us again take a beach for illustration. It is a two-dimensional domain and, hence, we would have two plus one corner points to work with. If the objective was to find the highest dune (giving the best view) the Simplex algorithm goes about as follows: Three people are placed reasonably close to each other, say 5m apart on a beach of 100m by 100m. They hold three flexible ropes, forming a triangle. For our local search the person who is at the lowest position with respect to sea level has to change position while the other two persons stay where they are. The person to move, naturally hoping to gain altitude, now walks towards the rope connecting the two stationary persons, crosses the rope in the middle and carries on to the other side until a new triangle of similar size has been established (so-called reflection). The objective, here the altitude, has now to be determined for the new position. In principle, the search is then repeated, namely the person that now happens to be at lowest altitude needs to move. This is done again and again until the highest point is found (or the maximum number of iterations has been reached). There are several situations which require special treatment, namely expansion (to realize faster gains), contraction (to counteract overshooting) and compression (to reduce the size of the Simplex and, eventually, to terminate the search).

The initial Simplex is typically derived from the exploration phase. For the beach we would ask all visitors, e.g. quasi-randomly distributed according to the Sobol, Appendix A, what their individual altitudes are. Heuristically, the person that is at the highest position so far makes a sensible starting point. The *n* closest neighbors may complement the Simplex, provided their distances relative to the size of the design space and with respect to the behavior of the objective are reasonable (not too small and not too large). Alternatively, little side-steps (say 5m at the beach) can be made by changing one free variable at a time (i.e., going parallel to the coordinate axes). This, however, requires *n* additional evaluations of the objective. Intuitively, it makes sense to undertake a handful of local searches in different areas wherever further gains can be expected from good candidates found during the exploration.

Appendix C - CAESES®

All examples shown in the paper were realized with CAESES®. Download the latest version from www.caeses.com (for both Windows and Linux) for a trial. CAESES comprises robust variable geometry, pre-processing, software connection to external simulation codes, post-processing and optimization & assessment, Fig.C-1. CAESES does not offer any CFD code but rather connects to any simulation code that can be run in batch mode, Fig.C-2.

Robust variable geometry as needed for simulation-driven design differs from traditional CAD. Traditional CAD focuses on production-level geometry, complex assemblies, manufacturing details as well as PLM. Consequently, CAD models contain many details that are required for production and relevant to the final product, but not important (and sometimes even detrimental) for the simulations of interest, *Harries et al. (2018)*. When pre-processing traditional CAD models for CFD mesh generation, they usually have to be de-featured and the "wetted" surfaces have to be extracted, cleaned up and repaired (e.g. making them watertight). This often is a manual process. In addition, changing one or several parameters in a traditional CAD system may lead to regeneration failures, *Harries and Abt (2019)*. Frequently, failed designs can be fixed manually within the CAD system without a lot of effort. However, too many "broken" variants (failure rates of 25% and higher have been observed) cause a real problem for the automated processes of shape optimization.

The specialized CAD engine as provided within CAESES is a flexible surface modeler that addresses the modeling of the geometry relevant and ready-to-go for the simulation(s). It has a strong focus on geometry variation via fully-parametric and partially-parametric modeling, or any hybrid thereof, so that, once a parametric model has been set up, it provides variants consistently (with regard to topology and name tags) and robustly (with regard to watertight CAD entities).

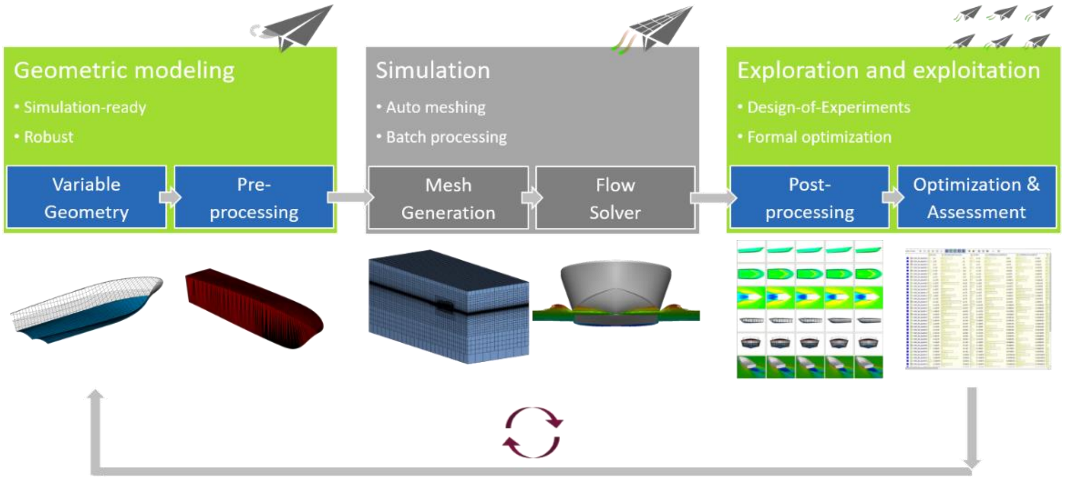


Fig.C-1: Illustration of components of an optimization process using CFD [see also Fig.12]

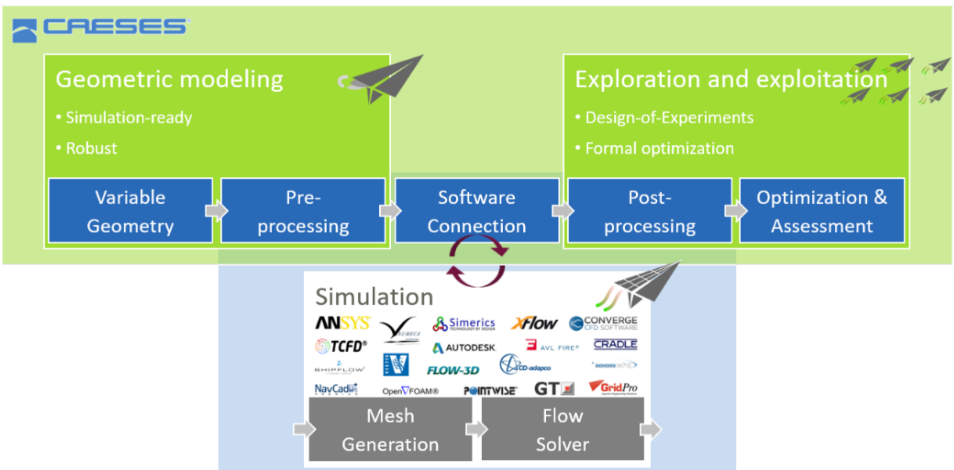


Fig.C-2: CAESES for practical optimization using external CFD codes

Opportunities of Using Open-Source FEA Software in Ship Structural Design

Marin Palaversa, University of Zagreb, Zagreb/Croatia, mpalaversa@fsb.hr
Pero Prebeg, University of Zagreb, Zagreb/Croatia, pprebeg@fsb.hr
Jerolim Andric, University of Zagreb, Zagreb/Croatia, jandric@fsb.hr

Abstract

The paper deals with characteristics and design of FEA software for ship structural analysis. Technological changes that occur in naval architecture and related disciplines that influence its relative performance are identified. The place of such tool within ship structural design is analysed and important requirements are formulated. Opportunities of use of open-source software are discussed. Importance of object-oriented programming for its code architecture is demonstrated. FEA software called OOFEM is selected to demonstrate the relevance of the aforementioned reasoning. At the end, classification societies' rules regarding finite element analysis of ship structures are analysed and used to assess OOFEM capabilities.

1. Introduction

Ship design is an iterative process where ship characteristics are defined and evaluated against requirements imposed by a ship-owner, state authorities and/or classification societies (CS). A step in this process is definition of ship structure, including its configuration (topology) as well as dimensions and materials for its constitutive elements (outer and inner shell, girders, stiffeners, etc.). Traditionally, this has been based on CS simple and easy-to-use formulae derived from their long-term experience with structural failures. Recently, the CS started to require finite element analysis (FEA) for most ship types as will be explained in chapter 5. This opens up a window of opportunities for application of other structural design tools beside the FEA such as sensitivity analysis and structural optimisation. They were proposed already in *Hughes et al. (1980)*. The proposed design method is based on first principles and use of computers to obtain an optimal structure, i.e. one that fulfils all the constraints and design objectives in the best possible way. The methodology consists of three modules: structural analysis module based on the finite element method (FEM), structural evaluation module to verify the obtained structural response against the imposed requirements (mainly from the CS rules) and structural optimisation module based on mathematical programming methods that seek to determine the optimal structure based on previously defined measure(s) of merit (design objectives) and design constraints. The proposed design method, called rational structural design, has since been applied to many real-life ship structures, *Zanic et al. (2000), Stone et al. (2017)* and has demonstrated that significant gains can be achieved with respect to the given design objectives (e.g. lower structural mass). An example of software that implements such an approach is MAESTRO, *Hughes and Paik (2010)*.

In this paper, we discuss the place of FEA in the rational ship structural design method and its interactions with other modules whence important conclusions regarding FEA software can be derived. After that, the benefits of having access to source code of such software and application of object-oriented programming (OOP) to its architecture are discussed and demonstrated on FEA software called OOFEM. At the end, we present CS requirements with respect to FEA and use them to assess capabilities of the aforementioned software.

2. Some characteristics of FEA software used in the rational ship structural design

2.1. Technological changes that require FEA software upgrades

The purpose of this subchapter is to outline some basic relationships of the linear static elasticity formulated by means of the FEM. This will serve us to identify some areas of research and technological changes which might require occasional upgrades to a FEA software if it is to remain up to date.

Many finite elements (FEs) in use today are based on an assumed displacement field *MacNeal (1994)* so expressions derived in sequel are based on that assumption (the same analogy holds for other FEs as well).

A new FE design usually starts with assuming distribution of displacements within the element:

$$u = [x]\{C\} \qquad (1)$$

where u represents general displacements (distribution of displacements within the element), $[x]$ is the basis function (e.g. for a rod element this can be $[1\ x]$) and $\{C\}$ is the vector of constant coefficients (for a rod element this could be $[C_1\ C_2]^T$). By defining a coordinate system for the element, we can obtain positions of its nodes and form the vector of nodal displacements $\{u\}$:

$$\{u\} = [X]\{C\} \qquad (2)$$

where $[X]$ is a matrix containing coordinate values of the nodes. By introducing (2) to (1) we get:

$$u = [x][X]^{-1}\{u_i\} = [N]\{u\} \qquad (3)$$

where $[N]$ is the matrix of shape functions. Equations (1) and (3) define the basic behaviour of the FE, which it is of ultimate importance for its accuracy.

In order to derive expressions for strains and stresses, we rely on the basic theory of elasticity and equations (1) and (3):

$$\{\epsilon\} = [L]\{u\} = [L][N]\{u_i\} = [B]\{u\} \qquad (4)$$

where $[L]$ is a linear derivative operator and $[B]$ is the strain-displacement matrix.

We can derive stresses from (4) and the elasticity:

$$\{\sigma\} = [D]\{\epsilon\} = [D][B]\{u_i\} = [S]\{u\} \qquad (5)$$

where $[D]$ is the elasticity matrix.

By employing principle of virtual work (or equivalent), we can obtain an expression for element's stiffness matrix:

$$[k^e] = \int_{V^e}[B]^T[D][B]dV \qquad (6)$$

where V^e is element's volume.

However, the element's stiffness matrix given in (6) is defined in element's own coordinate system (usually termed the local coordinate system) that generally is not coincident with the structure's coordinate system (usually termed the global coordinate system) so we need to "transform" $[k^e]$ to the "global coordinates":

$$[k] = [T]^T[k^e][T] \qquad (7)$$

where $[k]$ is the element stiffness matrix in the global coordinate system and $[T]$ is the transformation matrix. Now we can assemble structure's stiffness matrix [K] (also called the global stiffness matrix) by summing all element contributions:

$$[K] = \sum_{i=1}^{N_e} [k]_i \qquad (8)$$

where N_e is the total number of FEs in the structure. Having the global stiffness matrix *[K]* and structural loads $\{F\}$ (which are defined by a user and in the shipbuilding are mostly prescribed by the CSs), we can determine nodal displacements by solving the following system:

$$[K]\{U\} = \{F\} \qquad (9)$$

where $\{U\}$ is a vector of "global" displacements (displacements in the global coordinate system). Once we have $\{U\}$, it is possible to transform it back to the local nodal displacements $\{u\}$ and by means of equations (4) and (5) calculate strains and stresses. In ship structural design, we are predominantly interested in the FE stresses.

We can now analyse expressions from (1) to (9) and see what kind of changes might occur in the course of time and what are the sources of these changes because FEA software should be able to encompass them if it is to remain up to date.

Eqs. (1) to (3) define basis and shape functions. They are defined by an element designer and are primarily based on his reasoning about what kind of element's behaviours he wants to achieve or to avoid, *MacNeal (1994)*. In much of FEA software the shape function matrix *[N]* is explicitly defined in the code and FEA software usually starts its operations from this point. Alternatively, the strain-displacement matrix *[B]* can be specified and in that case matrix multiplication from equation (5) is not conducted by the software. The element designer can develop general FEA software or an industry specific one. There are many FEs developed specifically for ship structural analysis (e.g. beam FEs that can take into account the effective breadth as described in *Palaversa et al. (2016)* and available in MAESTRO). Thus, we might say that, as the time passes, new, more accurate and industry specific FEs will emerge and it will be necessary to implement them in the FEA software.

Once *[B]* and *[D]* are determined, we can calculate stiffness matrix for each FE by means of equation (6). This is mostly done using a numerical integration technique (e.g. the Gauss quadrature). The changes that might occur here are regarding number and arrangement of integration points (e.g. use of selective underintegration to alleviate shear locking), new quadrature techniques... These changes can come from the FEM research community or the mathematicians.

Operations defined by (6) and (7) are perfect candidates for parallelisation as they are independent of each other (calculation of stiffness matrix for one element does not depend on anything outside of that element and once its stiffness matrix is obtained it can be transformed to the global coordinates independently of other elements). This is an example of a change that has emerged from the computer science with latest hardware developments and software that wants to remain efficient in terms of computational time must be able to adopt these changes. Furthermore, expert knowledge that a researcher-naval architect might have about some characteristics of ship structures can be used here to enhance the existing code (e.g. some ship structures have long midship areas of roughly the same characteristics and the chances are that there are many elements with same properties that differ only in their position in the structure which can be used to skip some element stiffness matrices' calculations).

Eqs. (8) and (9) denote the most computationally and time-consuming operations. The first one is assembly of the global stiffness matrix. As the time for solving large systems of linear equations given by (9) has been reduced due to emergence of parallel-vector solvers, this operation together with the calculation of the element stiffness matrices may now represent a significant amount of the total CPU time, *Nguyen (2002)*. Most of the advancements in this field comes from the mathematicians in terms of matrix storage formats and matrix decomposition methods and computer scientists in terms of hardware and software that enables greater parallelisation and vectorisation of the code. Experts in ship structural design and analysis can help by identifying and exploiting regularities inherent to the ship structures to select a suitable matrix storage format and the most efficient decomposition method related to it that general FEA software might not employ.

Based on the foregoing, we can identify four sources of technological changes (Fig.1) that, in the long run, influence the performance and usability of FEA software for ship structural analysis: development of the FEM (mechanics), advances in ship structural analysis and design (naval architecture), emergence of new computational methods and algorithms (mathematics) and evolution of computer hardware and software (computer science). All this implies that selection and development of FEA software are a multidisciplinary task that requires occasional catching-up on the latest advances from fields that are usually non-native to a naval architect.

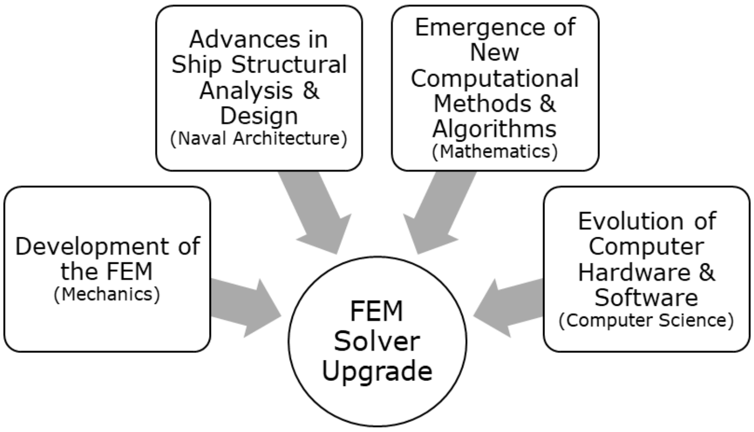


Fig.1: Four sources of technological changes that influence upgrades for a FEM solver in ship structural analysis and design

2.2. Structural analysis in the rational ship structural design

Although going into the details of the ship structural design process is not the topic of this paper, we will go through some basic steps and interactions between them to see whether some additional conclusions regarding a structural analysis tool can be drawn (for a detailed account on the rational ship structural design and terminology, refer to *Hughes and Paik (2010)*.

Ship structural analysis is usually conducted as a part of the ship structural design process to determine ship structural response which is necessary in proving that the proposed design fulfils all the imposed criteria (predominantly from the CSs). As already stated in the Introduction, this implies three distinct parts of the ship structural design process: analysis, evaluation and synthesis (that can be carried out by a human designer or by an automatic procedure that is able to find an optimal structure) as can be seen in Fig.2.

The FEM is usually employed to obtain ship structural response, i.e. to determine structural displacements, strains and stresses (as explained in the foregoing subchapter). In a parallel process, one must determine structural acceptance criteria (e.g. allowable stresses) to assess the design. The criteria mostly come from CS rules and can be implemented in suitable software, e.g. *Andric et al. (2019)*). After that, a human designer usually checks the obtained structural response characteristics against the required values and finishes the design if all the criteria are fulfilled (this is shown by dashed arrows in Fig.2). On the other hand, if an automatic synthesis procedure is to be employed (e.g. structural optimisation) then a more rigorous mathematical treatment of the problem is necessary or, in other words, a structural optimisation problem must be formulated. This includes definition of constraints' functions based on the determined acceptance criteria and the structural response as well as one or more objective functions. The functions of constraints bound the so-called feasible designs' space while the objective function(s) serve as a measure of "goodness" of the design. Based on the foregoing two sets of functions, it is possible to arrive to an "optimum" design. However, the "optimum" is, in general, found through a number of iterations (optimisation cycles) whose number can be very high and it is necessary to calculate structural response and criteria in each iteration as the structure's dimensions change (denoted by a dotted and a full line respectively in Fig.2). This implies that the number of FEAs can become prohibitively high due to the computational time and effort. An alternative approach is to approximate structural response characteristics in the vicinity of a design point, whose structural response is obtained by means of the FEM, and use the approximate values for determination of the constraints' functions, *Barthelemy and Haftka (1993)*.

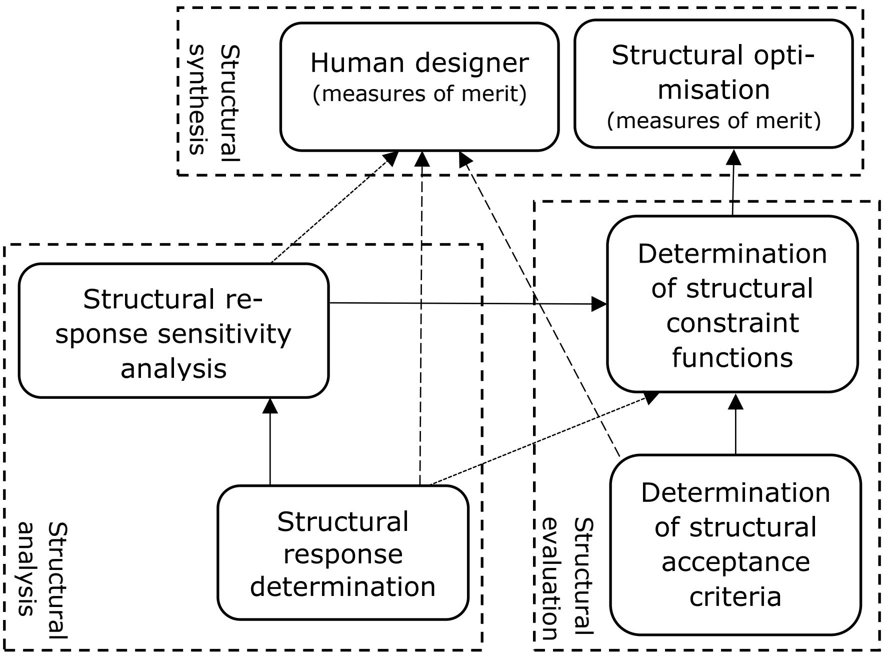


Fig.2: Scheme of a ship structural design process (solid lines denotes a typical data flow when structural optimization is used, dashed lines show a typical data flow with human designer and dotted lines show a non-typical data flow)

Consider Eq. (9) for a ship structure with all structural dimensions (design variables) defined and denote a design point that comprises the foregoing set of design variables by *P*. Based on them, FEA software can calculate structural response (e.g. vector of structural displacements $\{U^P\}$) as explained in the sequel, *Nguyen (2002)*.

We start with decomposition (e.g. Cholesky) of the stiffness matrix $[K^P]$:

$$[K^P] = [L^P][L^P]^T \qquad (10)$$

where $[L^P]$ is a lower triangular matrix and $[L^P]^T$ is an upper triangular matrix. By introducing (10) to (9), we can write:

$$[L^P][L^P]^T\{U^P\} = \{F^P\} \qquad (11)$$

It is possible to define a substitution vector $\{y^P\}$ such that:

$$[L^P]^T\{U^P\} = \{y^P\} \qquad (12)$$

By introducing (12) to (11), we obtain the following system:

$$[L^P]\{y^P\} = \{F^P\} \qquad (13)$$

where $\{y^P\}$ is the only unknown and (13) can be solved by the so-called forward substitution. Once we have $\{y^P\}$, it is possible to solve (12) for vector of nodal displacements $\{U^P\}$ by backward substitution. Once we have vector of structural displacements $\{U^P\}$, we can try to approximate its value in the vicinity of point P denoted as $P+\delta P$. E.g. we can use linear approximation based on the Taylor series expansion (with higher order terms neglected):

$$\{U^{P+\delta \mathrm{P}}\} = \{U^P\} + \left(\{x^{P+\delta \mathrm{P}}\} - \{x^P\}\right)\frac{d\{U^P\}}{d\{x^P\}} \qquad (14)$$

where $\{x^P\}$ and $\{x^{P+\delta P}\}$ are vectors of design variables and $\{U^{P+\delta P}\}$ is a vector of structural displacements. Thus, in order to calculate structural displacements at a distance δP from point P, we must obtain derivatives of structural displacements by design variables. Such an approach is called sensitivity analysis and it can be used not only in the structural optimisation (for functions' approximation) but also by a human designer to examine how changes in design variables influence response characteristics, design constraints or objectives (as depicted in Fig.2).

An intuitive approach to obtaining derivatives of structural displacements by design variables is to simply differentiate equation (9) at point *P, Adelman and Haftka (1992)*:

$$[K^P]\frac{d\{U^P\}}{d\{x^P\}} = \frac{\delta\{F^P\}}{\delta\{x^P\}} - \frac{\delta[K^P]}{\delta\{x^P\}}\cdot\{U^P\} \qquad (15)$$

and solve it for $d\{U^P\}/d\{x^P\}$. The right-hand side in the above equation is usually referred to as the pseudo-load vector because if we compare forms of Eqs. (9) and (15), we see that it stands in the position of structural loads. Once we obtain the pseudo-load vector and if we have access to the decomposed $[K^P]$ from (10), the solution of (15) is relatively cheap because it is necessary to carry out only operations given by (11) to (13) (approximately 90% of the total equation solving time is spent on the decomposition, *Nguyen (2002)*.

The cumbersome term, however, is $\delta\{K^P\}/\delta\{x^P\}$ as its evaluation requires differentiation of the global stiffness matrix by the full set of the design variables (number of design variables in a three cargo hold model of a ship structure is of order of few hundred). This is conducted by going back to the element stiffness matrices, differentiating them and then assembling their derivatives into the global matrix of stiffness derivatives:

$$\frac{\delta[K^P]}{\delta\{x^P\}} = \sum_{i=0}^{N_e}\left(\frac{\delta[k^P]}{\delta\{x^P\}}\right)_i \qquad (16)$$

There are two approaches to obtaining the latter term in the Eq. (16). The first is to analytically differentiate element stiffness matrices which implies having all the data from expression (6) available (and the Jacobian matrix if the element involved is isoparametric). This approach is denoted the analytical direct method in this paper. The second approach is to use the finite differences method and it is often termed as the semi-analytical method, *Adelman and Haftka (1992)*. The former approach is generally less prone to large errors while the latter is easier to implement. However, both imply that the sensitivity analysis solver has access to relevant intermediate results provided by the FEM solver. The authors are not aware of any application of the analytical direct method of sensitivity analysis to structural design of realistic ship structures. Alternatively, it is possible to determine $d\{U^P\}/d\{x^P\}$ by means of the adjoint method that should be more efficient in cases where the number of constraints is less than the number of design variables, *Adelman and Haftka (1992)*. When in its full fashion, the opposite is true for the ship structural optimisation problem. However, as the number of active/relevant constraints varies from one optimisation cycle to another, *Arora (1993)*, one might speculate about an interchanging use of the both methods within the solution to the same optimisation problem (if a mathematical programming method is used this should not be the problem as the problem is approximated anew in each solution sequence anyway).

As already mentioned, the foregoing approaches are not employed very often in ship structural design. More often, the overall finite difference method is employed. In that case the FEM solver can be used as a black box. However, computational costs involved with the use of the finite difference method for $d\{U^P\}/d\{x^P\}$ are much greater because all the operations given in (6) to (13) must be repeated one more time. Moreover, there is a possibility that the approximation cannot be achieved with the required level of accuracy at all because of the well-known "step-size" dilemma (see more in *Keane and Nair (2005)*). An alternative to the overall finite difference method when the FEM solver source code or relevant intermediate results are not available is the complex step method that eliminates the aforementioned "dilemma", *Keane and Nair (2005)*.

3. Open-source and object-oriented FEA software in the rational ship structural design

3.1. Open-source software in ship structural analysis

Open-source software has extensively been used for many purposes, from Internet browsers (e.g. Mozilla Firefox) and office suites (e.g. OpenOffice or LibreOffice) to engineering and scientific calculations and graphic tools (Salome-Meca, ParaView, OpenFOAM, etc.). At the beginning we shall mention some of its advantages and disadvantages usually found in the literature.

Open-source software tends to have lower total cost of ownership (TCO) than a similar proprietary software, *Crahmaliuc (2019)*. However, open-source software is rarely totally free, especially if it is a complex system that the users are not familiar with. Apart of the license fees that the users do not pay, the following costs must be taken into account: training costs (e.g. the full price of an OOFEM training course is 800 € (http://www.oofem.org/en/courses) while that of OpenFOAM is 1200 € (https://www.openfoam.com/training/schedule.php)), support (some organisations provide this for open-source software at a fee or, if not available or used, users themselves will likely spend considerable amount of time trying to solve the problem on their own) and documentation (some open-source tools are notorious for bad documentation).

Code reuse is another great benefit of open-source software. Greater collaboration and exchange enhance implementation of new features which means it can be adapted to user's specific needs, *Crahmaliuc (2019)*. This, however, can also come at a cost. Depending on current capabilities of open-source software one plans to use, the costs of adding features that one needs can become prohibitively high. Thus, it is important to thoroughly assess current capabilities of open-source software before the final decision is made.

Open-source software is continually evolving and less prone to bugs, *Crahmaliuc (2019)*. *Tuma (2005)* states that the "studies have demonstrated that code reviews", i.e. analysing source-code for problems, "can remove defects much more effectively then testing" and thus claims that code transparency leads to better quality assurance. On the other hand, open nature of the source-code raises concerns about its security. However, *Tuma (2005)* states that despite "experts on both sides of the open-source security debate contribute many compelling arguments … the bottom line is that open source software is not automatically more or less secure than proprietary software".

Another advantage usually emphasized by open-source software community is a free-of-charge support available through forums, mailing lists, discussion groups etc. "You get access to the support offered by technical community" while when using proprietary software "vendor support is conditional to maintenance subscription", *Crahmaliuc (2019)*. On the other hand, "when you run into problems, you only have the community of your fellow users to rely on for support which they provide on a voluntary basis", *Crahmaliuc (2019)*. We will see in the sequel that lack of inactivity of such communities can have a significant impact of one's progress and experience with the software.

An often-cited disadvantage of open-source software is that they usually lack a user-friendly GUI (graphical user interface). Thus, some of the platforms that have recently emerged, such as SimScale, use open-source solvers but provide their own UIs.

If we move our focus more towards engineering software used for engineering mechanics' calculations, a successful example is certainly OpenFOAM, which is widely used in computational fluid dynamics (CFD) although it is not limited to it. Its development started in academia in the 1990's and by 2004 it was officially released as open-source, *Chen et al. (2014)*. Before OpenFOAM, computer programs used in CFD had primarily been developed at universities and national laboratories and their lifetime had been that of 10 to 20 years, *Chen et al. (2014)*. Their development had been fuelled by numerous PhD students and postdoctoral associates which had led to continuity problems (as the students come and go) that had manifested themselves in programs' lifetime and code quality (the notorious "spaghetti code"), *Chen et al. (2014)*. Building larger communities of developers had been very difficult at the time due to the non-existent Internet and because the institutions where they had been developed had hesitated to share code due to large amount of resources that had been spent on the programs, *Chen et al. (2014)*. OpenFOAM managed to solve all of these problems: by going open-source and free and by employing object-oriented programming and C++. The former enabled creation of a large users' and developers' community that managed to exploit all the aforementioned benefits of the open-source (add new features through code reuse thus lengthening software's lifetime, fix bugs thanks to greater code transparency, etc.). The latter made the code easier to extend and maintain, more readable and more accessible by employing the so-called equation mimicking syntactical model, *Chen et al. (2014)*. In case of OOFEM that we deal with in the next chapter this can be seen in Fig.3. Formula for computing element's volume at a Gauss point in the last line of code is easily understandable to anyone, even if one cannot program.

```
double
Structural2DElement :: computeVolumeAround(GaussPoint *gp)
{
    // Computes the volume element dV associated with the given gp.

    double weight = gp->giveWeight();
    const FloatArray &lCoords = gp->giveNaturalCoordinates(); // local/natural coords of the gp (parent domain)
    double detJ = fabs( this->giveInterpolation()->giveTransformationJacobian( lCoords, * this->giveCellGeometryWrapper() ) );
    double thickness = this->giveCrossSection()->give(CS_Thickness, gp); // the cross section keeps track of the thickness

    return detJ * thickness * weight; // dV
}
```

Fig.3: Method for computing element's volume at a Gauss point

We can now reflect on the characteristics of open-source software just mentioned and on its relevance for the research community and industry. Based on this, we can try to formulate a set of assessment criteria for open-source FEA software used in ship structural design.

If we are to work with the code in any manner, it has to be readable. This will also help us later to trace bugs, rectify errors, etc. If we plan to add new features, the code must be easily extendible. This enhances the longevity of the software as the chances are greater that an easily extendible code will incorporate all the latest developments and thus survive over a long period of time than if one needs to spend hundreds of hours just to figure out its structure. Next, the code must be maintainable as its reliability will be seriously jeopardised if the bugs cannot be easily traced and rectified. All of these characteristics are supported by OOP. Thus the first criterion is that its architecture must be object-oriented.

The second criterion is related to its community of users. As we have seen, when open-source software is used, usually the only place to rely on for support is an Internet forum, a discussion group, a mailing list, etc. The kind of support one might need usually falls into one of the two categories: user or developer. In each of the categories, it is beneficial that other community members work in the same field and deal with the same types of problems. Moreover, this kind of information can give us a glimpse into possibilities of the software. E.g. if there are a lot of people concerned with, say, linear static analyses of ship structures but only few with ultimate strength and collisions, the chances are that there will not be many very refined options for different types of non-linear analyses, if there will be any at all. Or if there are many mathematicians and not so many engineers, than one might find very efficient solution algorithms and a poor library of FEs that could be used in engineering analyses. Thus, we can say that, regarding community, the following is important: its activity (we give an example of metrics that can be used to assess activity in Table I), size and type (scientists, engineers, what branches of engineering are represented, etc.). All of these criteria are relevant for the software development as well because if a project has been inactive for a long period of type, maybe it has died out. Also, if the number of developers that actively work on it is low or shrinking than this opens up questions regarding its future development. Generally, data about the community of users can be found by analysing their Internet forum, discussion groups (if they are not private), etc. If the project is hosted on GitHub, Bitbucket or similar, than data about software development (number and time of commits, number of forks, etc.) can be found using their metric or by devising one's own based on the data that is usually publicly available (we use some of these in Table I).

Table I: Indicators of project's activity for selected open-source software projects (data of 9.6.2020)

Software/ Project[1]	Forum activity[2]	No. of registered users[3]	Time to first answer (hrs)[4]	No. of threads with no answer (%)[5]	No. of commits[6]	No. of active forks[7]
code_aster	34	14	28.4	45.83	17	5
FreeFEM	9	3	2.9	25.64	92	37
OOFEM	2	1	6.5	40.00	9	21
OpenFOAM	251	N/A	8.3	20.00	N/A	N/A

The following criterion is a type of license. If one plans to use software for any other purposes than one's own use (e.g. resell it as a part of a greater package, redistribute it, etc.) then one should take a detailed look on what is included in one's license and what is not.

The next criterion is the target platform. Most of the engineering open-source software is primarily executed and built on Linux. If its use is planned on Windows or Mac, additional costs in terms of programming time can occur as the chances are that compiler used on Windows/Mac does not work in the same way as the one on Linux so some errors will have to be fixed before building it. However, there are projects where Windows versions are available as well which will also depend on the size of community of Windows users.

And the last criterion that can become significant in some cases is the programming language. The most widely used languages for engineering and scientific software are Fortran and C++ while Python has gained some popularity in recent years (although one might argue whether Python will ever be on

[1] Data in columns 2 to 5 are collected from project's official Internet forum except in case of OpenFOAM (that has multiple distributions) where the forum on https://www.cfd-online.com/ was used. Data in the last 2 columns are collected from project's GitHub repository (Bitbucket in case of code_aster).
[2] Number of posts in period not greater than 7 days from the date of data collection.
[3] Number of users registered in period not greater than 7 days prior to the date of data collection. For code_aster it was not possible to disambiguate newly registered users from those of Salome-Meca as they share the same forum. However, at the time of collecting the data, ratio of posts on forums associated with code_aster vs. those of Salome-Meca was almost 1:10. Additionally, Salome-Meca is very often used together with code_aster thus we might reasonably presume that among the newly registered users, the number of those who registered only because of Salome-Meca, is not significant.
[4] Time between the first post in a topic/thread and the first answer. The number was calculated as an average value of the aforementioned time in hrs for the top 5 topics/threads w.r.t number of views in forum with most posts (if there are multiple forums available).
[5] Number of threads/topics with no answer provided in period between 1 week and 1 month form the date of data collection divided by the total number of threads/topics in the same period multiplied by 100 to obtain percentage.
[6] Number of commits period not greater than 30 days prior to the date of data collection.
[7] Number of forks that had any activity in max 1 year prior to the date of data collection.

equal grounds with Fortran and C++ for computationally demanding tasks). In the past, the decision had to be made between Fortran with procedural programming and computational speed on hand and C++, object-oriented programming and compromises w.r.t computational speed on the other hand. However, Fortran has supported OOP since Fortran 2003 standard and some contemporary studies have shown very little difference in program execution time when C++ with an object-oriented code is executed in comparison with Fortran (see more in subsequent subchapter). Moreover, it is always possible to write computationally very demanding operations procedurally in Fortran and wrap it in a module used within a C++ object-oriented code. Thus, the selection of language has primarily influence time that needs to be spent if one is not familiar with the programming language the code is written in.

3.2. OOP in the design of complex FEA software

Before discussing benefits of OOP, it is useful to go through reasons that pushed early researchers towards this new programming paradigm that caught their attention in the 1980s. In, what many consider the first paper on OOP implementation of the FEM, *Forde et al. (1990)* list disadvantages of procedural programs: “notoriously complicated and error prone”, “complex control and data structures … tend to obscure fundamental intentions, making development and maintenance costly”. They continue by saying that “finite element analysis programs should be constantly changing to satisfy current and future demands of the engineering profession”. Thus important requirements imposed on FEA programs are extendibility and maintainability. This is facilitated if much of the code can be reused so they say that “one of the key issues for software developers is…how to structure their new software to maximise reusability in future renovations”. At that time, some concerns existed about computational speed and efficiency of object-oriented code vs. procedural code. *Dubois-Pelerin et al. (1992)* noticed that their object-oriented program written in programming language Smalltalk was considerably slower than a comparable program written in the procedural manner. *Mackie (2009)* states that “potential low execution speed of object-oriented code together with a lot of functional legacy code has led to a relatively slow and weak adoption of object-oriented practices in the FEM implementation”. Later he conducts a number of studies to test computational speed and acknowledges that most issues related to object-oriented programming have been resolved: new programming languages with new compilers and mature programming techniques allow FEM solvers written in the object-oriented manner to be equally fast as those written in the procedural manner, *Mackie (2009)* (e.g. an object-oriented program written in C++ is equally efficient as a procedural program and even an object-oriented program written in C# executes only 10% slower than the corresponding program in C++ or procedural code, *Mackie (2009)*).

The basic idea of OOP is to make software code more similar to the way people think or, as *Rouson et al. (2011)* write: “object-oriented thinking lends itself to an anthropomorphic view of objects”. Objects have their state (attributes) where the data are stored, behaviours (methods that define what an object does) and they communicate by means of messages.

Objects are defined by means of encapsulation, i.e. that all attributes (data) needed by object’s methods (operations) are stored within the object and generally only they can access object’s attribute. This is called data hiding and is a major part of encapsulation, *Weisfeld (2013)*. By employing this principle the code becomes more readable. It also enables equation mimicking mentioned in the foregoing subchapter. Further on, all methods and attributes of an object that can be accessed from the outside form object’s interface and are said to be public. The rest belong to object’s implementation and is set to be private. If an object (sender) needs some information from another object (receiver), it is supposed to send it a message by invoking one of its public methods. The receiver’s method then accesses data that contain the information asked and sends it back to the sender. Thus the whole communication is via objects’ interfaces. This is a very powerful concept because if we want to change object’s implementation, by retaining its interface the same nothing has to be changed in objects that communicate with the changed object. Consider, for example, an object that prints stresses into a file and another object that represents a FE. The FE object contains a method that calculates stresses and returns the stress vector. In order to print stresses into a file the print object

must invoke the method for stress calculation of the FE object and print the received stress vector into a file. However, the chances are that, in the course of time, new FEs will be added and their stresses might be calculated in a different manner. If the stress calculation method of the new FE object has the same name, arguments and parameters (i.e. if its interface remains the same), nothing will have to be changed in the print object. This improves software maintainability and extendibility considerably.
Another important feature of OOP is inheritance. It allows one object to use as its own attributes and methods of other object(s). Following the example given above, a number of FEs might calculate stresses in the same way. By using inheritance, one and the same method can be inherited by all objects (called child objects/classes) that have this property so that it does not have to be written many times again.

A feature related to inheritance is polymorphism that enables one object to change implementation (behaviour) of a method inherited from another object. Going back to the aforementioned example, a new FE might calculate stresses in a different way while still needing same information to do it and returning result in the format. Thus we can retain method's interface while changing its implementation in each child class/object. Both of the inheritance and polymorphism facilitate code reuse.

Better readability and code reuse enhance extendibility and maintainability. All of these are important for any complex software project to flourish as has been demonstrated in the foregoing subchapter. In the subsequent chapter, this will all be demonstrated on a real-life example.

4. OOFEM – object-oriented, open-source FEA software

OOFEM stands for object-oriented finite element method. It has been continuously developed since 1997 and today encompasses more than 230 000 lines of C++ code, *Patzak et al. (2019)*. It is open-source software published under GNU LGPL license. It will serve us here to demonstrate application of OOP concepts discussed in the foregoing chapter to realistic FEA software.

FEMComponent class is a base class for all building blocks of the finite element problem: elements, materials, boundary conditions, nodes, etc. In order to determine attributes and methods that this class should possess, one has to think about generalities common to all "FEM components" (namely, what is common to a finite element, a node, a boundary condition, etc.). If we take a closer look at Fig.4, we will see that FEMComponent's methods provide only for the most basic services: initializeFrom is used for component's initialization, checkConsistency is used for checking important properties of a component, printOutputAt and printYourself are used to print information to a file (or on screen), etc. Also, we can see that most methods are written in italics. The italic typeface is used, according to UML 2.0 standard, to denote an abstract (or virtual) method which means that no implementation is provided. Thus these methods are used only to provide a common interface and their implementation will be provided by child classes.

Moving down in the hierarchy tree we come to the Element class which represents a finite element. As can be seen, this class is a base class for all finite elements (bear in mind that FEM can be used for fluid and thermodynamic problems and this is why some of its child classes are FMElement and TransportElement). Again, we should think about those attributes and behaviours that are common to all elements. Thus we can find methods such as giveCharactersiticMatrix, giveCharacteristicVector, computeVolume, computeArea, giveIntegrationRule (e.g. in structural analysis this will usually be Gauss quadrature), etc. We can see that most of the methods are still abstract/virtual and their implementation will be provided in child classes. However, some of the methods are implemented already on a level this high. One of these methods is giveDofManager. Reason why this functionality can be implemented this high in the hierarchy is that all finite elements have points (called nodes) or sides where they are connected to adjacent elements. If we implemented this method in a class that is lower in the hierarchy, we would have to duplicate code as the set of identical procedures would be repeated in all subclasses (e.g. in FMElement, TransportElement, StructuralElement, etc.). This would cancel out one of the main benefits of OOP, namely code reuse. It also demonstrates probably the greatest challenge that a designer of object-oriented code faces, namely selection of a right hierarchical level for every method because if the method is implemented too late (too low in the hierarchy), it leads to duplicate code and if it is implemented to early than it might be inherited by classes that do not need it thus, unnecessarily, occupying space in computer's memory.

Going further down in the hierarchy tree, we come to StructuralElement class which represents a base class for all finite elements used in structural analysis (linear, non-linear, static, dynamic…). We can see that it provides interfaces (abstract methods) for typical operations encountered in solid mechanics when the FEM is used: computeMassMatrix, computeStiffnessMatrix, computeLoadVector, computeConstitutiveMatrixAt, etc. These methods, as well as those encountered in child classes of StructuralElement, can be compared to operations outlined in expressions (4) to (9) of subchapter 2.1. We can now consider computeStiffnessMatrix method to see why it is implemented at this level and why the way it is. computeStiffnessMatrix is denoted as abstract method and it appears in the hierarchy tree for the first time. Reason why it is not part of an earlier class is because not all finite elements possess a stiffness matrix (e.g. FEs used in fluid or transport problems) so if it were implemented, even the classes that do not need it, would inherit it. Reason why it is not implemented later is that in that case it would have to be rewritten in all child classes that possess a stiffness matrix as e.g. NLStructuralElement, Truss1d, beam elements, etc. Reason why it is declared as abstract lays in fact that FEs, in general, have different stiffness matrices that are evaluated in different way. Thus, its implementation should be provided by its child classes and its role here is to provide only a general interface.

The next level in the hierarchy tree is NLStructuralElement class and classes similar to it. The NLStructuralElement class represents a base class mostly for 2D and 3D finite elements, typically plates and shells of different kinds. We can see that computeStiffnessMatrix method is implemented at this level. The reason is that all these elements have the same procedure for obtaining their stiffness matrix which is calculated as an integral over element's volume of a product of element's strain-displacement matrix transpose ($[B]^T$), its constitutive matrix $[D]$ and the non-transposed value of $[B]$, all evaluated at element's Gauss points (if Gauss quadrature is used) as given by Eq. (6).

The hierarchy of Fig.4 ends with concrete finite elements such as MITC4Shell (a shell element), Quad1Mindlin (a Mindlin plate element), etc. The lowest classes in the hierarchy typically implement behaviours related only to them, e.g. giveMaterialMode method of PlaneStressElement class returns constitutive properties that are based on plane stress hypotheses thus they differ from shell, plate or plain strain constitutive relations.

We shall explain now why abstract methods are used so extensively in this code. It is related to the interface concept of OOP (see more in subchapter 3.2.) where interfaces are used for communication between class methods. Let's consider example of two methods, one belonging to class Domain and other to class PlaneStressElement. Class Domain keeps record of all FEM components (all elements, nodes, materials, etc. together with their properties). If the Domain class method wants to know stiffness matrix of a structural element, it will, put very simply, invoke its computeStiffnessMatrix method that will (somehow) calculate stiffness matrix and return it to the Domain method. This way the Domain method gets what it wants never knowing anything about the way stiffness matrix is actually calculated – all what it must know is the interface of the computeStiffnessMatrix, i.e. its name and the list of parameters (as declared in StructuralElement class which is a parent class of PlaneStressElement). If the interface is not changed, the code of the Domain method remains the same. However, this does not mean that the way stiffness matrix is calculated must remain the same. Another time the Domain method may want to obtain stiffness matrix of, say, MITC4Shell element. It will invoke its computeStiffnessMatrix (that has the same interface because MITC4Shell is also a child class of StructuralElement) that will have different implementation (different way of calculating stiffness matrix) but its interface will remain the same as will the code of the Domain class. This powerful concept allows for adding new finite elements while code in classes that communicate with the changed class remain the same which enhances code extendibility and maintainability as explained in subchapter 3.2.

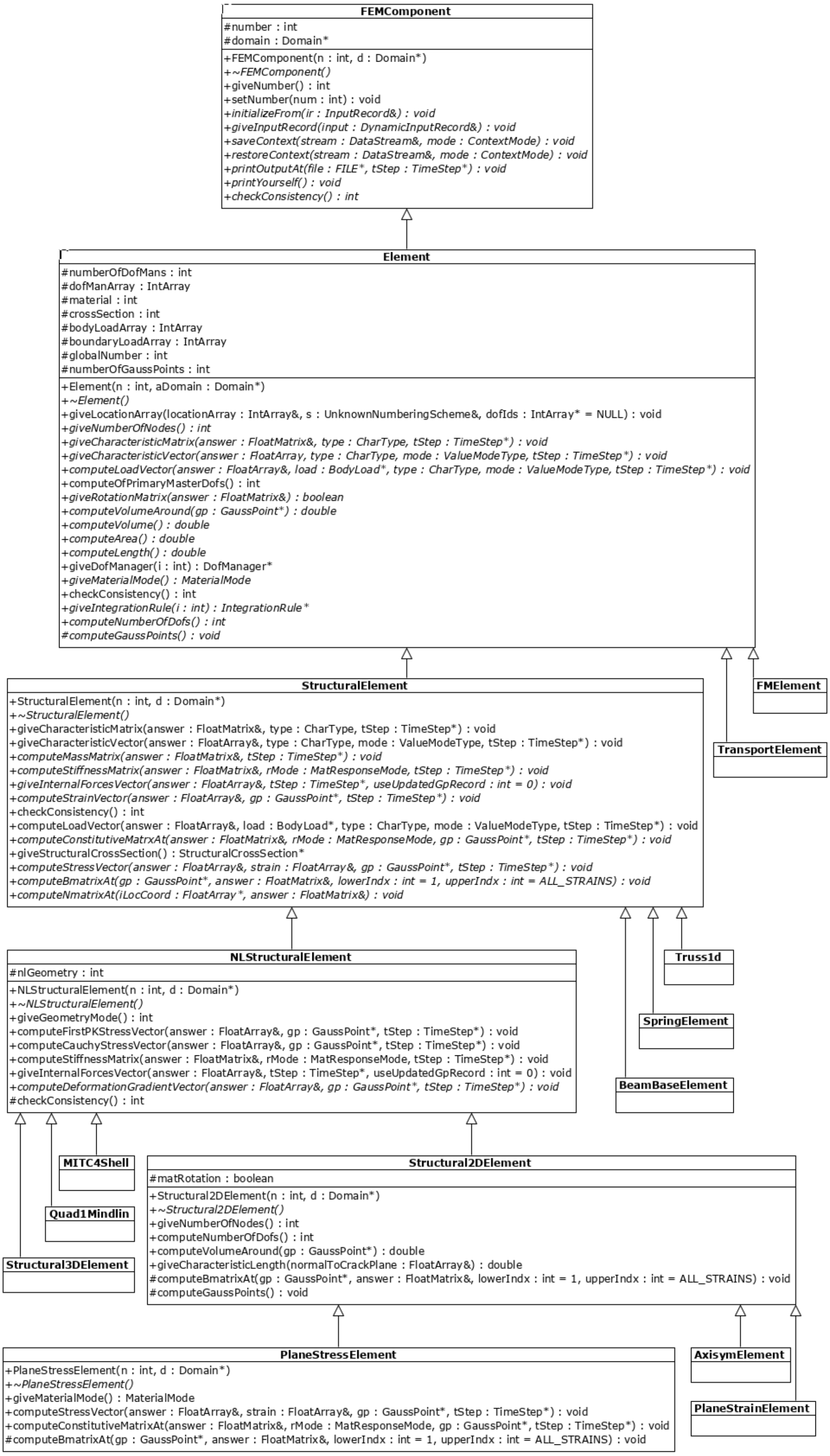


Fig.4: UML diagram of FEMComponent class partial inheritance tree

5. Classification societies' requirements on FEA and assessment of OOFEM capabilities

The FEM has been used in ship structural design for decades. However, it has been a recent trend that the CSs require strength assessment of ship structures by means of this method. IACS Common Structural Rules (IACS-CSR) *IACS-CSR (2018)* require FEA of cargo hold area for all bulk carriers and tankers whose length is 150 m or more. *DNVGL (2015)* requires FEA for many ship types, e.g. container vessels, ro-ro and car carriers, large cruise vessels (L > 150 m), etc. Traditionally, many CSs have required FEA in case of novel designs of ship structures or structural elements, for fatigue and buckling assessment, etc. Thus, they have established rules and acceptance criteria for results of the FEAs. In sequel, we describe the rules relevant for FEA software as given in *DNVGL (2015)*.

From a CS's point of view "any recognised software can be used provided that all specifications on mesh size, element type, boundary conditions, etc. can be achieved with this computer program" *DNVGL (2015)*. Regarding FEs, the following should be available to an analyst: rod (truss) elements, beam elements, membrane elements and shell elements *DNVGL (2015)*.

Rod (truss) element is defined as a line (1-D) element with axial stiffness only and constant sectional area along its length *DNVGL (2015)*. Beam element is a line (1-D) element with axial, torsional and bi-directional shear and bending stiffness and constant properties along its length *DNVGL (2015)*. Membrane (or plane-stress) element is a surface (2-D) element with bi-axial and in-plane stiffness and constant thickness *DNVGL (2015)*. Shell element is a surface (2-D) element with in-plane (membrane) and out-of-plane (bending) stiffness and constant thickness. Its inner angles should be between 45° and 135° and its aspect ratio should be kept close to 1 but should never exceed 3 for 4-node and 5 for 8-node elements. Its shape functions should include "incompatible modes" which offer improved bending behaviour. Use of triangular shell elements should be kept to a minimum, especially in areas of high stresses. The assessment against stress acceptance criteria is normally based on membrane (in-plane) stresses *DNVGL (2015)*.

Regarding loads, point (concentrated), pressure (distributed) and body (e.g. gravitational and inertial forces) loads should be supported. It should be possible to define three forces (one force in direction of each axis) and three moments (one moment about each axis) in every grid point *DNVGL (2015)*.

Regarding boundary conditions, fixed (clamped) and pinned ends as well as combinations of constrained translations and rotations should be supported. Special boundary conditions should be defined by means of rigid links and end-beams *DNVGL (2015)*.

Regarding analysis' reports, element types, reactions in boundary conditions, shear, in-plane, axial and von Mises stress should be available *DNVGL (2015)*. Additionally, plate and stiffened panel buckling analysis and results as well as results showing compliance with design criteria should be provided *DNVGL (2015)*.

As emphasized in chapter 3.1., when one starts to work with new FEA software, it is important to assess its capabilities as the difference between its current capabilities and the requirements directly defines the amount of work that has to be done. In ship structural design, the requirements primarily come from CSs rules because at the end of a design process these are to be fulfilled. Additionally, they might arise from one's research interests or integration with other software. As we are concerned with the rational ship structural design, we would like to use OOFEM as input for our sensitivity analysis tool. In this paper, we will only briefly touch on the assessment based on the rules just described. A detailed assessment of OOFEM capabilities w.r.t to the aforementioned CS requirements is given in *Palaversa et al. (2019)*.

Regarding representations of boundary conditions and load cases, we have not found any differences w.r.t CS requirements. When it comes to definition of loads, OOFEM can model both concentrated and distributed loads. However, distributes loads are always evaluated by means of numerical integration of the load vector (see more in *MacNeal (1994)*). In many ship structural project that we have participated in and that are aware of, the distributed loads are usually distributed equally at all nodes (“intuitively” as it is termed in *MacNeal (1994)*). This is probably not a major difference because CSs usually define distributed loads as constant but it might be confusing for an end-user. Thus, we have opted for such a definition of distributed load.

The biggest difference between CS requirements and OOFEM capabilities have been found in FEs it encompasses. When we started to work with OOFEM, it comprised the following 4-node quadrilateral membrane, plate and shell elements (terminology is according to *Patzak (2018)*) of interest in the analysis of thin-walled structures: PlaneStress2d, QdktPlate and MITC4Shell. Other 4-node quadrilaterals available in OOFEM are disregarded because of the following reasons: they are based on assumptions which differ substantially from characteristics of a ship structural problem (e.g. axisymmetric and plain strain elements such as L4Axisymm or Quad1PlaneStrain element), they have known issues when used in the analysis of thin-walled structures (e.g. Quad1Mindlin element exhibits a strong locking behaviour when used for thin plates, *Patzak (2018)*) or they are intended for substantially different types of analysis (e.g. Quad1PlateSubsoil element used in civil engineering). We have conducted a series of simple tests with these elements and their properties are unsatisfactory. Thus, our aim is to add a new shell element whose properties should not be inferior to shell elements widely used in the shipbuilding industry and available in standard commercial FEA software. Moreover, a beam FE based on Euler-Bernoulli theory will be implemented as these are normally used in ship structural analysis.

6. Conclusion

This paper has demonstrated that significant technological changes occur and influence relative performance of FEA software. The changes come from advances in several disciplines beside naval architecture such as mathematics, computer science and mechanics. FEA software that wants to remain up-to-date must be able to incorporate these changes.

Moreover, a researcher or engineer can use FEA software as a part of an overall design procedure. Sensitivity analysis and structural optimisation can help in achieving better projects. When used in conjunction with the two, FEA software must be able to share some intermediate results with them. This is not always possible when commercial software is used. Thus, use of open-source FEA software is proposed. Because it is difficult to find FEA software that will fulfil all requirements by a naval architect or a researcher in this field, the chances are that he/she will want to add some new features. This, together with the aforementioned need for constant maintenance, calls for implementation of object-oriented programming paradigm in design of its code architecture. An example of such software is given and briefly discussed. At the end, classification societies’ rules regarding finite element analysis are presented and used for assessment of FEA software called OOFEM.

References

ADELMAN, H.M.; HAFTKA, R.T. (1992), *Sensitivity Analysis of Discrete Systems*, Structural Optimization: Status and Promise, pp.291-315, AIAA

ANDRIC, J.; PREBEG, P.; ANDRISIC, J.; ZANIC, V. (2019), *Structural optimisation of a bulk carrier according to IACS CSR-BC*, Ships and Offshore Structures

ARORA, J.S. (1993), *Sequential Linearization and Quadratic Programming Techniques*, Structural Optimisation: Status and Promise, AIAA

BARTHELEMY, J.-F.M.; HAFTKA, R.T. (1993), *Function Approximations*, Structural Optimization: Status and Promise, AIAA

CSR (2018), *Common Structural Rules for Bulk Carriers and Oil Tankers*, IACS

CRAHMALIUC, R. (2019), Open-Source of Proprietary Software – What is Best for Users?, https://www.simscale.com/blog/2017/06/open-source-vs-proprietary-software/

DNVGL (2015), *DNVGL-CG-0127 Finite element analysis*, DNV GL, Hovik

DUBOIS-PELERIN, Y.; ZIMMERMANN, T.; BOMME, P. (1992), Object-oriented finite element programming: II. A prototype program in Smalltalk, Computer Methods in Applied Mechanics and Engineering 98, pp.361-397

FORDE, B.W.; FOSCHI, R.O.; STIEMER, S.F. (1990), *Object-Oriented Finite Element Analysis*, Computers & Structures 34, pp.355-374

HUGHES, O.F.; PAIK, J.K. (2010), *Ship Structural Analysis and Design*, SNAME

HUGHES, O.F.; MISTREE, F.; ZANIC, V. (1980), *A Practical Method for The Rational Design of Ship Structure*, J. Ship Research 24, pp.101-113

KEANE, A.J.; NAIR, P.B. (2005), *Computational Approaches for Aerospace Design*, John Wiley & Sons

MACKIE, R.I. (2009), *Advantages of object-oriented finite-element analysis*, Proc. Institution of Civil Engineers - Engineering and Computational Mechanics 162, pp.23-29

MacNEAL, R.H. (1994), *Finite Elements: Their Design and Performance*, Marcel Dekker

NGUYEN, D.T. (2002), *Parallel-Vector Equation Solvers for Finite Element Engineering Applications*, Springer

PALAVERSA, M.; PREBEG, P.; ANDRIĆ, J. (2019), *REMAKE Project Deliverable 2.1 - Definition of Requirements for OOFEM Upgrade*, FSB, Zagreb

PALAVERSA, M.; ŽANIĆ, V.; PREBEG, P. (2016), *Comparison of Beam Finite Elements Based on Euler-Bernoulli Beam Theory*, SORTA 2016

PATZAK, B. (2018), *OOFEM Element Library Manual.* CTU Prague

PATZAK, B.; SMILAUER, V.; HORAK, M.; HAVLASEK, P. (2019), *OOFEM training course*, CVUT, Prague

STONE, K.; McNATT, T. (2017), *Ship hull structural scantling optimization*, Progress in the Analysis and Design of Marine Structures

TUMA, D. (2005), *Open Source Software: Opportunities and Challenges*, Crosstalk – The Journal of Defense Software Engineering, pp. 6-10

WEISFELD, M. (2013). *The Object-Oriented Thought Process*, Addison-Wesley

ZANIC, V.; JANCIJEV, T.; ANDRIC, J. (2000), *Mathematical Models for Analysis and Optimization in Concept and Preliminary Ship Structural Design*, IMDC

Open Simulation Platform –
An Open-Source Project for Maritime System Co-Simulation

Øyvind Smogeli, DNV GL, Trondheim/Norway, oyvind@smogeli.net
Kristine Bruun Ludvigsen, DNV GL, Trondheim/Norway, kristine.bruun.ludvigsen@dnvgl.com
Levi Jamt, DNV GL, Trondheim/Norway, levi.jamt@dnvgl.com
Bjørnar Vik, Kongsberg Maritime, Ålesund/Norway, bjornar.vik@km.kongsberg.com
Håvard Nordahl, SINTEF Ocean, Trondheim/Norway, haavard.nordahl@sintef.no
Lars Tandle Kyllingstad, SINTEF Ocean, Trondheim/Norway, lars.kyllingstad@sintef.no
Kevin Koosup Yum, SINTEF Ocean, Trondheim/Norway, KevinKoosup.Yum@sintef.no
Houxiang Zhang, NTNU, Ålesund/Norway, hozh@ntnu.no

Abstract

The Open Simulation Platform (OSP) is an open-source industry initiative for co-simulation of maritime equipment, systems and entire ships. With steadily increasing complexity due to the increasing use of software in ship systems, and the integration of equipment and systems from many providers, it has become increasingly difficult to design, build, operate and assure ships and other maritime assets. The OSP sets out to change this by providing key tools and working processes for technical systems engineering. This paper presents the OSP background, its key deliverables, and three use cases that demonstrate the benefits of collaborative digital twin simulations.

1. Background and rationale

"How can the maritime industry prepare for the emerging risk to life, property and the environment coming from increasingly complex, software-controlled ships?" This question is taken from the DNV GL Technology Outlook 2030, *DNVGL (2020)*, chapter called "Impact on Maritime", and sets the scene for the complexity challenges faced by the maritime industry. What are then, these challenges? Referring to Fig.1, our accumulated experience and customer feedback shows that:

- It is increasingly difficult to design and optimize the integrated ship systems, since the overall system performance depends on the interplay between systems and software from different sources, and their actual behaviour cannot be predicted without actually simulating or observing them in operation together.
- It is difficult to assess whether the system is safe and fit-for-purpose. Even after it is put into operation, only a small number of the possible scenarios the system should be able to handle can feasibly be tested.
- With systems and software from many providers, commissioning and integrating them as one system onboard the ship is challenging, especially since there is no system integrator having the full overview and insight into the various parts of the system.
- In addition to complex systems, we are also dealing with complex operations and complex human-machine interaction, further increasing the challenge.
- Change management – both during construction and in operation – is increasingly difficult, since it is not possible to predict the consequences of a change in one software system or the other, connected systems and the ship systems as a whole. There are numerous stories about apparently simple system updates/patches gone wrong, since integration suddenly was broken unintended.

Why is this so? One apparent reason is that, while complexity in traditional mechanical systems is naturally limited by physical constraints and the laws of nature, complexity in integrated, software-driven systems easily exceeds human comprehension. How then, to assess these complex systems? One key takeaway is the fact that emergent system properties cannot be deduced; they must be observed, either in actual operation of the real system or in an appropriate system simulation. For all practical purposes, this means that simulation at a system level is key to safety assurance of complex systems, as presented in a precursor to the OSP initiative in *Ludvigsen et al. (2016)*. The same is true for other system properties such as fitness for purpose, reliability and performance.

To solve these challenges, and facilitate more efficient and effective design, commissioning, operation and assurance of complex, integrated systems, it is apparent that the maritime industry must stand together in a collaborative effort to find common ways of working and common standards. This is the background of the OSP Joint Industry Project (JIP), which was founded in 2018 by DNV GL, Kongsberg Maritime, NTNU, and SINTEF Ocean and subsequently joined by 20 industry partners spanning the maritime value chain. The grand vision of the OSP is to:

Enable collaborative digital twin simulations to solve challenges with designing, commissioning, operating and assuring complex, integrated systems.

For more in-depth information, access to the project deliverables and documentation, please visit www.open-simulation-platform.com.

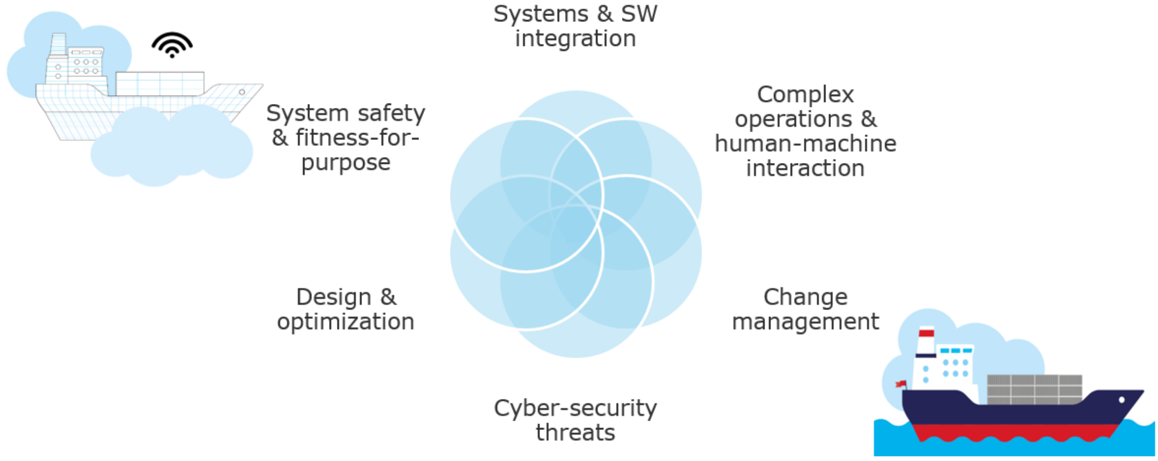


Fig.1: The challenges of complex systems

2. OSP key properties and structure

From the outset of the OSP initiative, it was clear that a new approach to system simulation would be needed. Some of the key properties and principles that have been at the heart of the development are illustrated in Fig.2:

- Model re-use: Simulation models and digital twin components are increasingly used by individual manufacturers in their internal working processes, but typically not made available to other stakeholders. To achieve scale and efficiency in building digital twin systems and assets we need to re-use these individual digital twin components also in other contexts.
- Protection of IPR: In order to share and re-use digital twin components, it is a prerequisite that no IPR contained in the component is exposed to other parties without the IPR owner consent. This requires black-boxing the individual components, and also making sure they cannot easily be reverse engineered. Critical IPR in a digital twin component would typically be the physics model of the equipment and its performance, and the control system software including HMIs. Various ways of black-boxing are discussed in *Cabos and Rostock (2018)*, and not further elaborated here.
- Co-simulation: A major challenge is to enable integration and orchestration of individual models and digital twin components originating from different stakeholders, simulation tools and platforms. Co-simulation solves this by decoupling the system simulation in its individual components, with each component containing a solver and executing independently of each other, synchronizing and exchanging information only at fixed intervals determined by the co-simulation engine.

- Common standards: To enable efficient construction of large co-simulations from a range of individual components, interoperability and common standards are needed. This pertains to both run-time requirements and component interfaces. We have set out to build on established standards wherever possible, only adapting and expanding on these when necessary for our maritime context.
- Open source: In order to drive standardization, ensure interoperability between stakeholders and tools, and enable common working processes, we strongly believe that a common, open-source software for co-simulation is needed. The idea is that any stakeholder can integrate the open-source software in their own working processes and tools, thereby simplifying cooperation and ensuring a common ownership and further development of the software.
- Collaboration: The challenge of complex systems is too big to be taken on by any individual stakeholder or tool. Therefore, the OSP focuses on collaboration and interoperability between tools, platforms and stakeholders.

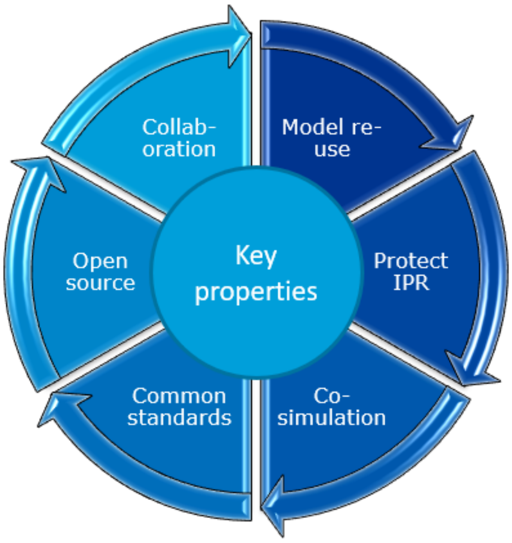


Fig.2: Key properties and principles of the Open Simulation Platform initiative

Staying true to these principles, the OSP JIP developed a number of open source deliverables to the maritime industry, all of which are further elaborated in subsequent chapters.

- OSP Interface Specification:
  Since we want to avoid re-inventing the wheel in any way, building on the Functional Mockup Interface (FMI) standard and a growing number of sister standards from the Modelica Association, https://fmi-standard.org/, appears to be a good choice. However, there are maritime specific needs that must be addressed. The first goal of the OSP Interface Specification is to reduce the effort required to construct a simulator and hence make it cheaper and more affordable, which can open up for a more widespread use of simulators in existing and new applications. The second goal is to make the connections between the simulator and a control system easier. Finally, we want to make models more available. The OSP Interface Specification is published free of charge under the CC-BY-SA license, with the ambition of maturing this into an established industry standard. A software tool for verification of model interfaces per the defined specification is also available.

- libcosim
  The OSP libcosim is an open source C++ library for co-simulation, supporting the OSP-IS. It is built on a combination of technical solutions from the OSP initiators and other established industrial solutions, in particular FMI. FMI has been driven primarily from the automotive industry, where they have similar challenges with complex, integrated systems and software. The number of modelling/simulation tools that support FMI export is already large and

increasing, enabling generation of so-called Functional Mockup Units (FMUs) that adhere to the FMI standard for co-simulation. The key working principle has been to use the best available technology and knowledge to meet the requirements for the OSP to efficiently serve its purpose. The libcosim source code is available on GitHub, free of charge, licenced under MPL 2.0.

- OSP Reference Models:
  This is a set of reference models including some of the most common marine systems and ship dynamics, adhering to the modelling guidelines and interface standards in the OSP Interface Specification. The reference models may be freely used and further developed by other parties, free of charge, and is available on GitHub together with OSP libcosim through the MIT open source licence.

- Use Cases:
  Three use cases, together spanning across the ship lifecycle, have been developed together with the JIP partners to test and verify the OSP software deliverables and standards, to explore value creation from the use of OSP, and to investigate possible business models.

3. OSP software

The open source software produced in the OSP JIP consists of the elements listed below, all further explored in the following sections:

- C++ co-simulation library – libcosim
- C and Java wrappers for libcosim
- Demo application
- Command-line interface
- Model interface validator (see Chapter 4.3)
- Example configurations (see Chapter 5)

3.1 libcosim C++ co-simulation library

The C++ library orchestrates the co-simulation of models that conform to the Functional Mock-up Interface (FMI). There is currently support for FMI 2.0 and FMI 1.0 and libcosim will be updated to support FMI 3.0 when this standard becomes operational. The structure of the co-simulated system can be defined using the System Structure & Parameterization (SSP) standard or the OSP configuration that makes use of the new model interface standard OSP-IS (OSP Interface Specification) developed in the OSP project. Both standards can be used to include FMUs, defining their interconnections, parametrization for simulation variables and individual model step sizes. The benefit introduced with OSP-IS is the possibility to defining FMU interconnections at a higher level, i.e. variable group connections rather than scalar connections. Additionally, it will enable automatic unit conversions and provides the a priori validation of the system configuration semantics utilizing an ontology that models the connection types, see Chapter 4. There are example configurations containing FMUs and co-simulation configurations exemplified in both OSP configuration and SSP available.

The conceptual architecture of the co-simulation is shown in Fig.3. The models are connected using the standardized FMI co-simulation interface. Based on the input-output variable mapping included in the configuration file, the OSP master algorithm routes the data between the models.

The co-simulation library includes a fixed-step master algorithm with a configurable base step size for data synchronization between the simulation models. It allows individually defined synchronization time steps for the models, with the requirement of being a multiple of the master algorithm base step size. During simulation, there are features to observe and manipulate simulation variables either through dedicated interfaces or with a scenario runner. It is possible to apply arbitrary operations to variable values without modifying the subsystem models themselves. Examples include arithmetic, logical operations, transformations, unit conversions and more. This is a necessity if variables connected between models must be summed or transformed by linear transformation or between coordinate frames. There are interfaces for client applications to include their own implementation of many of the included features.

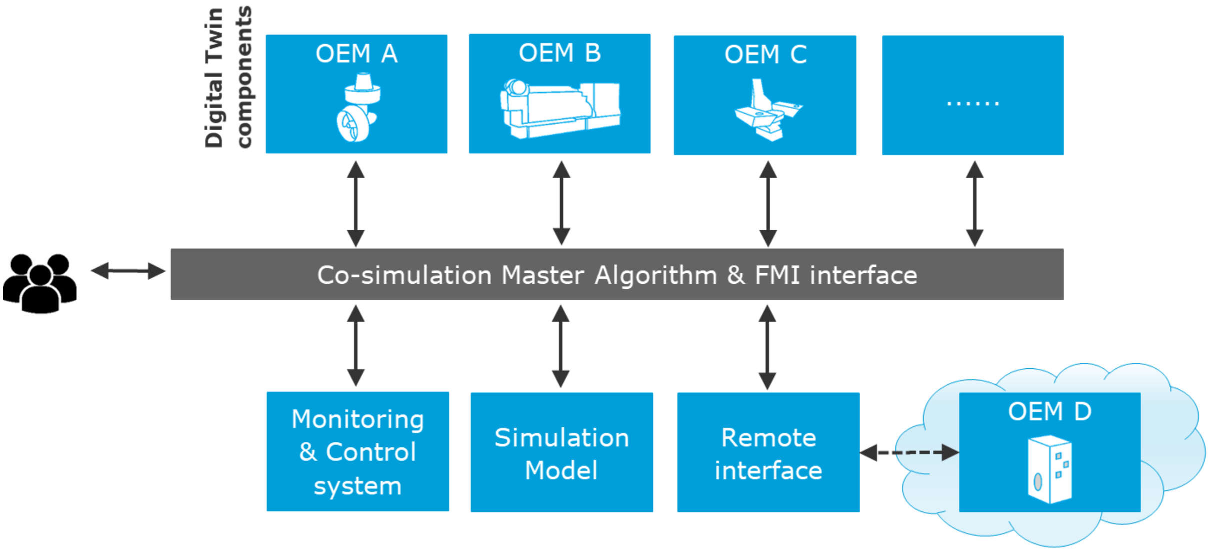


Fig.3: The OSP architecture, with digital twin components from various Original Equipment Manufacturers (OEMs), all connected in one co-simulation

Distributed simulation is currently enabled through integration with the software FMU Proxy developed at NTNU. Support for the Distributed Co-simulation Protocol (DCP) standard is planned for a future release.

The OSP software deliverables also includes two wrappers for libcosim that are helpful for developers who want to use libcosim with other languages than C++. The C wrapper, libcosimc, is applicable for many developers as most programming languages can integrate with C. The Java wrapper, cosim4j, makes libcosim available for Java developers by the use of Java Native Interface.

3.2 Demo Application

The demo application is developed to showcase how libcosim can be used in an application and to provide a graphical user interface to manage co-simulations. It allows for loading configuration, controlling the co-simulation and easy exploration of the co-simulation library's features. There are options for real time execution or simulation as fast as the hardware allows. During simulation, the user can observe simulation variables, manipulate these and visualise the simulation with plotting features. A scenario management feature allows change of the simulation variables automatically based on a scenario file. It is possible to plot and log the simulation variables of all simulation models in dedicated data files based on predefined rules, see Fig.11 for a screenshot of the user interface.

3.3 Command line interface

For the cases where a graphical user interface is not available or needed, libcosim comes as a command line application. It facilitates for running co-simulations of systems configured with SSP or OSP configuration. The command line application named cosim is suitable when multiple sets of simulations can be executed in parallel. An example for such use case could be in design, where there typically are multiple system variants and parametrizations of the system.

The command line interface also allows for simple debugging features. It can simulate single FMUs and show the model descriptions. Besides being a necessary simulation tool, the command line interface also serves as a demonstration for how client applications can make use of the co-simulation library.

4 OSP Interface Specification (OSP-IS)

Integration of models or sub-simulators to construct system simulations is challenging as there is no established interface standard. The FMI standard offers a partial solution to this problem by ensuring binary compatibility for the model connections, however semantic correctness is not addressed. OSP Interface Specification, www.opensimulationplatform.com, is developed to make model connections simpler and semantically correct. That is why OSP-IS at its core is a specification for connection of metadata to models and model interfaces. The metadata is specified in an ontology, and the connections between the metadata and the model interface signals are specified in a new file called OSPModelDescription.xml.

4.1 Main concepts in the ontology

The ontology consists of definitions of concepts which can be used to describe the model interface signals. The concepts also contain rules for connections, which can be used to deduce how models can be connected. The following is a short description of the most central definitions in the ontology:

- Variable: A variable has a unit, causality and data type.
- Units: The unit definition in the ontology is based on FMI 2.0 and the Units of Measure Ontology
- Causality: [a variable is an] input or an output
- Variable groups: Collection of variables that together form some real-world phenomena.
- Ports: A special case of variable groups that must contain exactly two variable groups, where one of the variable groups must contain variables with input causality and the other variable group must contain variables with output causality. This can be used to model connections according to bond graph theory, *Karnopp et al. (2020)*, which is a convenient way of dealing with model to model connections where energy (or mass) is transferred between the models.

4.2 OSPModelDescription.XML

Since FMI only deals with binary compatibility, it also lacks a specification for linking semantic information to the model and the model input/output variables. The OSP-IS ontology provides a representation of the semantic information that can be used to describe the variables. To use this semantic information to describe the variables, we have proposed a new XML format for linking the signals in the FMU modelDescription.xml to the concepts in the ontology, www.open-simulation-platform.com.

4.3 Validator

A validator has been developed to verify that the OSPModelDescription.xml is valid. The validator loads the OSPModelDescription.xml and the ontology, and checks that the configuration is correct. It can be used to verify that a given FMU with an OSPModelDescription.xml complies with OSP-IS. Or, it can be used to verify that the connections between two FMUs are semantically correct. Some consistency checks are handled through the xml schema, while the more advanced are handled by the validator. More information is available in the OSP-IS document.

4.4 Usage

One of the main benefits of OSP-IS is that it enables the implementation of multivariable connections with simultaneous verification of semantical correctness. Today, connection of two FMUs is a matter of connecting signal by signal, interpreting the signal names and deducing which connections are probably correct. Also, ensuring matching units might be a matter of consulting documentation, or possibly communication with the model developers, Fig.4.

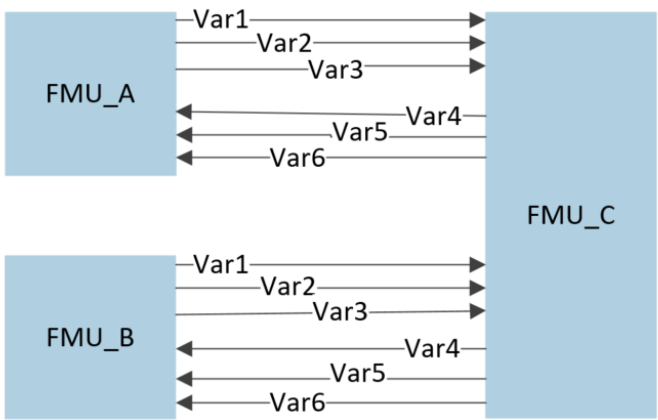


Fig.4: FMU variable connections today

With OSP-IS it is possible to connect groups of signals simultaneously by grouping signals representing the same real-world phenomena into one connector, Fig.5. Furthermore, it is possible to implement automatic checking of semantic correctness. Which could be used to guide the person performing the connections, possibly by disabling invalid connections as shown in Fig.6.

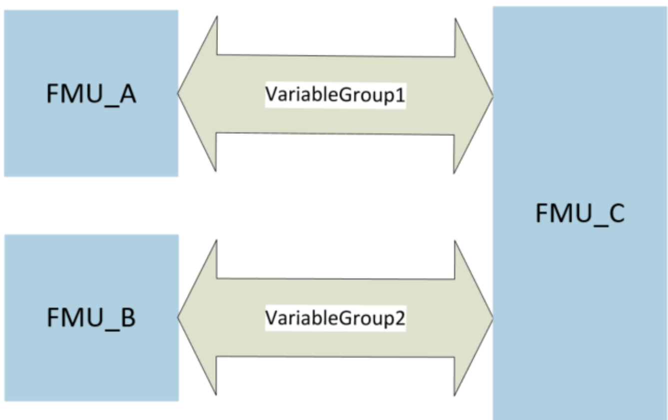


Fig.5: FMU multivariable connections

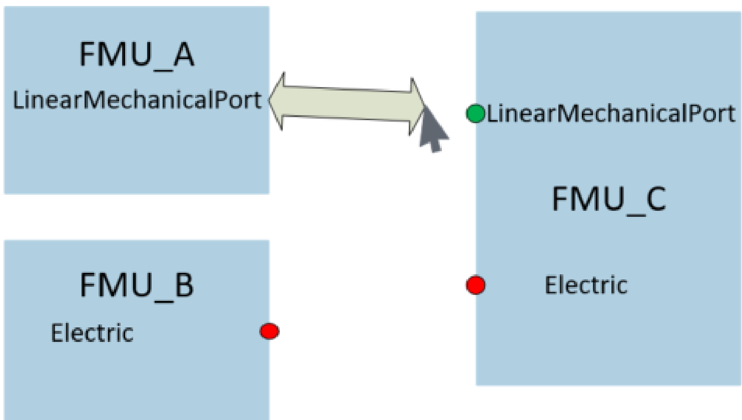


Fig.6: FMU multivariable connections with semantic correctness checks – example where one valid connection can be made.

Automatic unit conversion: The unit part of the ontology is developed such that it is possible to check if two signals are of the same unit. If they are, they can be connected, if they are not it is possible to check if they can be converted to the same unit and then connected. If they can be converted, the conversion factors can be retrieved from the OSPModelDescription.xml and a conversion can be inserted.

5. Maritime reference models

The motivation for providing maritime reference models is to lower the threshold for using OSP by providing example implementations that can be freely used and modified. The following principles have been followed:

- Models should as far as possible be generic and configurable
- Where possible, models should include both source code and compiled FMUs
- Models are implemented according to the OSP Interface Specification
- The models should support use-cases with various applications

Following this, the OSP maritime reference models is a collection of models representing typical maritime equipment. A set of example configurations are included where models are connected to each other and control systems in order to run model-in-the-loop (MIL) system simulations. The collection also includes examples of FMUs that provide network communication to send data to external applications, for instance control systems. This allows system integrators to connect external systems to the simulator without knowing the details of the communication protocol used by the supplier. Connections between models and control systems are handled like connections between models.

The reference models include system configurations and can be useful when testing the cosim demo app or your own implementations of libcosim and OSP-IS. The different configurations showcase different aspects of the Open Simulation Platform:

- Use of the OSP-IS, where standard variable groups are defined for more effective connections.
- Examples of the OSP configuration files, implementing variable groups, summation of signals outside FMUs and linear transformations between FMUs.
- Use of fmu-proxy for distributed co-simulation.
- Use of network communication inside FMUs.
- Use of the scenario runner to manipulate simulation variables
- Use of plotting configuration files to set up trending.

The following examples of system configurations are available:

- R/Y Gunnerus with path-following control
- R/Y Gunnerus with DP control and a crane (use case 3 below)
- R/Y Gunnerus with simplified models and DP control, which can be run on one PC or two PCs using network FMUs as communication.
- DP-Ship
- Launch and recovery system (LARS)
- Construction vessel with DP, a winch model and winch controller, and power system modelling
- Configurations with a sensor model FMU and matching NMEA communication FMU, e.g. a gyro sensor module and HEDTH NMEA sentence.

Table I shows an overview of reference model categories and tools/codes used in these configurations.

Table I: Maritime reference models

Model categories	Tools/codes
Vessel + environmental model	Simulink, VeSim, 20-Sim
Thrusters + main propulsion	C++, Simulink, Open Modelica, VeSim
Power system	Simulink, 20-Sim
Sensors and position reference system	C++
Deck machinery	SimulationX, 20-Sim
Control systems and estimators (DP, path-following)	Java, C++, Simulink, 20-Sim
Interface	C++

All models are available on Windows, while some are also available on Linux. Source code/model code for most models are available.

6. OSP use cases

To demonstrate the Open Simulation Platform the OSP Joint Industry Project has developed three use cases. The purpose has been to give users a sense of how the platform works and can generate value for the industry, as well as providing important feedback to the OSP development.

The use cases have developed simulation set-ups using OSP software and interface specifications for:
1. Design of a hybrid ferry propulsion system
2. Virtual commissioning of a coastal service vessel
3. Operational planning for a crane operation on R/Y Gunnerus

6.1 Design of a hybrid propulsion system

In the first use case, we demonstrate the capability of OSP as a collaboration tool in the conceptual design phase where the uncertainty of the design is high, and one needs to explore as many options as possible to find the optimized design to start with.

The target system for this use case is a hybrid power system in a passenger ferry, combining an electrical propulsion system with energy storage. This is an industrial area where technology is rapidly developing in order to minimize emissions and fuel consumption. Having batteries combined with electric and mechanical power distribution systems allows the plant to operate beyond the strict energy balance requirement as in conventional systems and allows a designer flexibly to choose the rated power of the main power sources.

However, this added degree of freedom in operation and design comes with challenges in the complexity of design. For example, the sizing of the energy storage system depends on the actual load profile and control strategy of the system. And the number of possible configuration and a set of dimensions for the power sources for evaluation increase significantly. In this regard, static analysis for the power balance does not provide the enough design input for the hybrid system. Furthermore, the knowledge of each main component of the system still remains with the specific supplier and makes it challenging for designer to create all the necessary knowledge to perform the system assessment.

The use case demonstrates how to tackle such challenges by using simulation models as a tool for description of requirements, evaluation of designs and system integration, and for assuring the transfer of knowledge through different stages of the design. Furthermore, system simulations enable the designer to efficiently explorer a large design space which is essential for designing a system with large flexibility in an optimal way. Open Simulation Platform allows co-simulation where multiple parties provide their own component models in FMU form to a system integrator without risk of disclosing their proprietary knowledge. The system model for the simulation is presented in Fig.7 and shows the partners that have delivered models to the set-up.

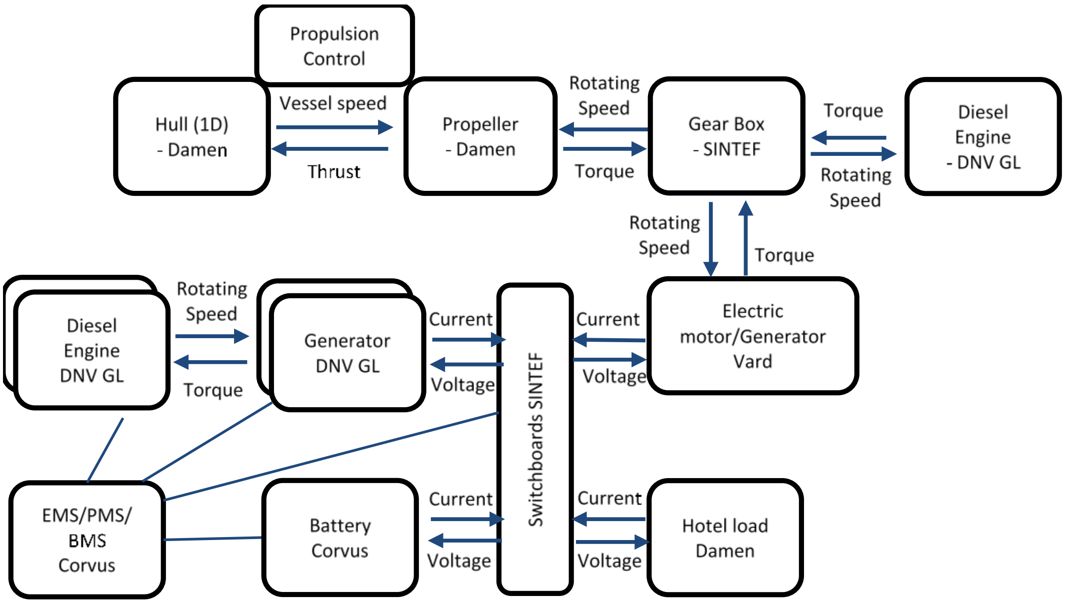


Fig.7: The system structure and interface for the components of the ferry with a hybrid power system

The first challenge for the design case was to choose proper models. A proper model has a certain level of fidelity to answer the question with acceptable accuracy and can be run fast enough with reasonable resources. When the system is multi-disciplinary, this is especially important as it is necessary to establish the balance of the fidelity among different component models.

A second requirement was to have a co-simulation environment that enables to run many simulations in parallel, this was solved by using the command line interface (CLI) cosim. Using cosim, retrieving the information from the FMU component model, testing the component and running system simulations could be easily integrated in a design tool developed as a demo application for the JIP project. In this integrated tool, a user can manage the FMUs of different sources, configure a system and scenario and run many simulations in parallel, and present the simulation results. All the created parameters, configurations, FMUs and results are saved in a database.

A simulation study is performed with two design parameters: rated power for the gensets and power upper limit for zero-emission operation of a given load profile for the passenger ferry. Combination of these two parameters are sampled using Sobol sequence sampling in order to effectively cover the design space. The simulation is run in two stages for each case. The first simulation determines the rated capacity of the battery depending on the power and energy usage during the simulation. Then, the second simulation is performed to find out actual fuel consumption and running hours which are key contributions to OPEX.

6.2 Virtual commissioning of a complex integrated system

The main objective of the virtual commissioning use case was to show that the OSP software, libcosim, could be used as a virtual commissioning tool for control system(s) software. When performing virtual commissioning, the suppliers of equipment to a vessel will share their models and digital twin components, including exact copies of control system software embedded in the physical systems, such that system simulations can run to perform integration testing with multiple scenarios far exceeding what is possible to test during a traditional sea trial. The concept is shown in Fig.8.

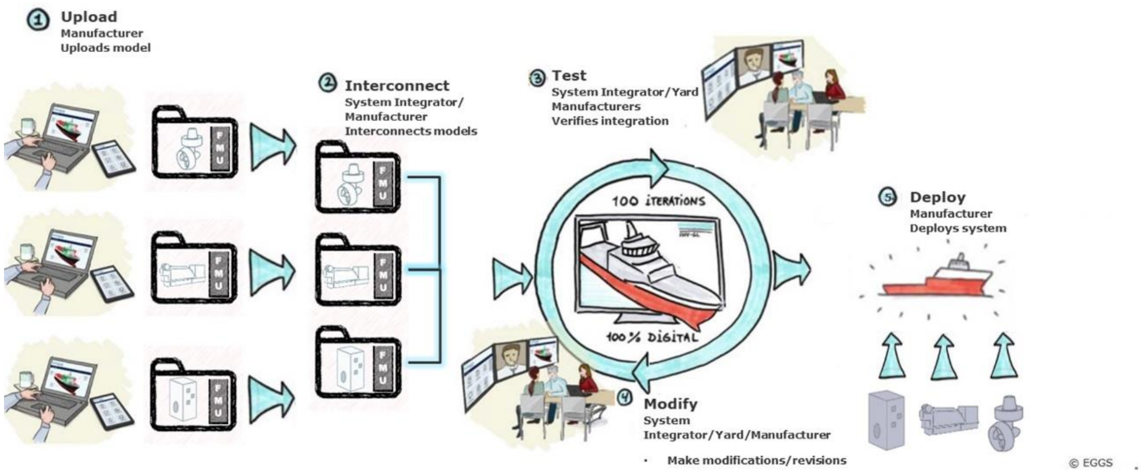


Fig.8: Virtual commissioning

A key element of the Open Simulation Platform is that stakeholders can collaborate in building the digital twin vessels based on common standards. This is necessary in order to create an efficient integration process.

In order to demonstrate OSP software for virtual commissioning, the use case has covered several aspects:

- Combining multiple control systems running in closed loop with simulation models (FMUs) where FMUs and control systems are delivered from multiple suppliers.
- Simulations run in real-time and models with sufficiently high fidelity to operate control systems as if they were controlling the real asset.
- Handling control systems with different sampling rates.
- Performing efficient interfacing, both from models to models, and models to control systems.
- Running software-in-the-loop-testing in the virtual setup to replicate testing during real commissioning.
- Virtualization of selected control systems and control system HMIs

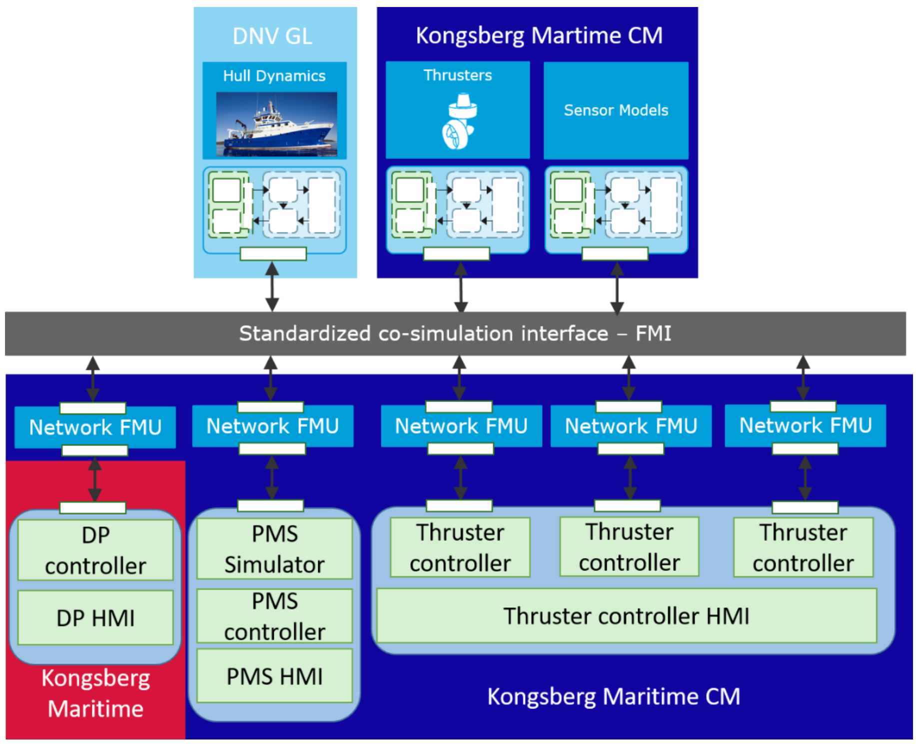


Fig.9: Virtual commissioning use case setup

The setup is shown in Fig.9 and includes two different control systems from Kongsberg Maritime CM (former Rolls-Royce Marine), and a K-Pos DP system from Kongsberg Maritime. On the simulation side, a hull model from DNV-GL, and thruster, power system and sensor models from Kongsberg Maritime CM has been included. Due to the large number of signals needed for the PMS in this case, the main part of the power simulation had to run on the control side, using an internal PMS simulator.

In addition, a set of network FMUs have been developed to take care of communication between the control systems and simulator. Network FMUs are network protocols wrapped as FMUs. These enable suppliers of control systems to implement their preferred protocol, wrap it as an FMU, and deliver it to the system integrator along with the control system and any model FMUs. This means that the simulator core does not have to support any specific interfaces. The communication between network FMUs and model FMUs is set up by the system integrator in the same way as model FMU to model FMU communication is set up. Also, since libcosim and the cosim demo app is open source, any supplier can use these to verify that models and Network FMUs are running according to expectation before supplying them to the system integrator.

To keep the communication modular and preserve system topology, it was decided to use several network FMUs for each control system, depending on the type of interface. As an example, the DP system requires sensor and position reference feedback from the simulator in order to perform station keeping of the simulated vessel. This information is normally communicated using one or more standard NMEA 0183 messages per device (sensor/position reference system). Since the number of devices varies from vessel to vessel, it was decided to make one network FMU for each device, enabling reuse and a one to one relationship between sensor models and network FMUs.

The main challenges when integrating systems from several suppliers will be the large number of signals for some systems and lack of standardization of interfaces and models in the maritime industry. The work done on connections in the OSP project has helped improve this, but further work on standardization and good tools for connecting control systems and models will clearly be very important in the future.

6.3 Co-simulation set-up for planning of an offshore crane operation

A strong motivation for the establishment of a system digital twin is easy access to simulation results during vessel operation. New tools and services for operational planning based on co-simulation is one application that can add significant value to vessel owners and operators assuring efficient and safe operations. A use case for operational planning was therefore established to investigate how the OSP software can:

- simulate an operation to support choice of operational strategies and system settings
- efficiently simulate changing environment (wind, waves and current) to investigate limiting conditions for specific operations
- handle the challenge of simulating tightly coupled systems (for example crane and vessel interaction)
- provide output for visualization tools considered essential for training and operational planning

The rationale behind using Open Simulation Platform for operational planning is that:

- Model protection enabled by use of FMI standard means less need to share system specific information (IPR) between different stakeholders
- That simulation tools can be easily updated by suppliers to reflect the current status of the vessel, through replacing or updating component models and software
- Reuse of models from the design and commissioning phase is facilitated by complying with common interface specifications, increasing the value of investing in validated and “flexible” models for a newbuild project

The NTNU research vessel R/V Gunnerus performing a crane lift operation for installation of subsea equipment was chosen as basis for this case study. Component models (FMUs) representing Gunnerus hull and main propulsion system were already available developed by Sintef Ocean and NTNU in previous research projects such as ViProMa, https://viproma.no/. In 2019 Gunnerus hull was lengthened midships and a new knuckle boom crane was installed, the set of FMUs are therefore updated with a modified hull model and a new crane model.

Re-implementing the use-case in OSP software environment demonstrated that libcosim and the cosim demo app was very efficient and stable for running simulations. The crane model runs at 1000Hz while the remaining models run at 20Hz. Challenges when establishing the use case emerged mainly from lack of documentation of the FMUs and their interfaces. In addition, the Sintef software VeSim generates FMUs that must be run in a distributed environment making debugging of system set-up slightly more challenging. Fig.10 shows the components included in the simulation set-up, where green frames marks components run by distributed co-simulation using the FMU-proxy framework, https://open-simulation-platform.github.io/libcosim/fmuproxy.

An updated interface configuration for the models and configuration according to OSP-IS was implemented using the configuration tool kopl, https://open-simulation-platform.github.io/kopl. The motivation was to allow for more efficient and verifiable reuse of FMUs and system configurations in the future.

A crane welded to a ship is tightly coupled through causality and pose challenges for time-domain simulations, *Sadjina et al. (2019)*. An optimal numerical solution would be to combine crane and vessel model in one FMU. To allow separation between suppliers, the use of separate FMUs was demonstrated in this use case. Coupling of crane and vessel was solved by having vessel motion as input to the crane and the reaction forces from the crane are input to vessel model. The vessel model represents hydrodynamic properties of hull calculated for a specific center of gravity. When changing position of crane and load on a vessel these properties will change as center of gravity changes, thus the current set-up must be used with care or be further developed for detailed planning of heavy lift operations. Methods to efficiently use co-simulation for tightly coupled maritime systems is suggested as topic for further research.

Scenario files are implemented and used to specify crane operation, environment and DP control parameters. Fig.11 show the cosim demo app plotting the crane tip position (NED frame, relative to vessel center of gravity).

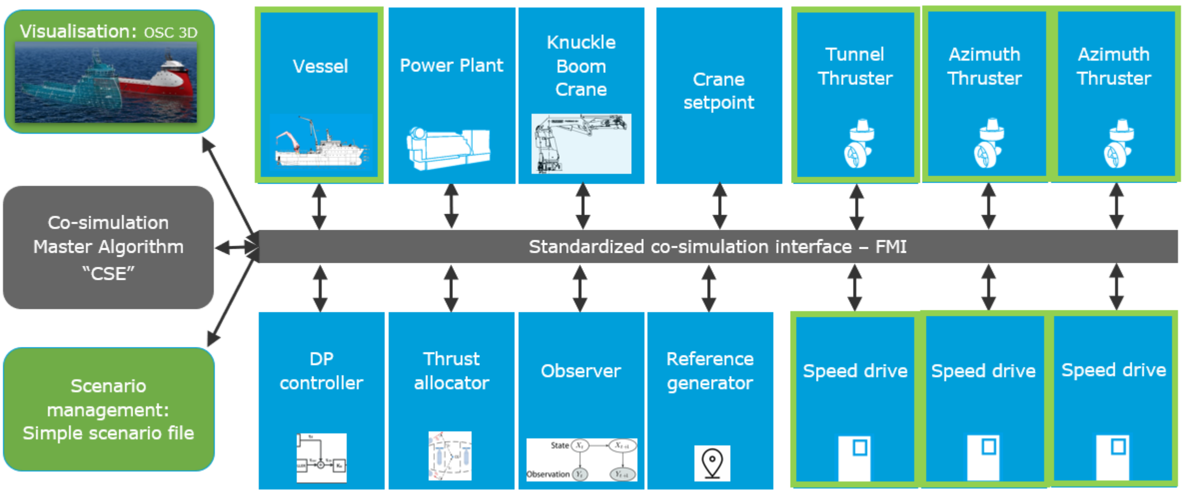


Fig.10: Simulation setup for use case 3

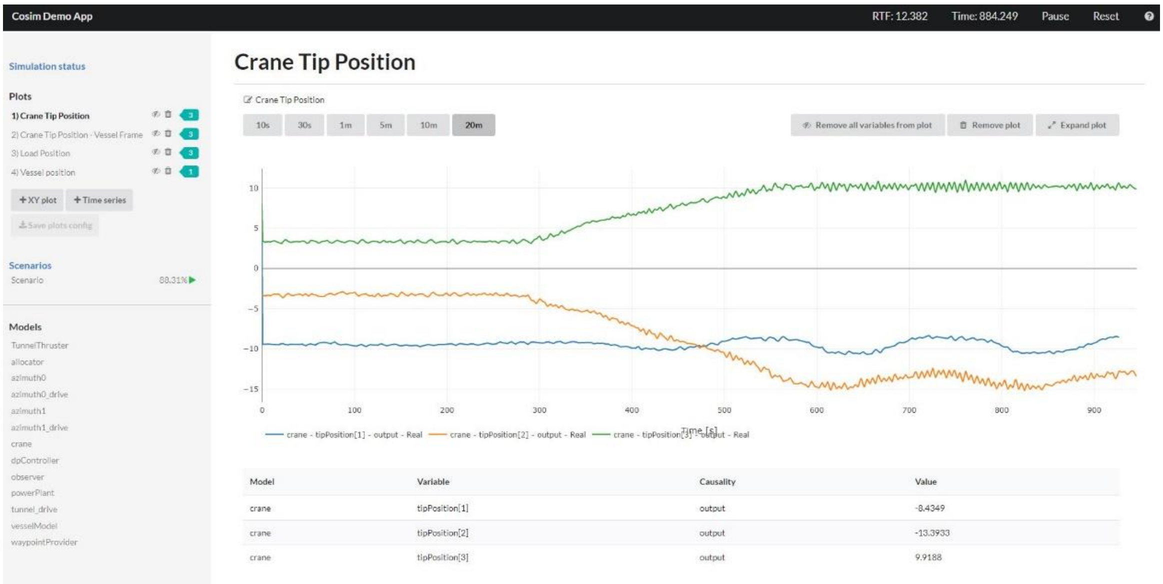


Fig.11: Time series of crane tip position from the Gunnerus case study

6.4. Next step: the OSP Open Source Community

Closing of the OSP JIP in Q2 2020 marks the start of a new phase of the initiative, where further development and maintenance will happen as a regular open source project. The ambition is to establish a living user community spanning both industry and academia, that will use the deliverables and contribute to their further development. The founding partners (DNV GL, KM, NTNU and SINTEF) have agreed to establish an OSP open source steering committee that will govern and maintain the software and specifications going forward. To succeed with the open source community, we will need contributions and commitment from both existing and new industry partners to maintain and develop the OSP components further. This can be supported through existing and new research projects building on or using the OSP, and we encourage the maritime industry to engage and contribute to develop these next steps.

7. Summary and benefits

The Open Simulation Platform is an open-source industry initiative for co-simulation of maritime equipment, systems and entire ships, aiming to help the maritime industry master the design, construction, operation and assurance of complex, integrated systems. Recently, several deliverables were made available on GitHub and through the project website, www.open-simulation-platform.com:

- An open source co-simulation library
- An interface specification for maritime simulation models
- A set of open source maritime reference models
- 3 use cases that demonstrate the usage and value.

The project consortium believes the OSP gives the industry a number of potential benefits, and aims to engage more stakeholders in further development of the open source project. Some of the key benefits are:

- The OSP gives industry a standardized way of running co-simulations and ensures interoperability between tools and platforms.
- Builds on established standards: FMI, SSP, DCP - all maintained and further developed by the Modelica Association
- The MPL 2.0 open source license model gives all organizations the opportunity of integrating the OSP deliverables in own tools and applications

- Community driven development on GitHub, facilitated by the OSP founding partners
- libcosim C++ co-simulation library can be integrated in any co-simulation architecture, local or cloud
- OSP Interface Specification enabling easy integration and connection of models

Acknowledgements

The OSP is supported by project grant numbers 295918 and 280703 from the Norwegian Research Council.

References

CABOS, C.; ROSTOCK, C. (2018), *Digital Model or Digital Twin?*, COMPIT Conf., Pavone

DNVGL (2020), *Technology Outlook 2030*, https://www.dnvgl.com/to2030/impact/impact-on-maritime.html

KARNOPP, D.; MARGOLIS, D.; ROSENBERG, R. (2012), *System Dynamics: Modeling, Simulation, and Control of Mechatronic Systems*, Engineering Pro Collection, Wiley

LUDVIGSEN, K.B.; JAMT, L.K.; HUSTELI, N.; SMOGELI, Ø. (2016), *Digital Twins for Design, Testing and Verification Throughout a Vessel's Life Cycle*, COMPIT Conf., Lecce

SADJINA, S.; KYLLINGSTAD, L.T.; RINDARØY, M.; SKJONG, S.; ÆSØY, V.; PEDERSEN, E. (2019), *Distributed Co-Simulation of Maritime Systems and Operations*, J. Offshore Mechanics and Arctic Engineering 141/1, p.011302

The Effectiveness of VR in Maritime Education

Harmen van der Ende, MIWB, Terschelling/Netherlands, harmen.van.der.ende@nhlstenden.com
Maarten van Asselt, MIWB, Terschelling/Netherlands, maarten.van.asselt@nhlstenden.com
Herbert Koelman, MIWB, Terschelling/Netherlands, Herbert.Koelman@nhlstenden.com

Abstract

This paper describes a research of first year students on the use of virtual reality in simulation. The Maritime Institute Willem Barentsz has positive experiences with Virtual reality with small groups of students, (max five) on a full-mission simulator with big screen and dedicated hardware. In this experiment two separate groups are tested with and without VR in a classroom layout with only a a computer, two small screens and a big screen. a total of four exercises were used and after these exercises the students had to fill in a questionnaire and this was analysed.

1. Introduction

VR simulation is commonly used for many years in different kinds of industries, like gaming, medicine, architecture, science and also in maritime applications. Generally, it is used for different purposes such as design and concepting. With the use of VR new designs can be explored before they are built. This way placement of the different systems and different objects can be chosen and easily moved around in order to make it fit and place it ergonomically. In addition, also design errors can be detected, *Hu (2018)* Another common use of VR is in sales, e.g. future customers can see beforehand what the design of their yacht will look like and they can change the interior and alter details to match their expectations in advance. Another example of a company who benefit of the use of VR, a combination between simulation, design and sales, is Heerema Marine Contractors the heavy lifting company. This company uses a very detailed crane simulator, https://heerema-production-content.s3.amazonaws.com/HMC/Media/Brochures/Simulation-Center-Brochure.pdf, where realistic mathematical models can be built into a complete model of their crane ships, including the load which needs to be lifted. These loads are always complex in shape and weight. With this model, different situations can be calculated in advance and this has the following benefits: optimizes the operational efficiency, interfaces to lower overall project costs, enhances execution predictability and timing, furthermore it minimizes unexpected issues. The simulator runs through failure cases and people can create mitigating solutions. Resulting in providing training for more complex projects when necessary. Additionally, it can be used to test new and innovative concepts. VR is also considered very important in education as training tool to improve skills, *Bertram and Plowman (2018).*

2. History

Since 1993 the Maritime Institute (MIWB) provide simulator training for all Nautical Colleges in The Netherlands. The institute is equipped with different types of full mission simulators: a full mission bridge simulator, a cargo simulator and an engine room simulator. These simulators are class A simulators, *DNVGL (2017),* and are used for sea time reduction of Dutch maritime students, *Uitterhoeve and Werner (2018).* The Class A bridge simulator is equipped with a 360° 3D view presented by Barco F35 beamers, https://www.barco.com/en/Products/Projectors, on a 12 m wide and 4 m high screen. These screens offer the possibility to project a 3D view in different environmental situations, for example: day, night or fog. The engine room simulator consists of a physical control room and an engine room in which the VR engine room is projected on a wide screen. This projection can be controlled by Microsoft Controller and is directly connected to the engine room model, interaction can be done from the physical control room or from the engine room on the screen, or both, this can be pre-set by the instructor. This 3D projection was developed in-house by the Maritime Institute and is connected to different models of the K-Sim Neptune engine room simulator. The first model developed by the Maritime Institute was the Neptune Mak-11 Container model: a 150 m refer container vessel. Followed by a Neptune MC90 V Suez max tanker, M11-Suez max tanker model and lastly, a fire fighting simulator model based on the M11 Suez max model. Currently an LNG model is in development. The LNG model will be a very realistic 3D model consisting of all systems present in the engine room and on deck at the bunker station and will be available as a system in the K Sim.

All functionality included in the simulator model is also present in the VR such as valves opening and closing, controls can be set, push buttons can be used, starters from local to remote, etc. All models are realistic machinery, but also CCTV and avatars can be part of the system in the VR world. The VR world is currently used for 16 years as part of full mission training and approximately 8000 Dutch students have been trained with the system. According to a survey done with students who have been trained with the VR worldview, they have experienced it as very realistic and are very content. It needs to be noted that these students have had at least 2 years of maritime education and some have already been at sea. Resulting from the positive experiences from using VR in the full mission simulator the VR feature was also added to the part task simulators. A part task simulator is a desktop simulator consisting of a computer, a 22-inch screen, a mouse and a keyboard. At the Institute eighteen part-task stations are installed which can be used for bridge, radar, Ecdis and GMDSS training. Additionally, it is also used for engine room training. On this parts task simulator an extra 55-inch screen and a Microsoft gamepad was added, VR glasses were also considered but were rejected. Motivation for this choice is described below.

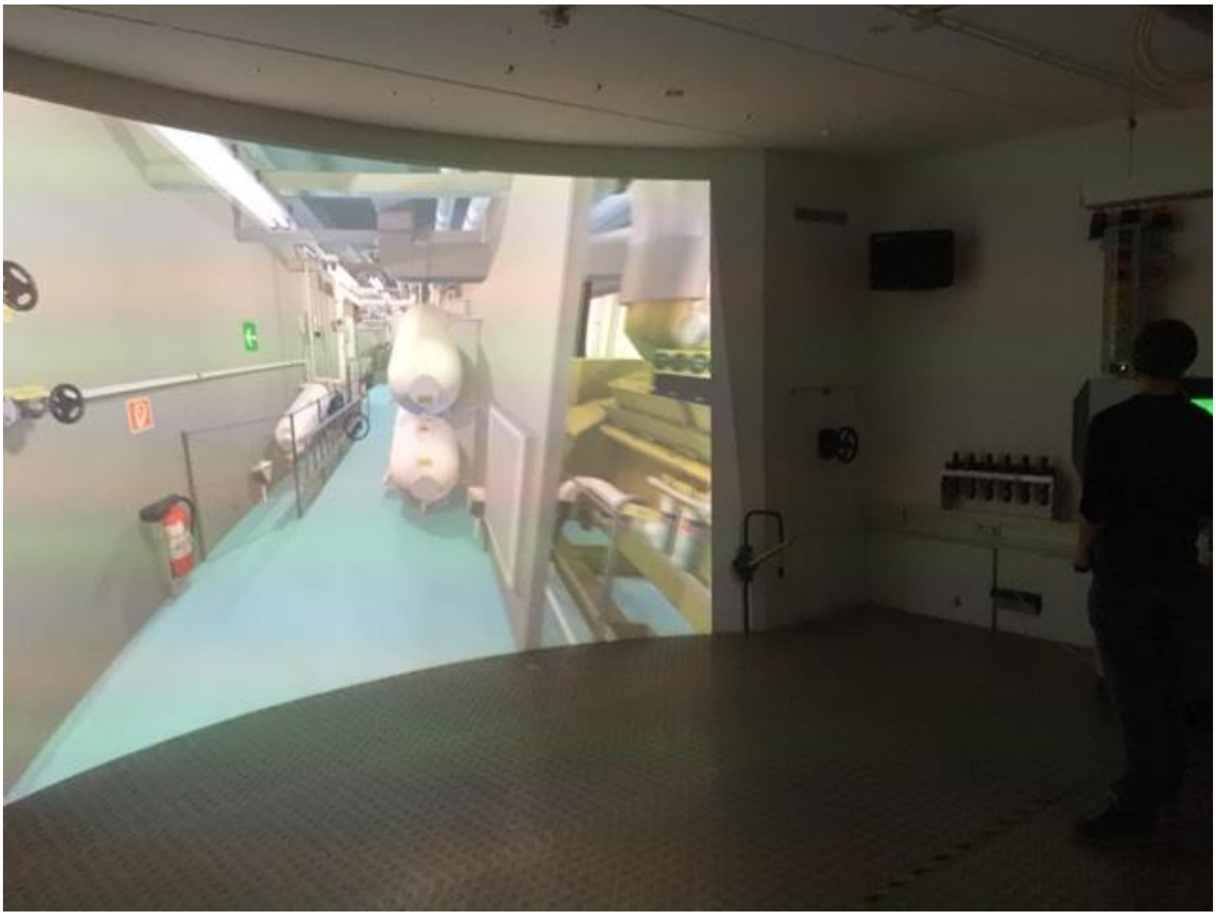


Fig.1: VR view engine room MSTC

2.1 3D world view

Two years ago, a decision had to be made between VR, https://www.vive.com/us/support/vive/category_howto/choosing-your-play-area.html, or VR on screen. Both options could be explored, because the full mission VR world was already present for a few years, Fig.1. Different pairs of VR Goggles were used on a VR engine room for testing.

Advantages VR on screen:

- Previous positive experience with the 3D world view.
- Size of the screen can be decided depending on the available space.
- Relative cheap investment.
- Instructor can monitor students' actions in the engine room.

Disadvantages VR on screen:

- Not emerged in the engine room world, less realistic.
- Outside distraction (from other students, environment).
- Gamepad, not realistic as in the real world

Advantages VR goggles:

- Realistic engine room experience
- Real movement in engine room (walking)
- No distraction of the surrounding
- Also, during the last years, VR goggles have improved significantly with regards to:
  - Increased speed of processors in the pc.
  - Improved resolution on graphical cards.
  - Decrease of weight, size and energy usage of VR glasses.
  - Improved network speed and WIFI.
  - Latency was less a problem.
  - Decrease of price.

Disadvantages VR goggles:

- Every station needs a special area to be setup in the play area; for 8 stations, this is not very practical.
- Battery life of the glasses is still relatively short and cannot last the whole exercise.
- Instructors have no view on what the students are doing in VR.
- Students can experience cybersickness. Cybersickness refers to the unwanted side effects arising from exposure to virtual environments (VE). It denotes an array of unpleasant physiological symptoms, such as nausea, dizziness, and lightheadedness, invoked in response to VR exposure, *Yu et al. (2019)*. This can influence the wellbeing of the student and the outcome of the training. For that reason, this is considered a big disadvantage of the use of VR glasses. This issue is less present in the newer VR models, but some students still experience cybersickness.

After a lot of consideration, about the space necessary, cybersickness with VR and looking at the investment a choice was made for 55-inch Phillips monitor screens with ambient lighting and a Microsoft Gamepad as controller. The monitor is attached on top of two 24-inch regular screens.

3. The use of engine room part task training of junior students

In the junior year of college for technical training the part task engine room simulator is used in the regular practical program. Every two weeks, the students have training on the simulator in groups of two students on eight part-task stations. One lecturer is present in order to provide instruction. Students can bring computers, books and may take notes during these exercises. All exercises are performed in groups of two students. Although the exercises are done in groups, every student must, individually, make checklists and Excel graphs of each exercise. At the end of the last semester an individual examination is done in which two random exercises are selected to assess the knowledge of the student.

For this training two different models of engine rooms are used: a CNTR M11 model of a container ship and the MC90V model of a Suez max tanker.

CNTR M11 container ship	MC90V Suez max tanker
Ms Elan Length over all 120 m Breath Moulded 19 m Draught 5,5 m Speed 17 kn Engine Mak 8M43C Continues service rating 8000 KW	Ms. Genmar Argus Length over all 305 m Breath Moulded 47 m Draught 30,4 m Speed 14 kn Engine Man B&W 5L90MC Continues service rating 18000 KW

The program consists of the following 8 exercises:

Exercise 1	Familiarisation exercise, starting a cool water system	Elan
Exercise 2	Start auxiliary engine, generator on the net, harbour condition	Elan
Exercise 3	Starting main engine, ready for departure,	Elan
Exercise 4	Departure to full sea speed, all systems running	Elan
Exercise 5	Fuel consumption and efficiency in different running modes	Elan
Exercise 6	Sankey calculations of the main engine	Elan
Exercise 7	Pressure volume diagrams and power calculations	Genmar Argus
Exercise 8	Fault finding in main engine by pressure volume diagrams	Genmar Argus

The container vessel Elan is used in the first four exercises and the main goal of these exercises is familiarisation with the systems, equipment, procedure training and how to start up a typical conventional container feeder vessel with a common four stroke main engine. The other exercises focus on common calculations of various items in the engine and on the ship. All these exercises are assessed with an individual practical simulation exercise in the last semester of the first year.

4. Research

In this section the experiment is described, how it is performed, the results of the experiment are discussed and statistically analysed.

4.1 Setup of the experiment

The study is performed as a semi experiment. Which means new situations are created and new experiences are gained, the effect of these new situations on the students is being observed.

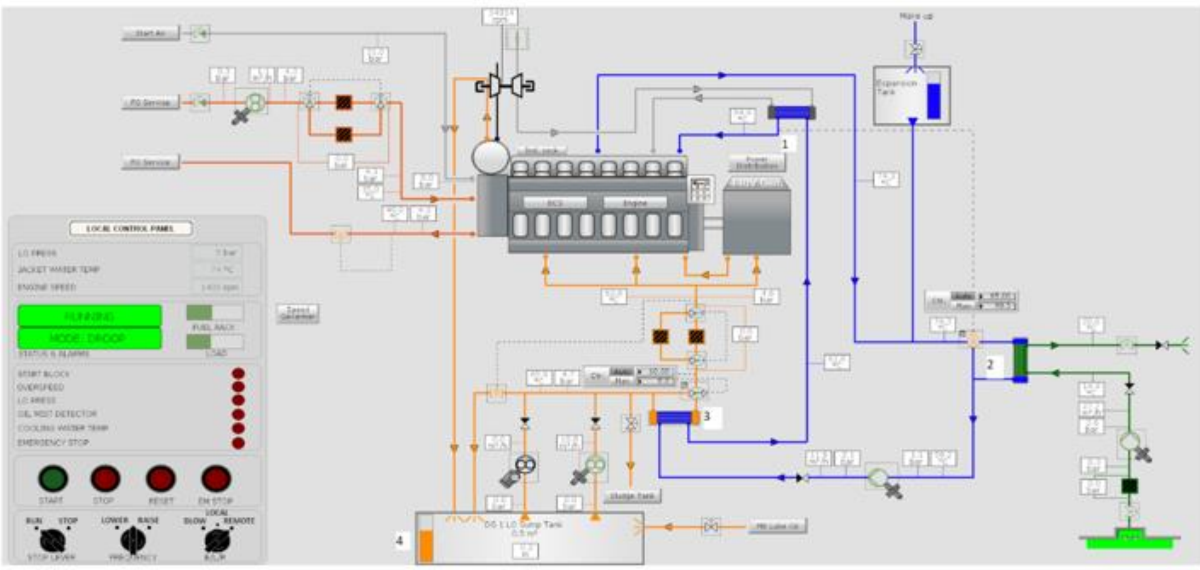


Fig.2: Mimic auxiliary engine

To increase the trustworthiness and validity of the research, the following measures are taken. Two groups are formed, the students are randomly picked per group. Each exercise is the same. Furthermore, the students have no maritime experience and have the same entry level.

The exercises involve starting the engine room systems, at first the seawater cooling water system needs to be started as part of the familiarization, followed by starting an auxiliary engine and connect it to the busbar. Lastly, bring the ship to harbor conditions and then make it ready for departure and subsequently the departure on to full sea conditions. These goals can be reached in different ways:

- Option 1: Only in the mimic systems (group number A)
  In option 1, the engine room can be started only with the mimic system. In this mimic system all necessary engine room components are presented and it's possible to open and close valves, start pumps, put controllers in automatic and different sensors readings are visible like temperature, pressure and flow. In addition, the main switchboard and power management, bridge and control room are available. In option 1, the 3D is not used.

- Option 2: Only in the 3D engine room
  Option 2 is only a startup in the 3D world. In the 3D world almost all systems are available, realistically and the layout is identical to the 'real' Elan. All the piping is mostly equal and interconnects different systems, therefore they can be used to figure out the lay-out. All valves can be opened and closed, and all machinery can be started and stopped. The bridge, control room and main switch board are not included. It is possible to block the mimic system interaction and still make it visible.

- Option 3: Combination of both is possible (group number B)
  Option 3 is a combination of both, the student can decide in which system he/she wants to start the different engine room systems, this can be used when a limited timetable is available. Some systems can be set fairly quickly in the mimic system and the rest in the 3D.

4.2 The experiment

During the experiment options 1 and 3 are used. The control group only used the mimic system of option 1. The second group used option 3 and could use both systems. The students were not told that they were part of an experiment. All exercises were exactly equal for both groups and the time available on the simulator was the same. Furthermore, the same stations were used as well as the same instructor and exercise instructions.

After the last exercise the students filled in a questionnaire. The reason behind the questionnaire is not told. It consists of the following items:

- General questions
  General questions are students' names and classes to find out what group and class he/she is in. This is necessary to find out in which group the student takes the test. The name is asked when more information is required.

- Practical questions
  The test consists of 10 questions about the machinery in the engine room. All questions are short questions with a photo and the student has to recognize the machinery on the photo and must choose from 4 answers.

First, the total group is analyzed. a total of 39 students have participated in this experiment and the average score of the total group is 6.38 out of 10 questions with a standard deviation of 1.914 with a high score of 9 and a low score of 3.

controle

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	3	4	10.3	10.3	10.3
	4	3	7.7	7.7	17.9
	5	5	12.8	12.8	30.8
	6	9	23.1	23.1	53.8
	7	5	12.8	12.8	66.7
	8	6	15.4	15.4	82.1
	9	7	17.9	17.9	100.0
	Total	39	100.0	100.0	

Statistics

controle

N	Valid	39
	Missing	0
Mean		6.38
Median		6.00
Std. Deviation		1.914
Minimum		3
Maximum		9

Fig.3: Total score of both classes, mean, minimum and maximum and score per question

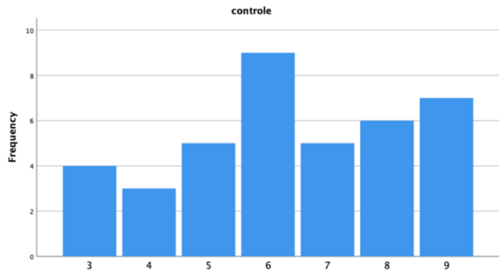


Fig.4: Bar graph, total score and number of correct questions

The lowest score was 3 correct questions with a percentage of 10.3 and the highest score of 9 correct answers with a score of 17.9%. The highest percentage was 53.8% with 6 correct questions.

If we look at the different groups, control group A and test group B the following results can be seen:

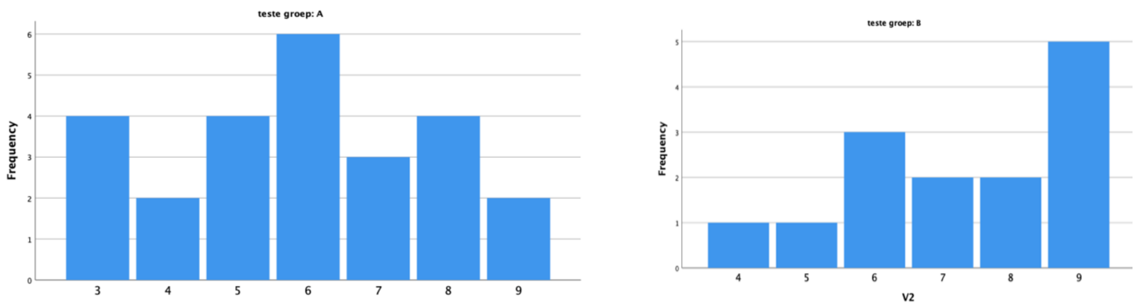


Fig.5: Control group A (left) and control group B (right)

teste groep			Frequency	Percent	Valid Percent	Cumulative Percent
A	Valid	3	4	16.0	16.0	16.0
		4	2	8.0	8.0	24.0
		5	4	16.0	16.0	40.0
		6	6	24.0	24.0	64.0
		7	3	12.0	12.0	76.0
		8	4	16.0	16.0	92.0
		9	2	8.0	8.0	100.0
		Total	25	100.0	100.0	
B	Valid	4	1	7.1	7.1	7.1
		5	1	7.1	7.1	14.3
		6	3	21.4	21.4	35.7
		7	2	14.3	14.3	50.0
		8	2	14.3	14.3	64.3
		9	5	35.7	35.7	100.0
		Total	14	100.0	100.0	

Fig.6: Test group 2

It is also interesting to distinguish how the individual questions are scored. Not all questions are of the same type and have different goals. In the table below can be seen that not all questions have the same result. Question 5, 7 and 10 have the lowest score and 2 and 3 the highest score.

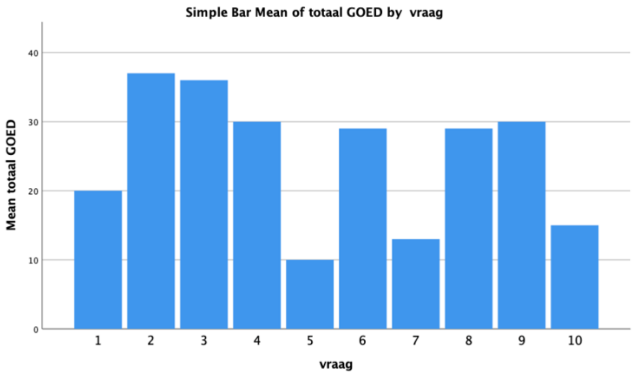


Fig.7: Correct answers total group

Now both groups are compared so the difference between the groups can be compared, Fig.8.

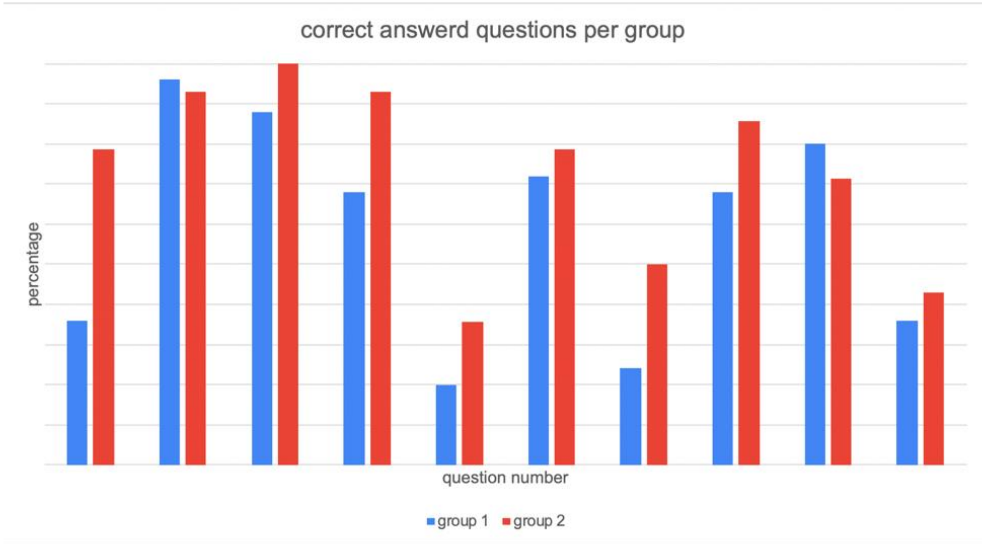


Fig.8: Percentage correct per group

When both groups are compared, per question and the percentage of correct answers, it stands out that the percentage of the second group is better in almost all questions, except question number 2 and 9. Still, questions 5,7 and 10 have the lowest score for both groups. To find out why these questions have the lowest scores, we have a closer look at the questions, Fig.9:

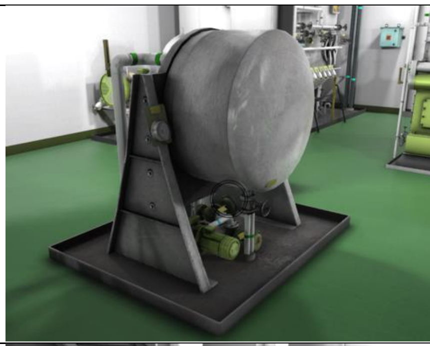
- Question number 5 shows a freshwater generator and this was not part of the training; so there has been no interaction with the freshwater generator. Students with the VR could have observed it during interaction with the VR and during the familiarization.

- Question number 7 is not only recognition, but also students need to have some knowledge on the layout and the operation of the system.

- Question number 10 not only needs recognition, but also knowledge and operation as well as insight in mimic layout and real machinery. Furthermore, a difference between other questions is the use of a picture of a real engine, and no VR animation.

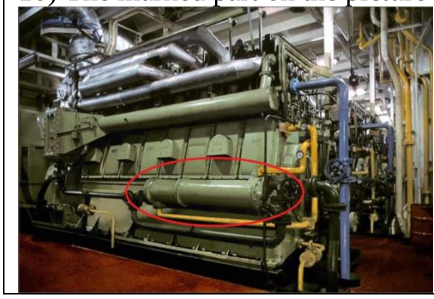
Now we compare the total result of both groups with an independent - Sample T Test. This is done to compare if the mean of both groups is the same. A sig-value less than 0.05 in difference in means would confirm this hypothesis. The test result, Fig.10 indicates that this hypothesis is rejected, and the Group B group is significant higher than group A

5) What is this?



7) Which valves should be opened to flow fresh water through the cooler?

10) The marked part on the picture is which number on the mimic?



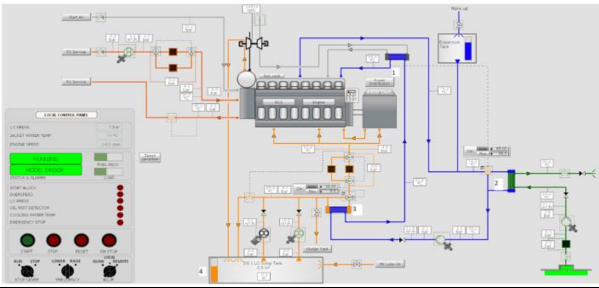


Fig.9: Pictures and questions

Group Statistics

	teste groep	N	Mean	Std. Deviation	Std. Error Mean
V2	A	25	5.88	1.878	.376
	B	14	7.29	1.684	.450

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means					95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
V2	Equal variances assumed	.052	.821	-2.324	37	.026	-1.406	.605	-2.631	-.180
	Equal variances not assumed			-2.398	29.631	.023	-1.406	.586	-2.603	-.208

Fig.10: Independent - Sample T Test

5. Conclusion

Since many years virtual reality is available on the market in very different forms. The MSTC Training Centre uses VR in the full mission engine room simulator where many students have used the VR in a full mission environment (control room, and engine room in 3D) in small groups of 4 persons. The experience of this VR application is very good. Recently goggles or screens could be added to the part task simulators in a classroom setting. At first, a decision had to be made about using either VR glasses or large monitors. After looking at all the pros and cons large 55-inch screens were added to the part task simulators.

A semi-experiment was setup with two groups of students; a control group and a test group (using the VR application). After both groups completing of the same exercises 10 knowledge questions were asked. With the the following results: the mean of the total group was 6.38, The mean of control group A was 5.88 out of 10 and the mean of test group B was 7.29 out of 10. This is almost 1.44 points higher than group A. When focusing on the individual questions the trend on both groups is the same, nonetheless some questions have in both groups a lower score than average. It shows that more knowledge-based questions are probably more difficult for all first-year students. Concluding from the results of these tests, it can be said that training with 3D improves the insight in knowledge and layout of the engine room and its machinery.

With a relative low investment (a computer and 55-inch screen for VR visualization) training and knowledge can be improved on the part task simulator of first-year students.

References

BERTRAM, V; PLOWMAN, T. (2018), *A Hitchhiker's Guide to the Galaxy of Maritime e-Learning*, COMPIT Conf., pp.7-21, http://data.hiper-conf.info/compit2018_pavone.pdf

DNVGL (2017), *Maritime Simulator systems*, DNV GL, Hovik, https://rules.dnvgl.com/docs/pdf/DNVGL/ST/2017-03/DNVGL-ST-0033.pdf

HU, L.; LIU, Z.Y.; TAN, J.R. (2018), *A VR Simulation Framework Integrated with Multisource CAE Analysis Data for Mechanical Equipment Working Process*, Computers in Industry 97, pp.85–96, https://www.sciencedirect.com/science/article/abs/pii/S0166361518300289

UITTERHOEVE, W.; WERNER, P. (2018), *The Dutch Perspective on the Use of Simulators and Sea Time Reduction in Maritime Education and Training*, ICERS 14, pp.4

YU, M.L.; ZHOU, R.G.; WANG, H.W.; ZHAO, W.H. (2019), *An Evaluation for VR Glasses System User Experience: The Influence Factors of Interactive Operation and Motion Sickness*, Applied Ergonomics 74, pp.206-213, https://www.sciencedirect.com/science/article/abs/pii/S0003687018302850

How to Achieve Smart Ship Design through the Industry 4.0

Rodrigo Perez Fernandez, SENER, Madrid/Spain, rodrigo.fernandez@sener.es
Jesus A. Muñoz Herrero, SENER, Madrid/Spain, jesus.munoz@sener.es

Abstract

In this paper, the challenges of the Industry 4.0 in shipbuilding are discussed and a solution from the CAD system is presented focused on the Smart Ship concept. Future Smart Ships must be connected, or they will not be smart. The connection of smart devices within a Smart Ship must be human controlled. The control should start from the design tools because they control the shipbuilding process from the early stages of the design up to the final production. The set of design tools, product lifecycle management and device must be interconnected among them and will be the platform for the smart ships connected to the Industry 4.0. The information shared in the scope of the Industry 4.0 must be managed by the human along the whole lifecycle of the ship, starting from the beginning of the initial design. This requires CAD tools to be prepared with specific characteristics to handle that information. This new ecosystem, opened, incorporating the new trend of technology but adapted to the specific environment of shipbuilding, will be the Smart Ship.

1. Introduction

Industry 4.0 has favoured the expansion of many technologies where application boundaries are very diffuse. Although some technologies may have very specific applications, they must be implemented as a whole when applied to Computer Aided Design, Manufacturing & Engineering (CAD/CAM/CAE) System (from now on referred just as CAD).

The CAD tool stands at the beginning of the design, but it manages many data that must be considered in advance for the further stages of the product lifecycle. Augmented Reality, Virtual Reality and Mixed Reality are closely related to the Digital Twin and interlaced with Big Data, which are generated by CAD tools and all surrounding solutions, which applies some cloud/edge/fog computing to these data in a merged technology between finite-state machines and Artificial Intelligence (AI) cognitive processes, *Muñoz and Perez (2017).*

To perform all these integrations in an agile manner requires a network which support different connections to add specific devices, i.e. Internet of Things (IoT), which can access to the data, creating and modifying them, in a different layer which affects to the basic information layer created by the CAD System in the shipyard. This network should be secure, but also open to allow distributed work, which must be tracked such that all design or process modifications are recorded an open, transparent, trusted and non-modifiable working method for all stakeholders, like: shipyard, engineering offices, classification society and ship owner.

Results of the design should be easily integrated with future manufacturing methods like 3D printing, generating printing orders directly from the CAD model.

Shipbuilding phases involve design and manufacturing, but an integrated Industry 4.0 CAD System should also be involved in operation and maintenance phases, i.e. it must cover all the vessel lifecycle end to end, from design to decommission.

When a ship comes for a repair or retrofit, sometimes a suitable CAD model is not available, *Muñoz et al. (2018)*. Trying to recreate a CAD model for e.g. an entire engine room from a scanned model can be a nightmare. Smart CAD tools can be used for processing the point clouds from scan to create CAD(-ready) geometries, which can be converted with increasingly less manual work, as intelligent CAD systems may learn from previous similar applications.

This briefly summarizes how Industry 4.0 technologies may be applied to a CAD System, whether through direct integration or in connected periphery applications.

2. What is the Industry 4.0

To be able to understand how Industry 4.0 became today's catchword, a look at its predecessors might give us a perspective on how this revolution is different:

- **The First Industrial Revolution** started in the United Kingdom introducing machines into production (1760-1840). This included the use of steam-powered engines and water as a source of power. This helped agriculture greatly and the term "factory" became widely used. The textile industry benefited particularly from the changes, and it was among the first to adopt such methods. It also constituted a huge part of the British economy at the time.
- **The Second Industrial Revolution** dates between 1870 and 1914 (although some of its characteristics date back to the 1850s) and introduced pre-existing systems such as telegraphs and railroads into industries. Perhaps the defining characteristic of that period was the introduction of mass production as a primary means to production in general. The electrification of factories contributed hugely to production rates. The mass production of steel helped introduce railways into the system, which in turn contributed to mass production. Innovations in chemistry, e.g. synthetic dye, also made major contributions. World War I put an end to this era. Mass production, of course, was not put to an end, but only developments within the same context were made and none of which can be called industrial revolutions.
- **The Third Industrial Revolution** dates between 1950 and 1970. It is often referred to as the Digital Revolution, and came about the change from analogical and mechanical systems to digital ones. Others call it the Information Age. The third revolution was, and still is, a direct result of the huge development in information and communication technology.
- **The Fourth Industrial Revolution** takes the automation of manufacturing processes to a new level by introducing customized and flexible mass production technologies. Machines will operate independently, or cooperate with humans in creating a customer-oriented production field that constantly works on maintaining itself. The machine becomes a rather independent entity that is able to collect data, analyse it, and advise upon it. This becomes possible by introducing self-optimization, self-cognition, and self-customization into the industry. Manufacturers and operators will communicate with computers rather than operate them.

Industry 4.0 can be considered as the set of technologies that applied to industry in general will allow the connection of machines, products and services to gather and analyse data enabling flexible, faster and efficient processes to produce cheaper and with high quality. The most relevant technologies considered in this scope are IoT, cloud computing, cybersecurity, augmented and virtual reality, additive manufacturing, autonomous robots, simulation and blockchain.

Smart Ships will embody the most disruptive technologies in this 4.0 revolution. However, what is a Smart Ship? It is a ship that integrates data from a wide variety of sources in order to contribute to, and improve, the operational efficiency and safety of the vessel. As we are running out of conventional approaches to make the ships more efficient, we should tap into the potential of Industry 4.0 to obtain quantum leaps in improving performance.

3. Adaptation of the maritime industry to the new technological challenges

"The platform is the key to success. The 'things' will get increasingly inexpensive, applications will multiply, and connectivity will cost pennies. The real value will be created in the horizontal platform that ties it all together, i.e. the new OS," *Taylor (2016)*.

Porter et al. (2015) described the new technology stack: "Smart, connected products require companies to build and support an entirely new technology infrastructure. This 'technology stack' is made up of multiple layers, including new product hardware, embedded software, connectivity, a product cloud consisting of software running on remote servers, a suite of security tools, a gateway for external information sources, and integration with enterprise business systems.” If a company wants to enter in this business, it must have products that connect with the products of other companies. So, the necessity for open products is increasing.

From our point of view, the mentioned platform it is not exactly an OS, but a set of interconnected applications, which share the same data about the product. Data that have been appearing since the early design phases and feed that single database, not only with the design attributes but also with the desirable values for optimal performance. Around this database, the applications will appear to control the life cycle of the components, the sensors that collect the information of operation or operational life of the same objects and devices.

The key is that all devices, sensors, components are not connected directly to the IoT, but connected to a management core that controls the data to interchange to IoT for those devices allowed to enter in that network. This allows exploiting that all the existing components in the ship have their respective model created, incorporated, and defined from the CAD system. We could say that CAD is the birthplace of the elements of the ship. Therefore, it is the CAD system that knows them and is best suited to assign their mission and level of participation in the world of IoT.

Future smart ships must be connected to IoT or they will not be smart. The connection of smart devices within a smart ship must be human controlled. The control should start from the design tools because they control the shipbuilding process from the early stages of the design up to the final production. The set of design tools, product lifecycle management and device must be inter-connected among them and will be the platform for the smart ships connected to IoT. The information shared in the scope of the IoT must be managed by the human along the whole lifecycle of the ship, starting from the beginning of the initial design. This need requires the CAD tools to be prepared with specific characteristics to handle that information. This new open ecosystem, incorporating the new trend of technology but adapted to the specific environment of shipbuilding, will be the Internet of Ships.

4. Digital Twin

Everybody speaks about the ‘digital transformation’ but there is little consensus of what it really means. Clarification may come from describing the three main pillars this transformation is based on:

- **Digital services**. The digital services include the generation of a digital twin of the ship, with all the relevant information. This mock-up is going to be of high value, not only for the Classification Societies, but also for the shipyards during the phase of construction, and for the ship-owners along the exploitation phase.
- **Customer experience**. Customer experience can be improved e.g. through digital certification and with a specific set of applications that will facilitate the tasks of inspection and maintenance accordingly with the Classification Society Rules.
- **Operations.** In this phase, effectiveness and efficiency of the interventions on board ships can be improved, as well as procedures of service provider such as Classification Societies.

Cybersecurity poses already a significant problem for ship operation. With increasing autonomy, smart on-board systems, and resulting data exchange, the problem is amplified. Consequently, both the digital twin and its fundamental derived benefit, the digital services, must consider aspects of cybersecurity and robustness of systems.

Inside the concept of cybersecurity, Classification Societies are defining new classification notations, such is the SYS-COM from the Bureau Veritas, *Pancorbo (2017)*, for ships with increased reliability of communication services ship and shore, supporting remote monitoring etc., Fig.1, ultimately enhancing safety and efficiency through IT.

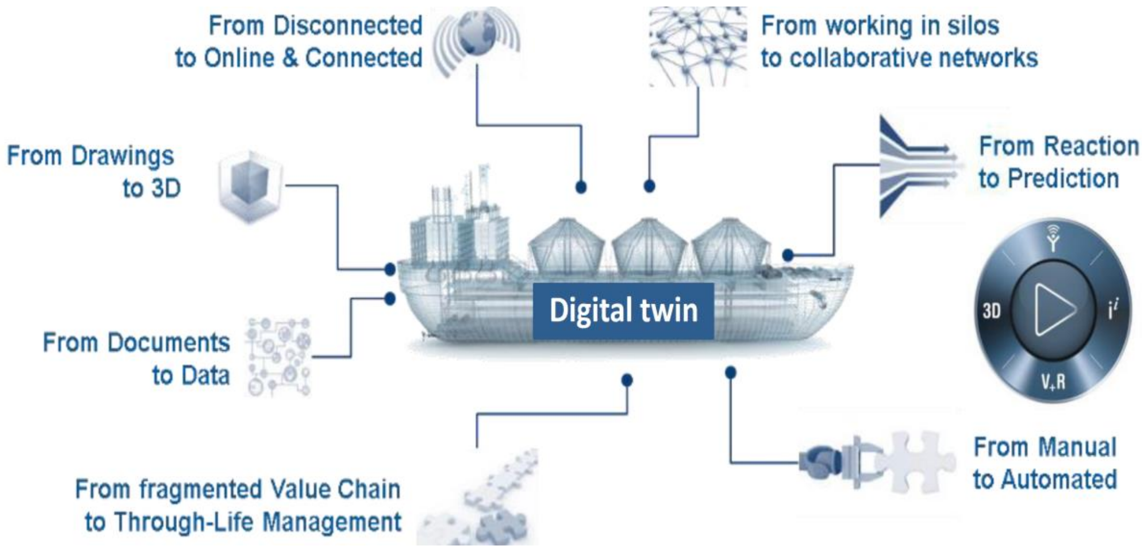


Fig.1: Applying the digital transformation into the Asset Integrity Management, *Pancorbo (2017)*

Other tools that had already been in use are those oriented to help to the customer experience of the ship-owner:

- **Mobile applications** can track the ship, handle its certificates, Class inspections, etc.
- **Specific cargo tools** verify that our ship is approved to transport a particular cargo.
- **E-Commerce** started as a web tool where ship owners can request the certification of ship drawings, obtaining an on-line price for it, pay for it electronically, all at the same time. It is specially oriented to improve ship owner experience, above all by taking reducing the time to get Class Certification.

5. Virtual Reality and Augmented Reality as part of Smart Ship design

Shipbuilding industry faces a global and highly competitive market scenario. Innovation (and particularly adapting and implementing innovation) is the key to success here. For example, Virtual Reality is not new. But its practical application in shipbuilding has spread now thanks to significant improvements in software and hardware. From a customer/designer point of view, current VR solutions contribute clearly to productivity and costs. And behind the VR application is the 3D CAD model.

Because of its many applications, consistency of the VR model is vital. CAD suppliers develop smart tools to use this model in realistic navigations in VR environments with multiple rationals, *Alonso et al. (2012)*. Some of the most important practices of VR in shipbuilding are discussed in the following paragraphs.

These tools are mainly used in the engineering department of a shipyard and the technical office. Sometimes also the production department employs VR tools, as design errors discovered in production are most costly. Designers are not yet regularly working with VR. However, it is very useful to check the model at the design stage, to avoid inconsistencies/errors but also to improve ease of assembly.

The first and maybe foremost benefit of Virtual Reality is for avoiding spotting design errors. Viewing the ship 3D model as realistic as possible has been proven to be very effective in spotting errors. From early design stages, to manufacturing and production phases, one can check all the elements, to see inconsistencies, interferences and collisions, to query about properties and attributes and to study design alternatives dynamically. User and VR model interaction has become much more intuitive, e.g. with tracking devices which give the sense of being inside the model, walking and moving in it, touching it.

VR viewers allow fast and easy navigation. Different modes of visualization, search and query options, inclusion of annotations and measurement of distances are useful features. Ergonomic aspects in the design can be studied and checked better having a Virtual Reality solution, with the possibility to include dummies.

Simulation is another important area of application of Virtual Reality in ship design, engineering and production. It is used for many different purposes, from the study of escape routes to the simulation of dismantling for maintenance of equipment items. The study of critical assembly tasks and how the surrounding elements are affected is helpful, too. Virtual Reality is very useful for the management, just to visualize the progress of the project. Once it is being built, the evaluation between the smart model and the real ship is essential and Virtual Reality can help to find issues to be solved as soon as possible avoiding further unnecessary costs.

On the other hand, it is common that ship design, engineering, and manufacturing are divided in different blocks that are subcontracted to different technical offices, contractors and shipyards. The shipyard collects all the information from workshops and the vessel itself, and needs to supervise that everything is correct in the design process. Virtual Reality allows the integration of all the data in an easy way, in a single model, which is much more effective than having many different files or bunch of data.

The benefits of Virtual Reality for commercial and marketing purposes are evident. This explains Virtual Reality has enjoyed ready reception in sales and promotion activities. Thanks to the extensive range of VR tools in the market, from minor and portable solutions to large and on-demand solutions, the present and future applications of Virtual and Augmented Reality are higher than we could imagine a few years ago. From a marketing point of view, the possibility to present a smart ship 3D model with high level of detail and resolution in an immersive Virtual Reality experience gives great value to sales activity. Many shipyards are already employing this technology, not only in their own facilities, but also in exhibitions, shows and fairs thanks to the portable solutions.

6. Smart Ships and connectivity - The concept of Internet of Ships

CAD tools are at the beginning of the product life cycle, but they also have a strong relationship with the production cycle, providing the information needed for construction in all aspects. Those CAD systems with a compact and homogeneous database can extend their contribution to both the life cycle management tools and the IoT connectivity management application of the ship's elements.

To enable all of this, CAD applications must have certain characteristics. CAD systems have to evolve to become global solutions. In our vision, the CAD system will not only be the heart of the design but also the vehicle of communication between the products, their manufacturers and their operation. Having this idea in mind, we wonder: what should be those characteristics that turn a CAD system into a global design solution? We think the answer is the integration of systems. It is impossible to grow enough quickly just with organic grew. Therefore, the companies must shape the future of their products thinking in terms of integration. The question then is how to get getting this integration implemented. To answer this question, it is necessary to think about the future and to identify the technological trends in this area.

We believe that connectivity, artificial intelligence and virtual reality are the keys to designing the future in the next ten years. We also think about the application of these trends in the maritime industries. Any future technology for software solutions, the world of IoT, Smart Ships and likely other technological developments of the products will be tied to these three trends. They are not disjointed roads; they are trends that converge and support each other.

IoT is simple: it is about connecting devices over the internet, letting them talk to us, applications, and each other. But IoT is more than smart homes and connected appliances. It scales up to include smart cities – think of connected traffic signals that monitor utility use, or smart bins that signal when they need to be emptied – and industry, with connected sensors for everything from tracking parts to monitoring crops.

But is the maritime sector ready for this revolution. Is it possible that this traditional and conservative sector moves into this technology? There is already evidence that the shipbuilding industry is no stranger to these developments and is already connecting some ship equipment to the Internet. But this is not the only field of application of the IoT for shipbuilding. Different systems in a ship shall be connected to each other in order to share information. E.g. measurements of inclination of the ship will instruct the ballast system to get the upright position of the ship. This means that ballast pumps must be connected to the centralized control of the ship in order to receive the orders.

In the initial design stage, it must be considered which systems or components of the ship need to be smart and which not. Not all devices have to form part of the IoT of a ship.

As in a smart home or smartphone, there smart ships will be equipped with a network of sensors that capture a range of voyage information, including:

- Location
- Weather
- Ocean current
- Status of on-board equipment
- Status of cargo

Ship owners can monitor a vessel's status in real time and apply analytics to current and historical data to make decisions enabling more efficient operation, saving time and fuel. Sensors and IT technologies are facilitating the introduction of new applications at sea, like energy distribution, water control and treatment, equipment monitoring in real time.

The aim is to start this technological revolution in the design and production phases in order to build efficient, safe and sustainable vessels. In a decentralized sector, like the maritime, where often the engineering and production are in different locations and where critical decisions cannot wait, the 'Internet of Ships' or connection through the network of critical components in the design / shipbuilding, starts to look like something that the sector cannot obviate.

The idea is to monitor all those parts in which early detection of events allows us to make the right decisions. In this sense, the available sensors during the early stages of construction of the ship, allow us to identify if the construction of the vessel is according to the design we have created with CAD. If we can reduce materials or use another material, if we must change anything according with naval architecture calculations …

The continuous monitoring integrated with a ship design CAD will reduce costs, avoid mistakes, and make decisions in real time from the shipyard, design offices or from remote locations. Nowadays, CAD solutions can be used in pocket tools, making it the indispensable ally in this new technological revolution. The shipbuilding process generates a lot of information and data. A priori it may seem impossible to have all this data in real time, but the new processors, simpler and smaller, with a good connection to the Internet, make it possible. The data management is, however, only one aspect of the the Internet of Ships. Another aspect is energy efficiency.

But Internet of Ships not only covers the stages of design or production. The different equipment items to be installed in the ship will have to be prepared for connections between them in order to provide the necessary information that the components must to share for the normal operation. Once the sensors are in the components where we want to monitor information, we can obtain information throughout the life of the ship. Internet of Ships is presented as a solution capable of detecting when a component on a ship is close to fail and must be replaced, when we take the ship to repair, when we have to paint again, when corrosion has reached a certain limit, etc. And all this from a smartphone and in time to avoid unforeseen failures. Internet of Ships reaches this sector to ensure profitable production and safe, efficient and sustainable processes for all types of ships (fishing vessels, tugs, tankers, cargo vessels, ferries, dredgers, etc.).

There should be two different networks of connections. An internal network where the systems and components of the ship work together to get the best operation possible. Different smart devices, components and systems will share their states and needs for a better operation. Another connection is needed to the external world, providing the link from the smart ship to the centralized intelligence of the ship owner. This connection will send and receive information necessary for the better operation of the ship. Information necessary to for repair or replacement of components will be transmitted from the components themselves. To enable this approach, the CAD tools must be able to manage the amount of data generated, know the relationship between the components and the recipients of the information shared.

7. Conclusions

Now, you may have an idea of the scope and complexity of the changes that Industry 4.0 demands. In most cases, it is an unavoidable step that must be taken to remain competitive. Parts of the necessary transformation may appear at present too risky and complex step; some technologies are in early stages for industrial implementation, but it is important to be aware of them, monitor them and consider them early and design software systems to be ready for their integration at a later stage.

With 5G, it is expected that IoT will grow in exponentially unfolding its true potential, which will also give new momentum to the Digital Twin.

There is another important piece of the puzzle: the security, veracity and trust in the data transactions between all the different participants. Cloud technology and cybersecurity are two pieces of the game and without them, the envisioned ecosystem will never work. Data transactions should be guaranteed, and all process steps should be tracked employing blockchain technology. This data centric model will be complemented with Big Data analytics and AI:

- Extracting conclusions from that big amount of data acquired from IoT devices.
- Enabling better predictive maintenance thanks to the capability to predict failures or detect problems early.
- Using expert systems analysing current activities of human experts in order to derive best business practice for whole companies or industry segments.

At the centre of this ecosystem lies a 3D model, created mainly with CAD tools. The importance of CAD will rather increase than diminish. Digital twins link the 3D CAD model and the rest of Industry 4.0, the real world and the virtual world, both living concurrently during the evolution of the construction and extending this world of possibilities to the entire lifecycle of the product.

CAD tools, as an important part of this environment, should evolve to be easily linked to these technologies. It is also important that the future CAD systems are adapted to a new generation of users that demands a different interfaces and workflows. This will require an important effort.

This big world is now open. Shipbuilding is starting in it, but the potential is clear and to remain competitive, it is mandatory to study how these technologies can improve processes, resources, workflow, cooperation between stakeholders. The tools are here, now it is necessary to study and analyse our particular maritime case. There is no global solution for all industries, not even a global solution for a particular industry, but digitalization is on hand and it is mature enough to start implementation and be part of our strategy for the future.

Future Smart Ships or Smart Shipyards must be connected to IoT, or they will not be smart. Probably the cheapest option to achieve it is using RFID tags, https://en.wikipedia.org/wiki/Radio-frequency_identification. They are an affordable solution for warehouse tracking systems, helping shipyards with material traceability and lowering material collection times for MBOM (manufacturing bill of material) orders. This kind of tags can be used not only in warehouses to control stock, but also in workshops and on board, allowing exploiting economies of scale for RFID readers and writers.

If software systems fail, because these tags are out of the system, a manual reader can help controlling the storage easily. This will support robust production lines. Once the software problem is solved, the SCS (self-contained systems) stock can be updated without any extra effort.

On a final note, the envisioned eco system comes at some entry barriers:

- Initial study of the requirements
- Initial investment
- Selection of software packages which fulfil all requirements of efficiency for each group (CAD designers, production operators, purchase people...) and, at the same time, is fully integrated with other parts of this complex system.

References

ALONSO, V.; PÉREZ, R.; REIDAR, T.; SANCHEZ, L. (2012), *Advantages of using a Virtual Reality Tool in Shipbuilding*, 11th COMPIT, Liege

MUÑOZ, J.A.; PÉREZ, R. (2017), *CAD tools for designing smart ships in the world of the Internet of Things (IoT)*, SMART SHIP 2017, London

MUÑOZ, J.A.; PÉREZ, R.; GUTIERREZ, J.R. (2018), *Design Rules Evaluation through technologies of treatment of Big Data*, 17th COMPIT, Pavone

PANCORBO, J. (2017), *Bureau Veritas: Evolution towards a digital industry*, Revista Ingeniería Naval, pp.75-80, June

PORTER, M.E.; HEPPELMANN, J.E. (2015), *How smart, connected products are transforming companies*, Harvard Business Review, October

TAYLOR, S. (2016), *10 Predictions for the Future of IoT,* Big Data Quarterly, March

Ship Hull Optimization – An Attempt to Determine Position and Course of the State of the Art

Volker Bertram, DNV GL, Hamburg/Germany, volker.bertram@dnvgl.com
Emilio F. Campana, CNR, Rome/Italy, emiliofortunato.campana@cnr.it

Abstract

The paper surveys the state of the art of ship hull optimization and trends to extrapolate to likely future applications. It looks both at frontier industry applications and recent R&D papers. The paper looks at individual elements of hull optimization, such as geometry model (level of detail), hydrodynamic model, optimization algorithms, and objective functions (and constraints). For all elements, a gap analysis shows room for improvement or further R&D efforts.

1. Introduction

Increasing fuel prices and IMO regulations to curb CO_2 emissions put pressure on ship owners to obtain more fuel-efficient ships. Various publications, e.g. the 2009 IMO Greenhouse Gas study, *Buhaug et al. (2009)*, have identified ship hull optimization as one of the "low-hanging fruits" for significant fuel savings. The changing economic and regulatory framework have contributed to the recent success of ship hull optimization. So have changing technical possibilities, such as the evolution of high-fidelity CFD (computational fluid dynamics) and affordable high-performance computing.

Hochkirch and Bertram (2012) describe the historical development of ship hull optimization to the state of the art. *Nowacki (2003)* and *Birk (2003)* give more fundamental overviews of optimization methods, *Bertram (2003)* and *Harries (1998)* on applications for ship design. Our focus in the present paper lies on the most recent state of the art in industry application, looking at unharvested fields, not yet exploited technical possibilities as evidenced in research applications. We will show that more progress is to be expected in virtually all elements of ship hull optimization, but most notably on more realistic objective functions and constraints.

2. Elements of optimization

2.1. Geometry model

The state of the art in ship hull optimization in industry applications typically considers the bare hull only, often focusing on the bulbous bow only, sometimes on the stern only and sometimes on the whole ship hull. Rudder, propeller and possibly propulsion improving devices (PIDs) are designed separately. In CFD analyses, complete models of hull, propeller, rudder and as applicable PIDs are more frequently found, then often with a simplified propeller model, where body forces mimic the effect of the propeller but its geometry is not explicitly modelled. The trend is generally towards more complete models, *Peric and Bertram (2011)*, in the CFD models, but the variation in geometry is so far still limited to the hull.

Asymmetric sterns can be seen as a special case of a PID. Here the geometry model and the CFD model release the usual assumption of symmetric hulls. The larger design freedom converts into additional savings in calm-water fuel consumption for realistic operational profiles of 1-5%, depending on initial design, propeller loading, ship type and optimization model, *Hochkirch (2017)*, *Van der Ploeg and Schuiling (2018)*.

The level of detail of CFD models generally omits bilge keels, welds and other minor geometries. It also considers generally a standard roughness, and looks at the geometry as designed, which always differs (somewhat) from the as-built geometry and the in-service roughness.

For optimization, special geometry models are used to keep the numbers of evaluated geometries relatively low. Such "parametric models" describe families of hull surfaces using control parameters, leading to much fewer free parameters than e.g. direct modification of hull surface points, Fig.1. "Parametric modeling for shape optimization can be subdivided into fully-parametric and partially-parametric modeling. In fully-parametric modeling the entire shape is defined and realized by means of parameters while in partially-parametric modeling only the changes to an existing shape are described parametrically," *Harries et al. (2015)*. See also *Harries (1998,2014)* for a more detailed discussion of parametric hull modelling. Alternatively to using a parametric model for the hull, moderate changes from a base design can be described by parametric models, *Peri et al. (2001)*. This reduces the effort, but also the design freedom.

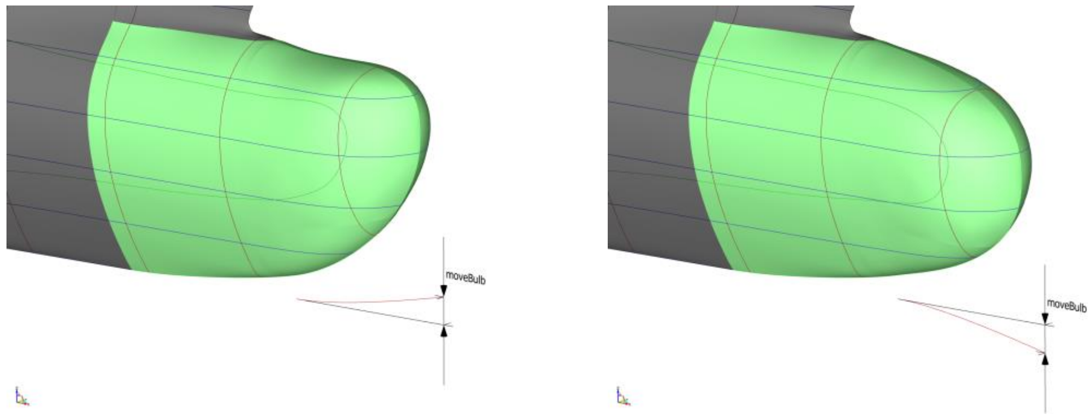


Fig.1: Partially parametric model for bulbous bow optimization varying one parameter, *Harries et al. (2015)*

So-called adjoint methods calculate the effect of a shift of a hull surface perpendicular to the surface, *Stück (2012)*. They indicate quickly where a hull change will most affect the objective function, but are not popular in industry, as the amount of shifting for an optimum solution is not given and direct morphing of ship hulls following patterns from adjoint methods generally leads to hull shapes that violate design-for-production principles. Parametric hull models, on the contrary, are set up such that the resulting families of hulls are "reasonable" and generally don't cause issues with production.

2.2. Hydrodynamic model

The prediction of resistance or power requires a hydrodynamic model. More accurate "high-fidelity" models come at the price of large computational effort, and vice versa fast "low-fidelity" models have moderate to high errors.

For ship bows, the wave making, inviscid part of the resistance dominates. Hence, potential-flow models can be justified. State of the art in industry still seems to rely on such potential-flow models, typically so-called fully nonlinear wave resistance codes, as described in *Bertram (2011)*.

For ship sterns, viscous effects dominate, requiring RANSE (Reynolds-averaged Navier-Stokes equations) solvers which involve a much higher computational effort. If the ship speed is relatively low (e.g. for a tanker), wave making may be neglected and simpler double-model models may be employed to optimize a stern.

If the propeller is geometrically modelled, the computational effort increases typically by one order of magnitude. Industry application thus generally use a body-force propeller model to optimize the hull, which appears to be a sensible trade-off between accuracy and computational effort, unless the propeller itself or a nearby PID shall be optimized, Fig.2.

But in many cases, the whole ship hull is to be optimized and omitting either wave-making or viscosity introduces significant errors. The trend is towards complete hull models and using RANSE solvers.

Current state-of-the-art CFD computations have sufficient resolution and sufficiently accurate free-surface and turbulence to obtain good agreement with model-test validation cases, *Peric and Bertram (2011)*. There is wide consensus that the same CFD approaches will predict full-scale resistance and power at least as accurate as traditional model-test extrapolation.

For computational efficiency, often a first round of selection with simplified models reduces the search space, and a second round with high-fidelity models then limits the variations to balance overall computational effort and achieved gains.

RANSE solvers find solutions iteratively, where pressure-velocity coupling, free-surface deformation and other quantities need to be updated until all equations are fulfilled. Computational effort can be reduced by starting with a good approximation of the final solution. So far, RANSE solutions in optimization applications seem to start always from scratch for each new candidate design. Re-using solutions from similar design candidates as starting point could reduce computational effort, but this is subject to research.

Industry is interested mainly in yearly fuel costs, which is linked to different operational scenarios (variations of draft and speed) and the full-scale ship in self-propulsion conditions. We expect that optimization projects will then move from model-scale simulations of resistance to full-scale simulations in self-propulsion condition in the future. The scale effects are important not only for the boundary layer, but also for wave breaking, *Hochkirch and Mallol (2013)*.

The main hurdle for better hydrodynamic models in practical optimization projects is affordable high-performance computing power. This will be discussed further below.

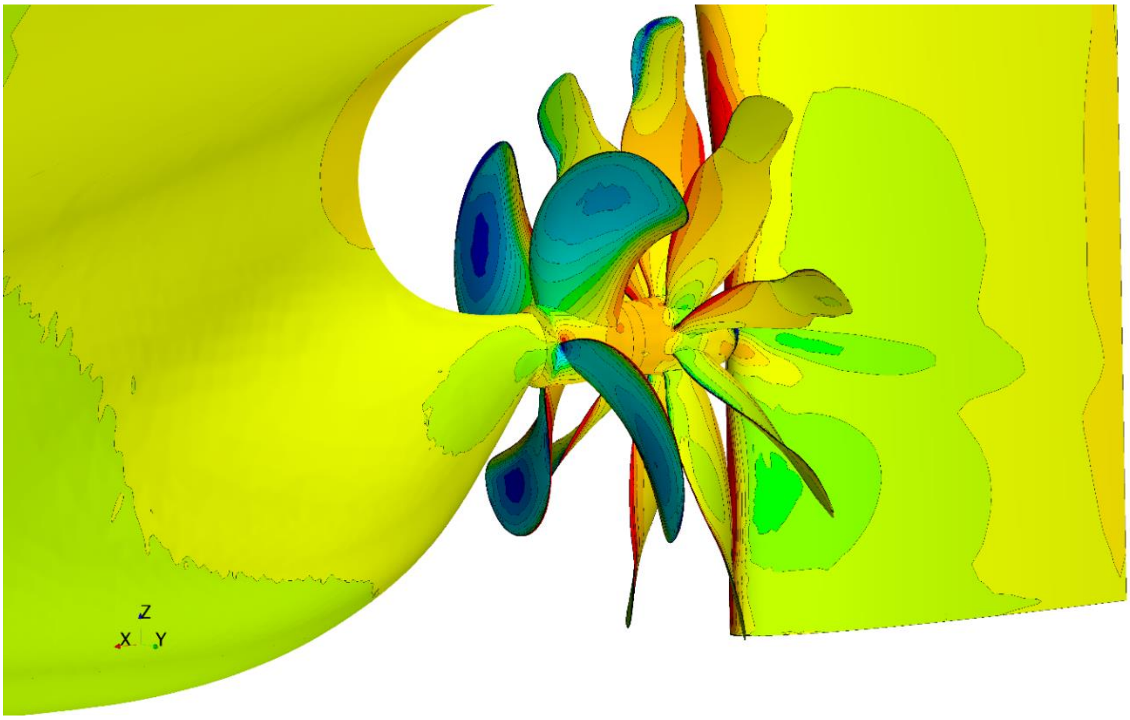


Fig.2: Trend towards "complete" models with hull, propeller, rudder, PID at full scale with unsteady free-surface RANSE solvers

2.3. Optimization strategies

Optimization methods per se are a domain for mathematicians and computer scientists. There is no shortage of optimization methods and theoretical advantages and disadvantages, see e.g. *Birk (2003), Peri and Campana (2005)*. Some methods work better for certain problems, some better for other problems. In general, experts in the mathematics of optimization have little understanding of ship design and ship designers have little understanding of the mathematics of optimization. Hence, ship designers generally opt for commercial or public-domain optimization packages, may try a few of the available options in such a package and stick with the option that worked best for one project until they see a reason to try something different.

For industry projects, two commercial packages seem to dominate:

- CAESES (formerly Friendship Framework), www.caeses.com
  Applications for ship hull optimization are found e.g. in Birk and Harries (2000), Harries (2003), Abt and Harries (2007).
- modeFRONTIER, www.esteco.com/modefrontier
  Applications for ship hull optimization are found e.g. in *Maisonneuve (2002), Giassi et al. (2003)*

These frameworks offer user-friendly graphical user interfaces and suites of optimization algorithms. The frameworks can be used for Design of Experiments (DoE) (a.k.a. concept exploration) and for formal optimization. The combination of both is dubbed multi-criteria optimization, where DoE is first used for an understanding of the design space and a requirements elucidation with the customer, and then the single objective (often an artificial weighted average of various directly meaningful objectives) and the constraints are formulated and the formal optimization can proceed.

Evolutionary optimization algorithms, such as Genetic Algorithms, are notoriously inefficient in single-processor computing. However, they are very easy to parallelize and robustly find global optima. With the advent of widely available parallel processing, evolutionary algorithms have become the standard choice in ship hull optimization. Research continues to improve the efficiency of evolutionary algorithms, e.g. *Diez and Peri (2009).*

Meta-modelling using response surfaces are subject to research, *Alvarez (2000), Peri (2013)*. The idea is using surface fitting on computed designs and using the functions for the fitted surfaces to interpolate or extrapolate to generate virtual other design evaluations, avoiding expensive CFD evaluation. The trend from the approximated response surface can then guide the search to the optimum of the current response surface. The approach is likely to become standard business practice for future hull optimization, as the individual CFD evaluation becomes increasingly computationally expensive with the trend towards better hydrodynamic models.

2.4. Objective functions and constraints

The formulation of the right optimization objective is arguably the biggest issue in industry practice. It is often a pragmatic, rather than intellectual challenge: What do we really want:

- Minimum calm-water resistance at design draft and design speed
  This is the classic objective of first and academic applications, but this objective may lead to aftbody geometries with poor propeller efficiency and overall higher power demand and larger main engines than designs identified as “sub-optimal”. Even if the penalty is only small, the design may then perform well in model tests for the contractual design point, but poorly in off-design conditions. But most ships operate virtually all the time in off-design conditions (lower draft, lower speed).
- Minimum calm-water resistance or power at a given operational spectrum
  This is the state-of-the-art approach in industry projects, *Hochkirch and Bertram (2012), Hochkirch et al. (2013)*. The objective function considers an operational profile instead of just one design (or contract) point. A “point” in this respect is the combination of load condition and speed. Ships are frequently operated at lower speeds or partial drafts (with varying trim angles). The objective then considers for a “best” compromise between these conditions, yielding the lowest yearly fuel costs (in calm water). The propeller is generally not directly modelled. Either

the resistance is optimized, assuming unchanged propulsive performance (poor approximation of reality) or the propeller is considered in a simplified body force model (satisfactory model for hull optimization).

- Something else
  If we look at yearly fuel bill, ship-in-service condition should be considered, in varying spectra of seaways, the specific fuel oil consumption of the main engine at various power settings etc. But maybe the customer is more interested in EEDI (Energy Efficiency Design Index) certification and wants perhaps the smallest engine possible, and once that has been selected from a catalogue with discrete options, we can fix the engine power and the ship achieving contract speed at contract conditions as a constraint and re-optimize the hull for in-service fuel consumption? Setting the objective function is not easy…

Research is active on various aspects of current business practice:

- Optimization for in-service ambient conditions
  We have progressed from optimization for contract condition to optimization for realistic operational conditions. The logic continuation is considering also realistic ambient spectra, rather than adding a traditional 15% for all designs to cover added power in waves and with progressing hull roughness. *Kleinsorge et al. (2016)* argue the case for considering realistic sea state scenarios looking at likely life-time service in hull optimization. The added savings may be considerably, at least for full hull forms like tankers or bulkers: "A recent case study indicates the potential for further fuel savings if performance in waves is considered in hull design, especially for bulk carriers and tankers. Two hull variants of a Handymax bulk carrier were investigated: one with a conventional bulbous bow, one with a rather straight bow profile. The assessment considered a typical trade and a simplified operational profile with two load conditions (fully loaded and in ballast) and two speeds (service speed and slow steaming). The design with the straight bow contour reduced average fuel consumption by 3%," *Shigunov and Bertram (2014)*, Fig.3.

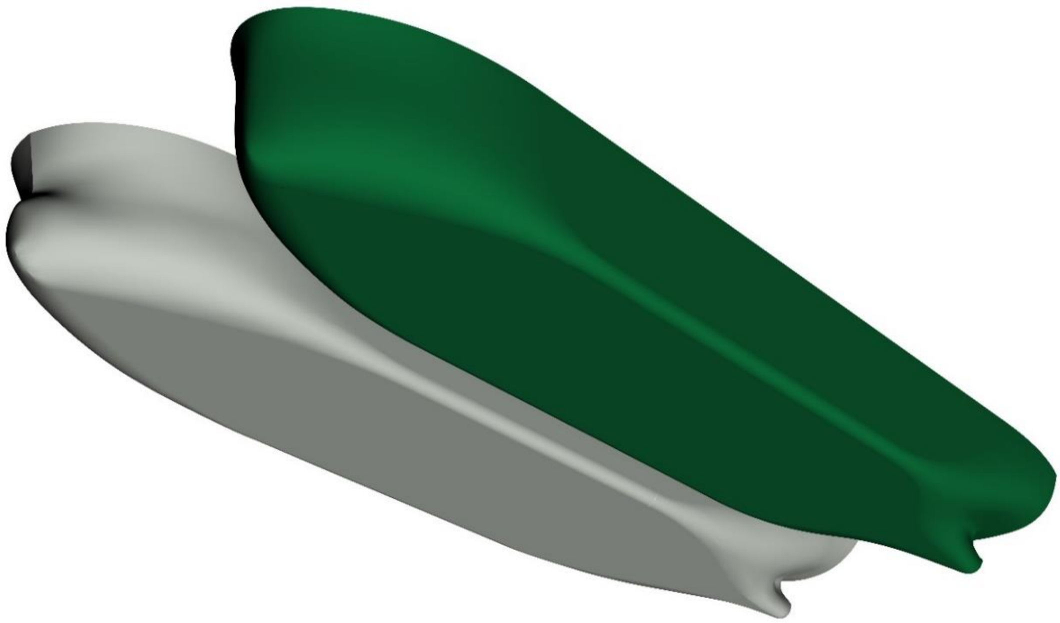


Fig.3: Classical bulbous-bow (grey) and optimized straight-bow (green) design for bulk carrier bows, *Bertram (2014)*

- Consideration of realistic in-service hull roughness
  Changes in hull roughness due to fouling may be significant during 5-year intervals between docking. The changes in roughness change both resistance and propulsive efficiency (via the

wake). Few CFD analyses for ships have considered varying hull roughness (so far for different hull coating roughness), *Demirel et al. (2014), Vargas and Shin (2017),* and none so far for ship hull optimization.

- The past is certain, but the future is not. If we optimize a hull for an operational profile, ideally we should do so for the ship in future service. We may make some educated guesses for operational speed, e.g. decreasing due to increased fuel prices, but these will by nature be approximative. Uncertainty is then unavoidable and we may want to include some measure of robustness in our optimum design, in the sense that it performs well in a range of probable scenarios. It is likely that future business practice will reflect uncertainty considerations. Work by *Diez and Peri (2010,2011),* give a glimpse where we might head.

2.5. Accessibility

Ship hull optimization has drifted over the decades from research to industry specialists. It is likely that the process is not yet finished and we will see in due time a transition from specialists to the wider design public.

CFD analyses for power predictions have become increasingly accessible for a wider public. The codes have become more user-friendly and robust, *Peric and Bertram (2011)*. Computing hardware and software licenses has become widely affordable with rent by core-hour business schemes by major software vendors, *Hildebrandt and Reyer (2015)*. Skills in creating good CFD models (grids, initial conditions, process control parameters) have been translated into macros for typical ship types, such that now anybody can perform numerical power predictions in "virtual trials" requiring only the geometry of the ship and the specification of drafts and speeds, *Hochkirch and Hahn (2017)*.

Optimization shells such as CAESES and modeFRONTIER are already user-friendly. Bringing them together with cloud computing, license rent-by-hour schemes and suitable macros for parametric hull descriptions could allow online ship hull optimization for anybody in a similar manner.

3. Conclusion

Each year, we get more computing power for our money. The added computational power allows more sophisticated models. There is no shortage of ideas how to make optimization projects computationally more expensive: better turbulence models, finer grids, more detailed operational spectra, inclusion of directly modelled turning propeller, etc.

As a heretic thought, we may ask to what extent and accuracy we should optimize ship hulls if in the end the ship owner is interested only in the as-built ship and not the as-designed ship. If sister vessels differ in sea trial measured power for same speed by 5-6%, *Krapp and Bertram (2016),* where does that leave us with typically reported 4-5% improvements in hull optimization projects? Quite likely, with progress in tracking as-built condition, e.g. accurate 3d laser scans of built ships and hull roughness measurement in drydock, we may understand differences between sister vessels better.

Panta rhei – everything flows. Ship hull optimization is likely to progress on several fronts. Combine this with changing economic frameworks (e.g. fuel price) and it appears likely that tomorrow we will re-optimize yesterday's "optimized" hulls.

Acknowledgements

In the writing of this paper, we drew on the expertise and insights of many colleagues who have contributed over the year much to our knowledge of the field of practical hull optimization. We express our gratitude to Matteo Diez, Stefan Harries, Karsten Hochkirch, Daniele Peri, and Auke van der Ploeg.

References

ABT, C.; HARRIES, S. (2007), *A new approach to integration of CAD and CFD for naval architects*, 6th Conf. Computer and IT Applications in the Maritime Industries (COMPIT), Cortona, pp.467-479, http://data.hiper-conf.info/compit2007_cortona.pdf

ALVAREZ, L. (2000), *Design optimization based on genetic programming*, PhD thesis, University of Bradford, https://www.brad.ac.uk/staff/vtoropov/burgeon/thesis_luis/chapter3.pdf

BERTRAM, V. (2003), *Optimization in ship design*, in Birk and Harries (2003), pp.29-56

BERTRAM, V. (2011), *Practical Ship Hydrodynamics*, Butterworth & Heinemann

BERTRAM, V. (2014), *In-service optimization offered by seakeeping software*, The Naval Architect, November, p.18

BIRK, L. (2003), *Introduction to nonlinear programming*, in Birk and Harries (2003), pp.57-86

BIRK, L.; HARRIES, S. (2000), *Automated optimization – A complementing technique for the hydrodynamic design of ships and offshore structures*, 1st Conf. Computer and IT Applications in the Maritime Industries (COMPIT), Potsdam, pp.72-90, http://data.hiper-conf.info/compit2000_potsdam.pdf

BIRK, L.; HARRIES, L. (2003), *OPTIMISTIC – Optimization in ship design*, Mensch & Verlag http://www.wegemt.com/download/wegemt/39th_WEGEMT_Summer_School_on_Optimistic_Optimization_in_Marine_Design.pdf

BUHAUG, Ø.; CORBETT, J.J.; ENDRESEN, Ø.; EYRING, V.; FABER, J.; HANAYAMA, S.; LEE, D.S.; LEE, D.; LINDSTAD, H.; MARKOWSKA, A.Z.; MJELDE, A.; NELISSEN, D.; NILSEN, J.; PÅLSSON, C.; WINEBRAKE, J.J.; WU, W.Q.; YOSHIDA, K. (2009), *Second IMO GHG study 2009*, Int. Maritime Organization (IMO), London, http://www.imo.org/en/OurWork/Environment/Pollution-Prevention/AirPollution/Documents/SecondIMOGHGStudy2009.pdf

DEMIREL, Y.K.; KHORASANCHI, M.; TURAN, O.; INCECIK, A.; SCHULTZ, M.P. (2014), *A CFD model for the frictional resistance prediction of antifouling coatings*, Ocean Engineering 89, pp.21-31, https://pure.strath.ac.uk/portal/files/41144140/Demirel_etal_OE2014_CFD_model_frictional_resistance_prediction_antifouling.pdf

DIEZ, M.; PERI, D. (2009), *Global optimization algorithms for robust optimization in naval design*, 8th Conf. Computer and IT Applications in the Maritime Industries (COMPIT), Budapest, pp.296-310, http://data.hiper-conf.info/compit2009_budapest.pdf

DIEZ, M.; PERI, D. (2010), *Two-stage stochastic programming formulation for ship design optimization under uncertainty*, 9th Conf. Computer and IT Applications in the Maritime Industries (COMPIT), Gubbio, pp.76-89, http://data.hiper-conf.info/compit2010_gubbio.pdf

DIEZ, M.; PERI, D. (2011), *Optimal hull-form design subject to epistemic uncertainty*, 10th Conf. Computer and IT Applications in the Maritime Industries (COMPIT), Berlin, pp.245-252, http://data.hiper-conf.info/compit2011_berlin.pdf

GIASSI, A.; MAISONNEUVE, J.J.; BENNIS, F. (2003), *Multidisciplinary design optimisation and robust design approaches applied to the concurrent design*, 2nd Conf. Computer and IT Applications in the Maritime Industries, Hamburg, pp.162-174, http://data.hiper-conf.info/compit2003_hamburg.pdf

HARRIES, S. (1998), *Parametric design and hydrodynamic optimization of ship hull forms*, PhD thesis, TU Berlin, Mensch & Buch Verlag

HARRIES, S. (2003), *Geometric modeling and optimization*, in Birk and Harries (2003), pp.115-138

HARRIES, S. (2014), *Practical shape optimization using CFD*, Whitepaper, Friendship Systems, https://www.caeses.com/wp-content/uploads/2014/11/PracticalShapeOptimizationUsingCFD_Whitepaper_FRIENDSHIP-SYSTEMS_Nov2014.pdf?pk_c=whitepaper-download-nov-2014

HARRIES, S.; ABT, C.; BRENNER, M. (2015), *Upfront CAD – Parametric modeling techniques for shape optimization*, Int. Conf. Evolutionary and Deterministic Methods for Design, Optimization and Control with Applications to Industrial and Societal Problems (EUROGEN), Glasgow, pp.1-20

HILDEBRANDT, T.; REYER, M. (2015), *Business and technical adaptivity in marine CFD simulations – Bridging the gap*, 14th Conf. Computer and IT Applications in the Maritime Industries (COMPIT), Ulrichshusen, pp.394-406, http://data.hiperconf.info/compit2015_ulrichshusen.pdf

HOCHKIRCH, K. (2017), *Optimization of ships with asymmetric aftbodies*, Whitepaper DNV GL, https://www.dnvgl.com/maritime/publications/optimization-of-ships-with-asymmetric-aftbodies.html

HOCHKIRCH, K.; BERTRAM, V. (2012), *Hull optimization for fuel efficiency – Past, present and future*, 11th Conf. Computer and IT Applications in the Maritime Industries (COMPIT), Liege, pp.39-49, http://data.hiper-conf.info/compit2012_liege.pdf

HOCHKIRCH, K.; HEIMANN, J.; BERTRAM, V. (2013), *Hull optimization for operational profile – The next game level*, 5th Int. Conf. Computational Methods in Marine Engineering (MARINE), Rome, pp.81-88

HOCHKIRCH, K.; HAHN, C. (2017), *Cloud-based numerical towing tanks – Anytime, anywhere and for anybody*, 11th Symp. High-Performance Marine Vehicles (HIPER), Zevenwacht, pp.66-69, http://data.hiper-conf.info/Hiper2017_Zevenwacht.pdf

HOCHKIRCH, K.; MALLOL, B. (2013), *On the importance of full-scale CFD simulations for ships*, 12th Conf. Computer and IT Applications in Maritime Industries (COMPIT), Cortona, pp.85-95, http://data.hiper-conf.info/compit2013_cortona.pdf

KLEINSORGE, E.; LINDNER, H.; WAGNER, J., BRONSART, R. (2016), *Ship hull form optimization using scenario methods*, PRADS Conf., Copenhagen, https://www.lsb.uni-rostock.de/fileadmin/uni-rostock/Alle_MSF/LSB/Publikationen/PRADS_066.pdf

KRAPP, A.; BERTRAM, V. (2016), *Hull performance analysis – Aspects of speed-power reference curves*, 1st Hull Performance & Insight Conf. (HullPIC), Pavone, pp.41-48, http://data.hullpic.info/HullPIC2016.pdf

MAISONNEUVE, J. (2002), *Towards ship performance improvement using modeFRONTIER*, 5th Num. Towing Tank Symp. (NuTTS), Pornichet, http://www.uni-due.de/imperia/md/content/ist/nutts_05_2002_pornichet.pdf

NOWACKI, H. (2003), *Design synthesis and optimization – An historical perspective*, in Birk and Harries (2003), pp.1-28

PERI, D. (2013), *Automatic tuning of metamodels for optimization*, 12th Conf. Computer and IT Applications in Maritime Industries (COMPIT), Cortona, pp.51-62, http://data.hiper-conf.info/compit2013_cortona.pdf

PERI, D.; ROSSETTI, M.; CAMPANA, E.F. (2001), *Design optimization of ship hulls via CFD techniques*, J. Ship Research 45/2, pp.140-149

PERI, D.; CAMPANA, E. (2005), *Global optimization for safety and comfort*, 4th Conf. Computer and IT Applications in Maritime Industries (COMPIT), Hamburg, pp.477-486, http://data.hiper-conf.info/compit2005_hamburg.pdf

PERIC, M.; BERTRAM, V. (2011), *Trends in industry applications of CFD for maritime flows*, 10th Conf. Computer and IT Applications in Maritime Industries (COMPIT), Berlin, pp.8-18, http://data.hiper-conf.info/compit2011_berlin.pdf

SHIGUNOV, V.; BERTRAM, V. (2014), *Prediction of added power in seaway by numerical simulation*, 9th Int. Conf. on High-Performance Marine Vessels (HIPER), Athens, pp.102-113, http://data.hiper-conf.info/Hiper2014_Athens.pdf

STÜCK, A. (2012), *Adjoint Navier–Stokes methods for hydrodynamic shape optimisation*, PhD thesis, TU Hamburg, https://tubdok.tub.tuhh.de/bitstream/11420/1063/1/adjointNavierStokesMethodsForHydrodynamicShapeOptimisation_arthurStueck.pdf

VAN DER PLOEG, A.; SCHUILING, B. (2018), *Improving an already optimized ship by making its stern asymmetric*, 17th Conf. Computer and IT Applications in Maritime Industries (COMPIT), Pavone, pp.84-97, http://data.hiper-conf.info/compit2017_pavone.pdf

VARGAS, A.; SHIN, H. (2017), *Modeling of ship resistance as a function of biofouling type, coverage, and spatial variation*, 2nd Hull Performance and Insight Conf. (HullPIC), Ulrichshusen, pp.264-281 http://data.hullpic.info/HullPIC2017_ulrichshusen.pdf

Adapt, Adapt, Adapt:
Recent Trends in Multi-fidelity Digital Modelling for Marine Engineering

Riccardo Pellegrini, CNR-INM, Rome/Italy, riccardo.pellegrini@inm.cnr.it
Andrea Serani, CNR-INM, Rome/Italy, andrea.serani@cnr.it
Simone Ficini, University of Roma Tre, Rome/Italy, simone.ficini@uniroma3.it
Riccardo Broglia, CNR-INM, Rome/Italy, riccardo.broglia@cnr.it
Matteo Diez, CNR-INM, Rome/Italy, matteo.diez@cnr.it
Jeroen Wackers, Ecole Centrale de Nantes, Nantes/France, jeroen.wackers@ec-nantes.fr
Michel Visonneau, Ecole Centrale de Nantes, Nantes/France, michel.visonneau@ec-nantes.fr

Abstract

The paper presents some recent trends in multi-fidelity digital modelling for marine engineering applications. Digital modelling is achieved by machine learning methods, namely multi-fidelity surrogate models, trained by computational fluid dynamics (CFD). Adaptative approaches are discussed for radial basis functions and Gaussian process models. Simulation-based design optimisation problems are presented to discuss the use and effects of different adaptivity concepts: (1) adaptive refinement of the computational-domain discretization in CFD; (2) adaptive sampling of the design/operational space; (3) adaptive selection of the fidelity used for the surrogate model training in a multi-fidelity environment; (4) adaptivity of the models to noise. Model adaptation allows for the efficient training of machine learning models, reducing the computational cost associated to building the training sets and improving the overall accuracy of the digital representation.

1. Introduction

Digital models allow for accurate performance prediction, design optimisation, control, uncertainty quantification, operation planning, and predictive maintenance of complex systems (such as ships, marine structures, offshore wind farms). Digital representations may use high-fidelity computational models, low-fidelity and/or reduced-order models, machine learning methods, spanning a variety of approaches, accuracy levels, and computational costs. Physics-based and data-driven models may be integrated and assimilated in modern digital twins, where the digital model is continuously updated with sensor data to accurately represent the status of its real-world physical counterpart (therefore realizing the digital twinning). Integration and assimilation can be realized through (supervised) machine learning methods such as surrogate or neural network models, where the underline input/output relationship between significant parameters is inferred from simulation results and/or sensor data. A surrogate model or metamodel is a model of a model, where the input/output relationship of a complex (accurate and computationally expensive) model is inferred by a simpler (generally less accurate and computationally inexpensive) model for fast assessment. One may think of the resistance of a ship varying its speed, evaluated by computational fluid dynamics (CFD) simulations. The simpler surrogate model that one can imagine is an interpolation of available resistance/speed values provided by CFD. For more complex problems and applications, such as uncertainty quantification and simulation-based design optimisation, the training of such models can be expensive, due to the large amount of data/simulations required for an accurate representation of the desired input/output relationship. For models trained by numerical simulations, several approaches can be used to reduce the time and therefore the cost of the training procedure. These approaches include the use of multi-fidelity surrogate models with adaptive sampling procedures.

Multi-fidelity (MF) approximation methods have been developed, with the aim of combining the accuracy of high-fidelity solvers with the computational cost of low-fidelity solvers, *Peherstorfer et al. (2018)*. A MF surrogate model uses mainly low-fidelity simulations and only few high-fidelity simulations are used to adjust/correct the model to improve its prediction accuracy. A hierarchy of models is usually required to build the MF approximation. In CFD-trained models, different fidelity levels can be obtained by varying the physical model (e.g. potential flow, PF, or Reynolds averaged

Navier-Stokes, RANS, equations) and/or the numerical accuracy of the solution (e.g. grid discretization and/or convergence tolerances), see e.g. *Beran et al. (2020)*. The training efficiency of the MF model and its final accuracy depends on several degrees of freedom (solvers setup, data points, etc.). Adaptive methods that automatically optimise these settings can greatly increase the potential of MF methods. For example, one of the key aspects to consider when building a MF surrogate model is the definition of the design of numerical experiments (DoE) required for the training. The DoE includes both the definition of the desired training points (operating conditions, design parameters, etc.) and fidelity level for the analysis (PF, RANS, grid resolution, etc.). The proper trade-off between model accuracy and computational cost of the training process should be carefully identified. This can be achieved using adaptive sampling methods, which train the model with a DoE that is not defined a priori but dynamically updated, exploiting the information that becomes available during the training process. By adding training points where it is most useful, adaptive sampling approaches increase the model accuracy while reducing the computational cost associated to the training, *Liu et al. (2018)*.

The objective of this paper is to present recent trends in MF surrogate modelling for CFD applications using adaptation. Specifically, the fidelity of CFD solutions is automatically adapted during the model training using an adaptive grid refinement approach, *Wackers et al. (2017)*. Adaptive sampling methods are used to run simulations only where it is most informative, *Volpi et al. (2015)*, and selecting the proper fidelity level, *Serani et al. (2019)*. Finally, the MF model self-adapts using regression to cope with the presence of noise in the CFD output. Results are shown and discussed for stochastic radial basis functions (SRBF) and Gaussian processes (GP) models. Three applications are presented and discussed, namely the shape optimisation of a naval destroyer, a hydrofoil, and a roll-on/roll-off passenger (RoPax) ferry. The paper shows: (i) how adaptive grid refinement adds computational grid nodes only where it is most useful, improving the capability of capturing the flow features and allowing to define an arbitrary number of fidelity levels in a MF environment; (ii) how the adaptive sampling procedure can efficiently manages the training process, adding training points to the DoE only where it is most informative and using the fidelity that is most convenient; and (iii) how self-adaptivity of the surrogate model by selecting regression parameters can effectively cope with the possible presence of noise in the training data.

2. Multi-fidelity surrogate modelling

In this section some definitions and concepts of MF surrogate modelling are briefly recalled. Details of the mathematical formulation with examples may be found in *Volpi et al. (2015)*, *Serani et al. (2019)*, and *Wackers et al. (2020a)*. Consider an input/output relationship of the type $f(\boldsymbol{x})$, where f is a desired performance metric and $\boldsymbol{x} \in \mathbb{R}^D$ is a vector of dimension D collecting design and/or operational parameters. Let the true function $f(\boldsymbol{x})$ be assessed by numerical simulations with fidelity levels denoted by l as in $f_l(\boldsymbol{x})$ where, in general. an arbitrary number N of fidelity levels can be used. Here, the following hierarchy is used: $l = 1$ indicates the highest fidelity and $l = N$ indicates the lowest fidelity. The MF surrogate model $\hat{f}$ is built as a low-fidelity surrogate model $\tilde{f}_N$ "corrected" with the surrogate models $\tilde{\varepsilon}_l$ of the inter-level errors (or discrepancies) between consecutive fidelities, i.e. $\hat{f}_1 = \tilde{f}_N + \tilde{\varepsilon}_{N-1} + \cdots + \tilde{\varepsilon}_1$. Note that each fidelity level has its own training set, which is used for modelling the

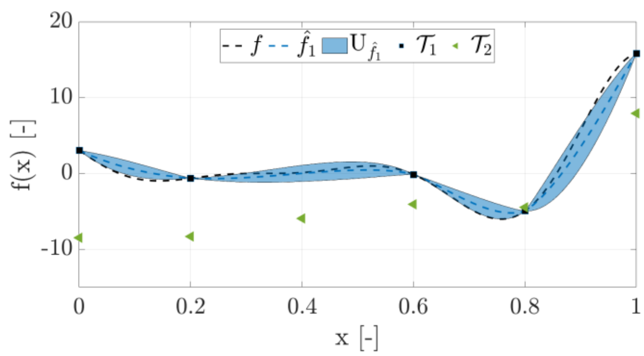


Fig.1: MF surrogate model with two fidelities (N = 2), *Serani et al. (2019)*

functions and/or discrepancies. In this work SRBF and GP surrogate models are used, which provide the prediction along with the associated uncertainty $U_{\tilde{f}_l}(\boldsymbol{x})$ of each level. The resulting uncertainty $U_{\hat{f}_1}(\boldsymbol{x})$ associated to the MF surrogate model prediction is a combination of the prediction uncertainties of each level. Fig.1 shows an example for one design/operational variable (D = 1) and two fidelity levels (N = 2).

3. Adaptivity of the computational-domain discretization: adaptive grid refinement

A first step towards adaptivity comes from the adoption of adaptive grid refinement methods for the definition of the CFD mesh, *Wackers et al. (2017)*. With this method an initial coarse grid is built and then its cells are divided based on the water surface position and on second derivatives of pressure and velocity. Thus, cells are added only where it is most useful to improve the capture of the flow features. The refinement level is expressed by a real value parameter that is user controlled, allowing to define arbitrary levels of fidelity. Fig.2 shows the refined mesh for the 5415 DTMB hull, it is worth noting that the grid is finer close to the hull and at the air-water interface. Fig.3 shows another advantage of using an adaptive grid refinement method, when a shape optimisation is demanded it is possible to modify the shape within a coarse grid, see Figs.3a-b, reducing the problems connected to negative volumes and excessively skewed cells. Then, during the simulation, the grid will be refined up to the desired level, Fig.3c. Further details can be found in *Wackers et al. (2020b)*.

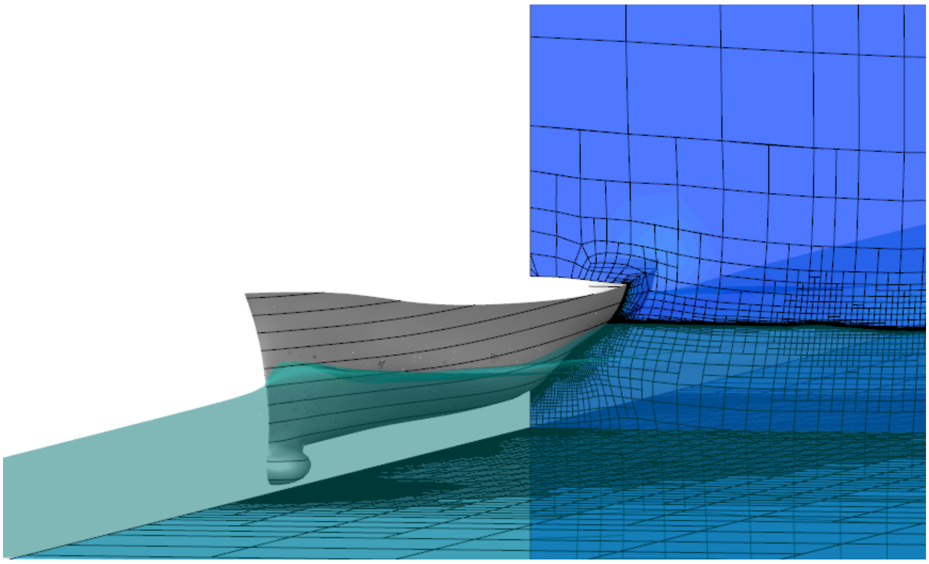
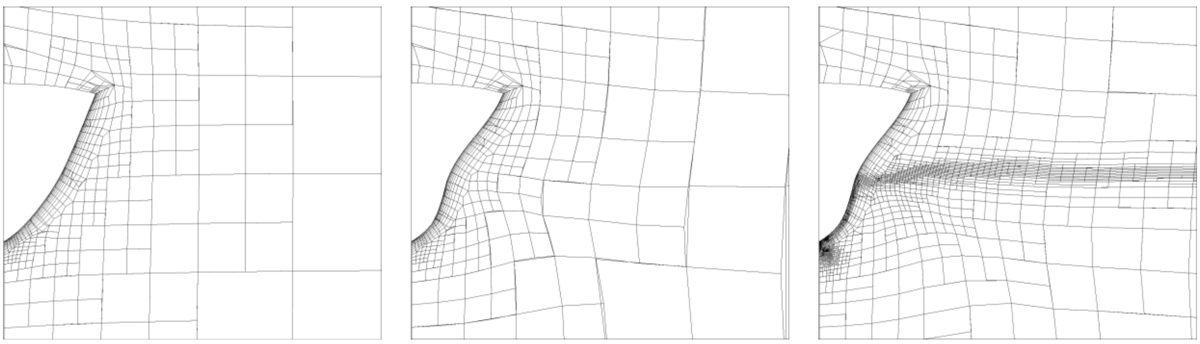


Fig.2: Shape optimisation with adaptive grid refinement example: 5415 DTMB hull optimisation

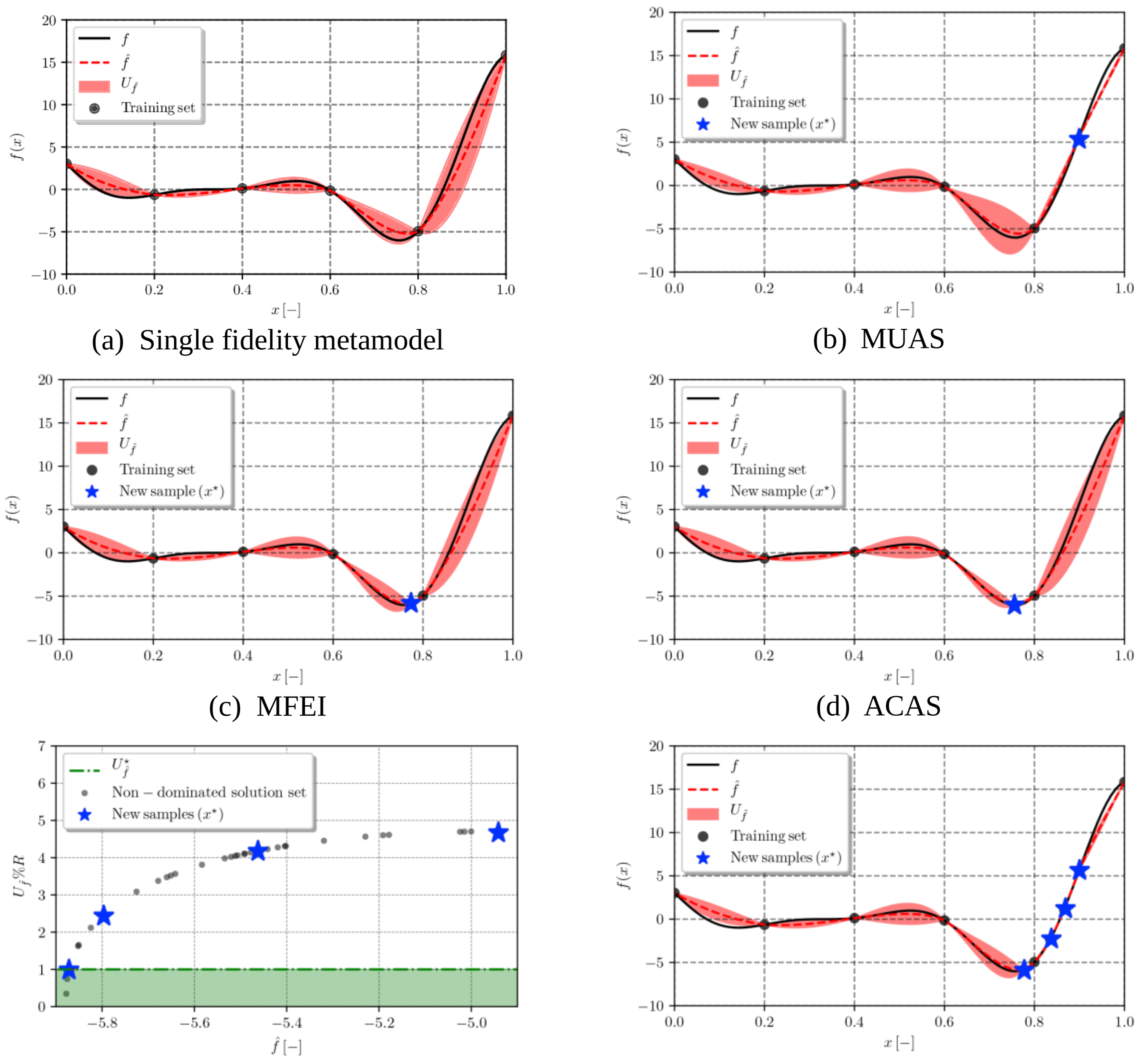


(a) (b) (c)

Fig.3: Shape optimisation with adaptive grid refinement process: (a) original, (b) deformed, and (c) refined mesh, *Serani et al. (2019)*

4. Adaptivity of the design/operational space exploration: adaptive sampling

A second type of adaptivity stems from the method used to dynamically update the surrogate model training set. A relatively small-size initial training is defined. Then, the metamodel accuracy is improved by adding samples only where it is most informative. In general, in a sampling process a trade-off exists between the global exploration of the design/operational space and its local exploitation (e.g. investigating an optimal solution). Depending on the aim of the design/operational space exploration (e.g. assessing globally the performance as in uncertainty quantification or finding an optimal design as in design optimisation), exploration or exploitation may be preferred, or a combination of them. To illustrate the differences between exploration- and exploitation-oriented sampling approaches, the following methods are discussed, considering the design optimisation context, *Serani et al. (2019)*: i) maximum uncertainty adaptive sampling (MUAS, exploration oriented); ii) MF expected improvement (MFEI, exploitation oriented); iii) aggregate criterion adaptive sampling (ACAS, combining exploration/exploitation); iv) multi-criteria adaptive sampling (MCAS, combining exploration/exploitation). All sampling methods consider the uncertainty associated to the prediction as a metric for the accuracy of the local representation of the desired function. For instance, the prediction uncertainty is an indication of the discrepancy of solutions provided by different kernels in SRBF and therefore is taken as an estimate of the model accuracy. Similar considerations apply for GP, which provides directly with the variance (uncertainty) associated to the prediction.



(a) Single fidelity metamodel

(b) MUAS

(c) MFEI

(d) ACAS

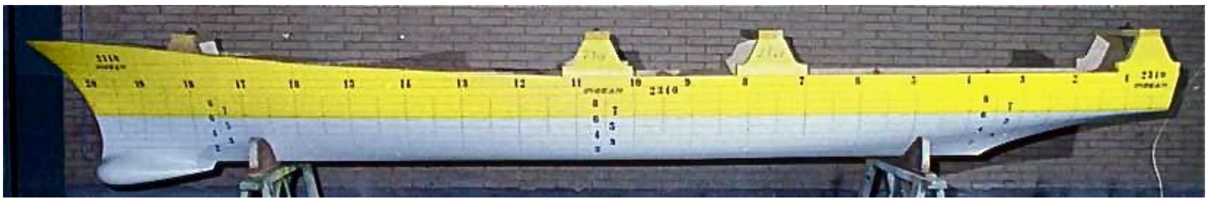
(e) MCAS, Pareto front in the desired function and prediction uncertainty space

(f) MCAS, new samples

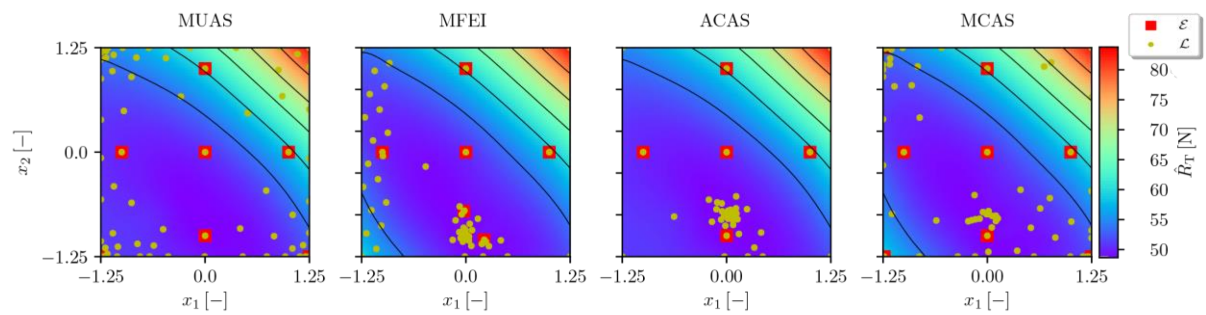
Fig.4: Adaptive sampling methods, *Serani et al. (2019)*

Fig.4 shows how new training points are identified following the different methods. MUAS identifies the point with the maximum prediction uncertainty, see Fig.4b. MFEI identifies the point with the highest probability of further reducing the desired objective function, see Fig.4c. ACAS identifies the function minimum considering the lower bound of all possible predictions, see Fig.4d. Finally, MCAS first identifies a Pareto front composed by points with large uncertainty and small objective function value. Then, the Pareto front is sub-sampled with a user-defined number of points, which represents the new samples, see Figs.4e and 4f. MCAS can be seen as a multi-objective extension of ACAS and allows to perform the parallel infill of new training points, thus taking advantage of high performance computing systems that allow to run more simulations at once. All these sampling methods can be used either with SRBF or GP, since both models provide the function prediction with the associated uncertainty.

To demonstrate the effects of the adaptive sampling method on a design optimisation procedure, the shape optimisation of the DTMB 5415 (bare hull) is shown and discussed for total resistance in calm water at Fr = 0.30 and Re = $1.18 \cdot 10^7$. The DTMB 5415 model, see Fig.5a, is an open-to-public early concept of a USS Arleigh Burke-class destroyer, widely used for towing tank experiments, CFD studies, and as hull-form optimisation benchmark. CFD simulations are performed by RANS with the ISIS-CFD code, *Queutey and Visonneau (2007)*. Computational grids are created through adaptive grid refinement and mesh deformation to take into account the need for high- and low-fidelity, as well as the different geometries needed for shape optimisation. Fig.5b shows the final iteration of the sampling strategies, MUAS explores the design space and provides an almost uniform sampling of the domain without any consideration of the objective function minimum; MFEI clusters the samples in the region close to the minimum and provides some samples in the left side of the design space, identified as promising at the beginning of the sampling; ACAS strictly focuses on the region around the minimum with a quite limited exploration of the design space; finally, MCAS provide a reasonable trade-off between exploration and exploitation, focusing on the region around the minimum but also exploring more broadly the design space. Solutions are close; MFEI identifies the best shape with a total resistance reduction of 4% compared to the original hull. The resulting reduction of the bow wave can be seen in Fig.6. The adaptive sampling procedure allows for finding an approximate optimal solution with a quite small computational cost, enabling shape optimisation also in the case of limited computational budgets. Further details can be found in *Serani et al. (2019)*.

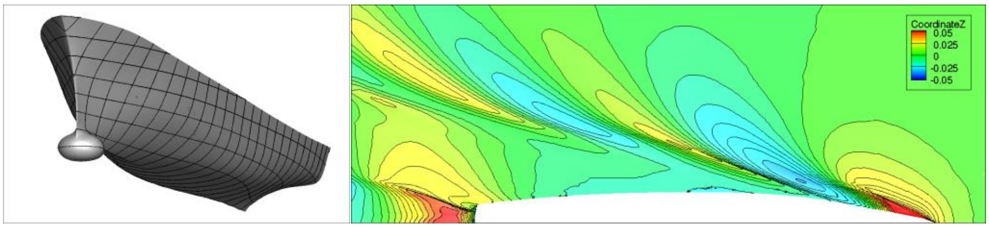


(a) Geosym replica of the DTMB 5415 (CNR-INSEAN model 2340)

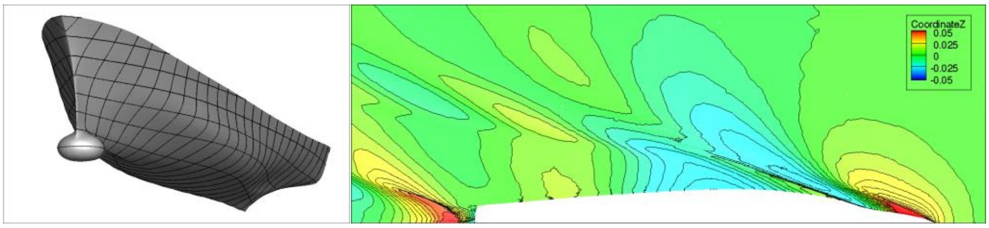


(b) Final iteration of the adaptive sampling methods with SRBF metamodel

Fig.5: DTMB 5415 optimisation (Fr = 0.30 and Re = $1.18 \cdot 10^7$) with SRBF based MF metamodel, *Serani et al. (2019)*



(a) Original



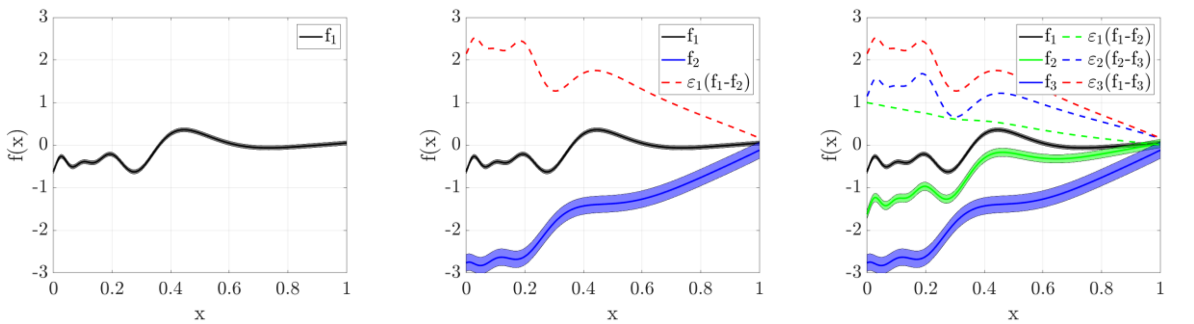
(b) Optimised

Fig.6: DTMB 5415 optimisation (Fr = 0.30 and Re = $1.18 \cdot 10^7$): comparison between original and optimised hull forms, *Serani et al. (2019)*

5. Adaptivity of the fidelity used for training: The *N*-fidelity approach

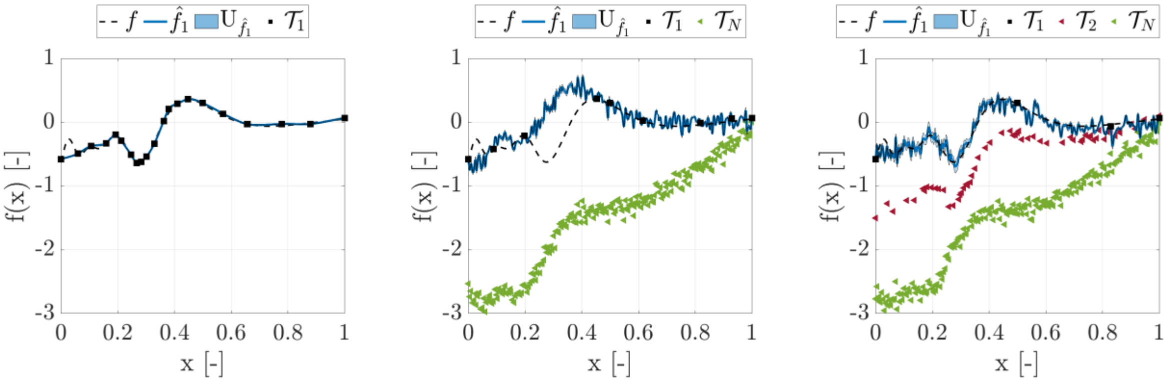
In the MF context, the use of more than one fidelity level adds an additional degree of freedom in the adaptive sampling process. Once the design/operational space regions that need additional training points are identified, it is also necessary to define the fidelity level able to produce the greatest benefit in terms of accuracy versus computational cost. This selection can be made automatically and adaptively as the surrogate model training progresses, e.g. based on the ratio between the prediction uncertainty and the computational cost associated to each fidelity level.

As an example, results with N = 3 fidelity levels are shown for an analytical test (extended from *Clark et al. 2016*) and a NACA hydrofoil shape optimisation for reduced drag coefficient with the MUAS method (the interest reader can be referred to *Wackers et al. 2020a*). The NACA is optimized at constant lift coefficient $C_L = 0.6$ and Re = $8.41 \cdot 10^6$. Fig.7 shows the analytical test problem with the high- (f_1), medium- (f_2), and low-fidelity (f_3) levels along with an artificial noise band for each fidelity (representative of CFD output noise). Increasing the number of fidelity levels improves the capability of the MF surrogate model to represent the desired function (see Fig.8). Similar consideration can be made looking at the NACA hydrofoil optimisation problem. Fig.9 shows the hydrofoil grids (see Figs.9a-c), along with the noise associated to the CFD computations for one design variable (see Figs.9d-f).



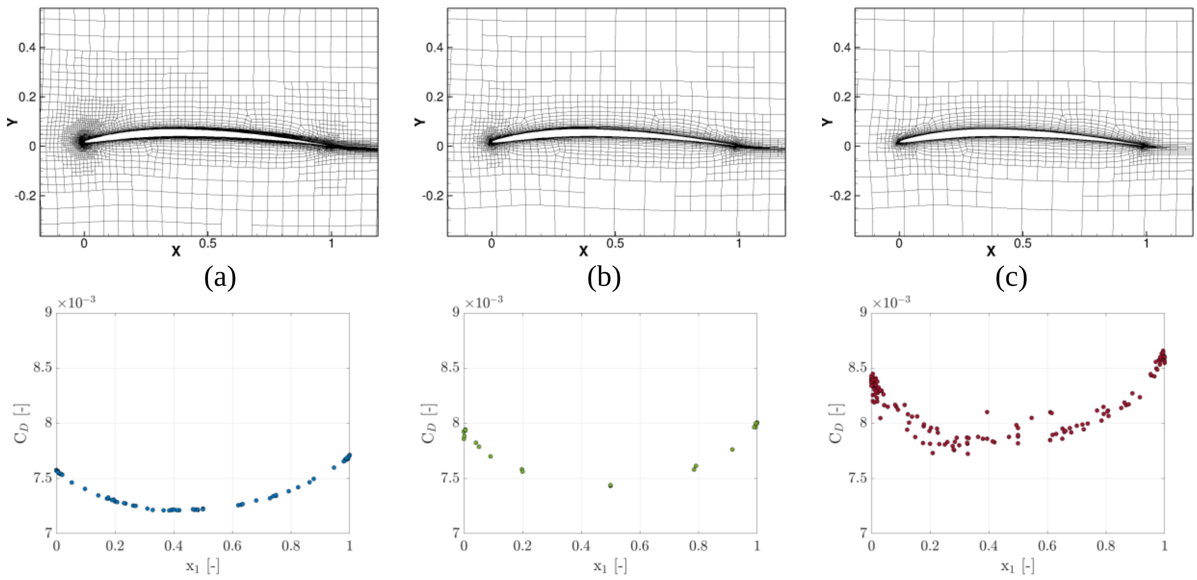
(a) N = 1 (b) N = 2 (c) N = 3

Fig.7: Analytical test problem with different number of fidelities and noise bands, f_1 and f_2 functions without noise are taken from *Clark et al. (2016)*. Example taken from *Wackers et al. (2020a)*



(a) $N = 1$ (b) $N = 2$ (c) $N = 3$

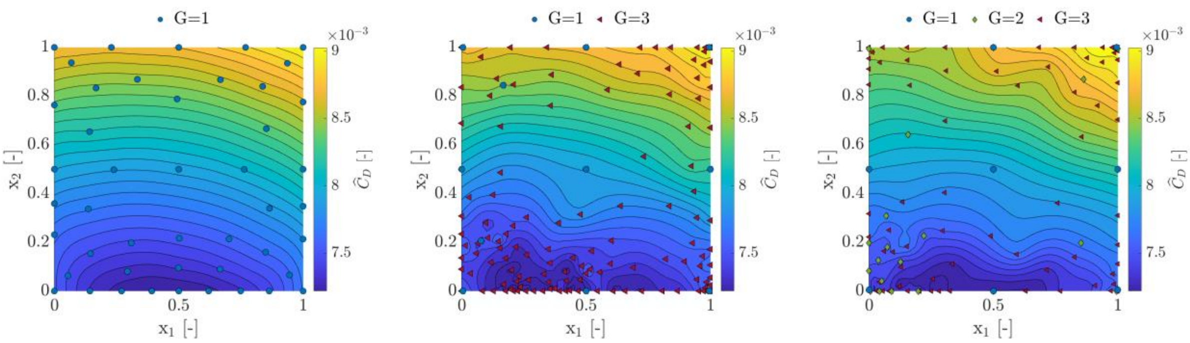
Fig.8: Analytical test problem: final iteration of the adaptive sampling procedure with SRBF and MUAS method, *Wackers et al. (2020a)*



(a) (b) (c)

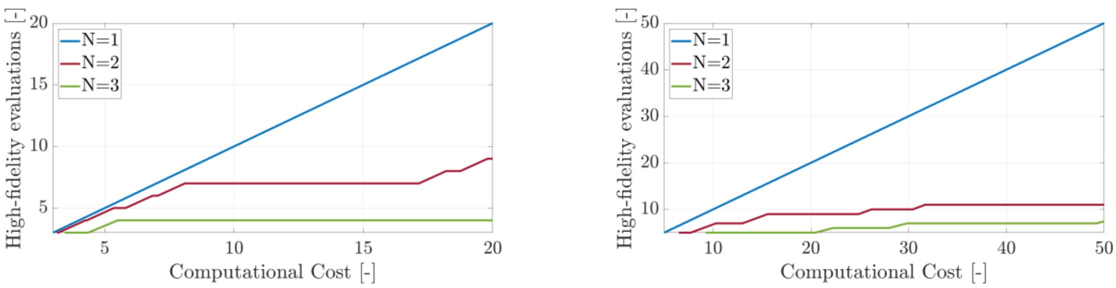
(d) (e) (f)

Fig.9: NACA hydrofoil shape optimisation ($C_L = 0.6$ and Re = $8.41 \cdot 10^6$): high-, medium-, and low-fidelity grids (a, b, and c) and CFD outputs (d, e, and f) fixing one design variable to highlight the numerical noise, *Wackers et al. (2020a)*



(a) $N = 1$ (b) $N = 2$ (c) $N = 3$

Fig.10: NACA hydrofoil shape optimisation ($C_L = 0.6$ and Re = $8.41 \cdot 10^6$): final iteration of the adaptive sampling procedure with SRBF and MUAS method, *Wackers et al. (2020a)*



(a) Analytical test problem (b) NACA hydrofoil shape optimisation

Fig.11: High-fidelity training set size varying the number of fidelities, *Wackers et al. (2020a)*

Fig.10 shows that the introduction of an additional fidelity level (thus moving from N=2 to N=3) produces a smoother response surface. Finally, Fig.11 shows that the higher the number of fidelity levels the greater the reduction of the high-fidelity evaluations and therefore the cost.

A further example using four (N = 4) fidelity levels is provided for the resistance minimization of a RoPax ferry at Fr = 0.245 and Re = $1.017 \cdot 10^7$. Details can be found in *Wackers et al. (2020b)*. For this example, a multi-grid approach is used (see Fig.12) instead of adaptive grid refinement. CFD simulations are performed with the χnavis RANS code, *Broglia and Durante (2018)*. Fig.13 shows the surrogate model prediction and the associated uncertainty at the final iteration using the ACAS sampling method. The sampling is strictly focused on the global minimum region and the overall surrogate prediction uncertainty is low. The MF surrogate model provides a prediction error at the minimum close to 10% and an objective function improvement equal to 12.7%.

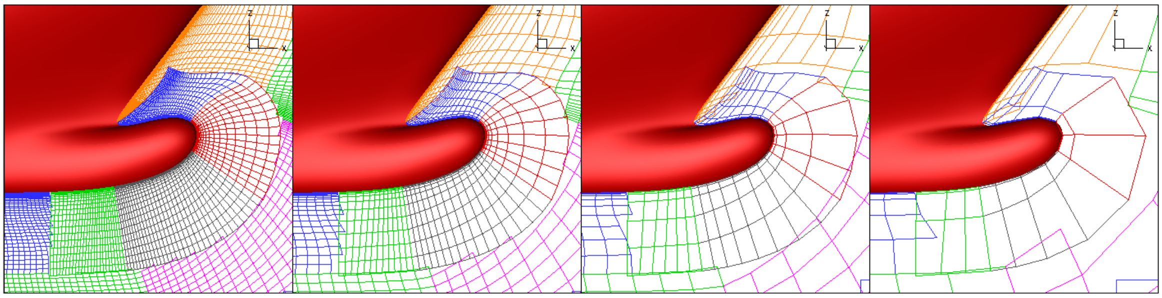
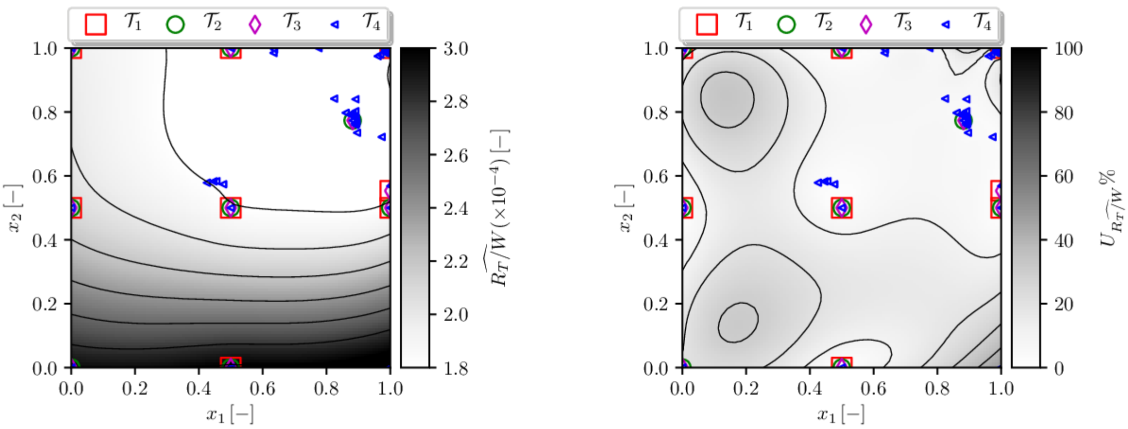
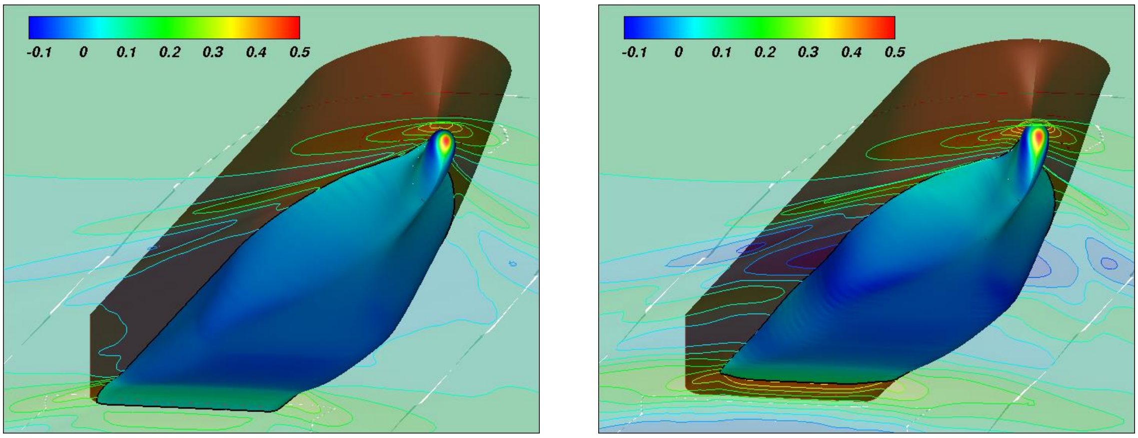


Fig.12: Shape optimisation with multi-grid: example with four levels of refinement, bulb region, *Wackers et al. (2020b)*



(a) Resistance metamodel prediction (b) Resistance metamodel prediction uncertainty

Fig.13: RoPax optimisation (Fr = 0.245 and Re = $1.017 \cdot 10^7$): final iteration of the adaptive sampling procedure, *Wackers et al. (2020b)*



(a) Original (b) Optimised

Fig.14: RoPax optimisation (Fr = 0.245 and Re = $1.017 \cdot 10^7$): comparison between original and optimised hull shapes, pressure field, and wave patterns, *Wackers et al. (2020b)*

Fig.14 shows the comparison between (a) the original and (b) the optimised hull. The optimised hull is characterized by a reduction of the submergence of the stern region and a less pronounced bulbous bow.

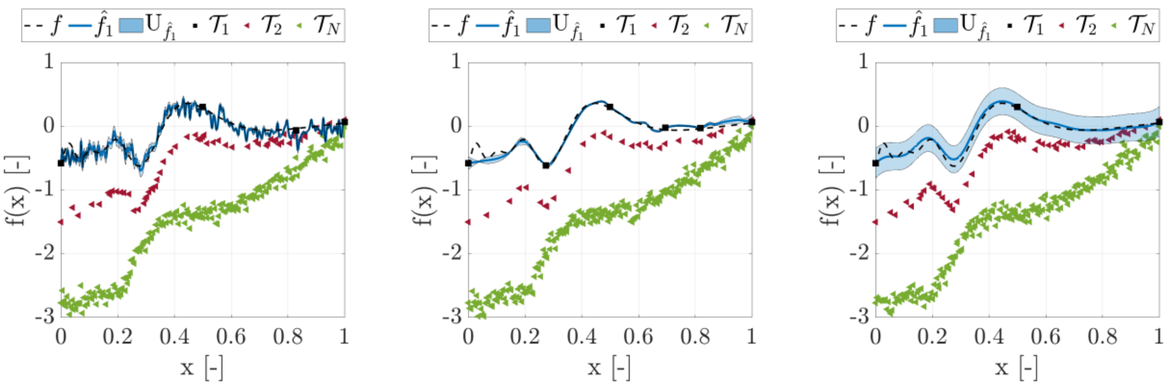
Overall, adding fidelity levels and managing adaptively their contribution to the model training increase significantly the optimisation procedure efficiency.

6. Adaptivity of the model to noise: Auto-tuning using regression

Generally, the presence of numerical noise in CFD outputs is somehow unavoidable and cannot be neglected a priori (see e.g. Figs.9d-f). The use of a surrogate model that performs an exact interpolation of the training points may lead to overfitting and loss of accuracy, since the surrogate model reproduces all the numerical (non-physical) fluctuations in the data, see e.g. Figs.8b-c. To overcome this problem and filter out the noise, regressive surrogates may be used. Regressive models generally require the definition of several tuning parameters. Model auto-tuning involves the automatic selection of these parameters. Furthermore, when dealing with noisy data it is also useful to have a surrogate capable to assess the “amount” of noise (*e.g.* as standard deviation of the data).

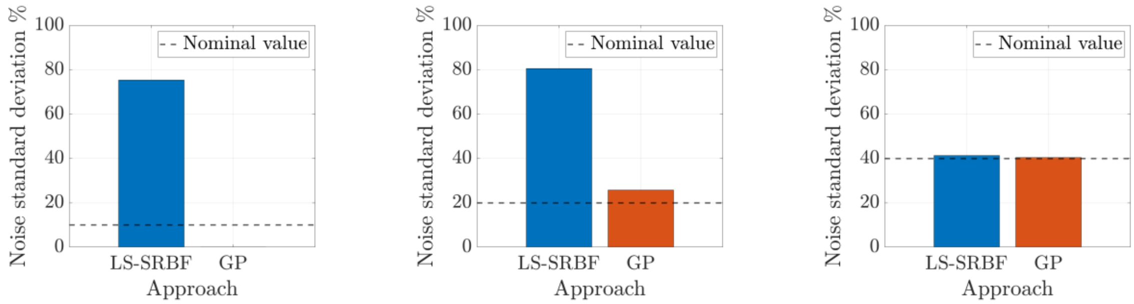
Here, SRBF uses least-squares regression (LS-SRBF) to filter the noise out. The auto-tuning capability is based on a leave-one-out cross-validation procedure with the automatic selection of the number of SRBF centres based on an error metric. The assessment of the amount of noise in the data is based on the root mean squared error between the training set and the surrogate model prediction. GP auto-tuning is based on the identification of an optimal set of internal parameters as the result of a log marginal likelihood maximization. GP uses an internal parameter to explicitly model the standard deviation associated to the noise in the training set. Further details can be found in *Wackers et al. (2020a)*. Results with three fidelity levels are shown for the analytical test problem (see Figs.15 and 16, where I-SRBF indicates the interpolatory version of SRBF) and the NACA hydrofoil optimisation (see Figs.17 and 18, where again I-SRBF indicates interpolation), described in the previous section.

Figs.15 shows the final iteration of the adaptive sampling procedure comparing I-SRBF with LS-SRBF and GP for the analytical test problem. The use of LS-SRBF as opposed to I-SRBF allows to filter out the noise in the training set, achieving a better representation of the desired function. GP achieves similar results to LS-SRBF. Fig.16 shows the predicted standard deviation of the noise in the training set. LS-SRBF and GP are not able to assess properly the noise for the high-fidelity training set, since very few samples are available (small training set size). GP achieves quite accurate results for the medium-fidelity training set. Finally, both LS-SRBF and GP are able to identify accurately the noise in the low-fidelity training set (characterized by the largest training set size).



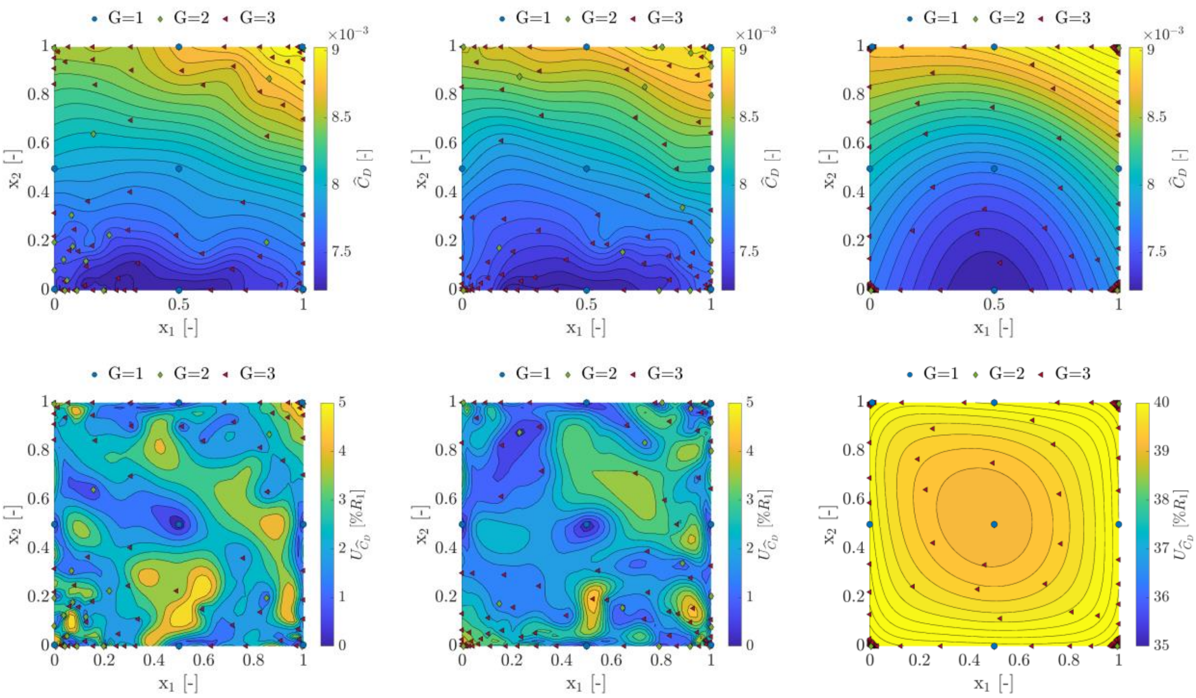
(a) I-SRBF (b) LS-SRBF (c) GP

Fig.15: Analytical test problem: final iteration of the adaptive sampling procedure with different metamodels, *Wackers et al. (2020a)*



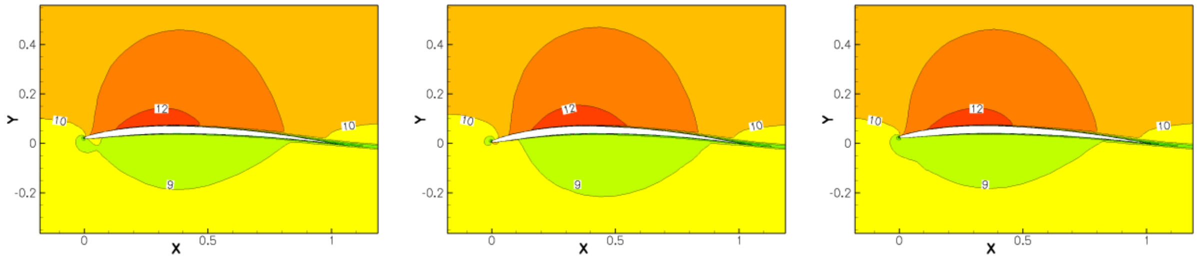
(a) High-fidelity (b) Medium-fidelity (c) Low-fidelity

Fig.16: Analytical test problem: noise standard deviation of the modelled training sets



(a) I-SRBF (b) LS-SRBF (c) GP

Fig.17: NACA hydrofoil shape optimisation ($C_L = 0.6$ and Re $= 8.41 \cdot 10^6$): final iteration of the adaptive sampling procedure, metamodel prediction (top) and prediction uncertainty (bottom), *Wackers et al. (2020a)*



(a) I-SRBF (b) LS-SRBF (c) GP

Fig.18: NACA hydrofoil shape optimisation ($C_L = 0.6$ and Re = $8.41 \cdot 10^6$): contour of the velocity magnitude of the optimised hydrofoil shapes, *Wackers et al. (2020a)*

Figs.17 shows (top) the surrogate model predictions and (bottom) the associated uncertainties at the final iteration of the adaptive sampling procedure using I-SRBF, LS-SRBF, and GP for the NACA problem. LS-SRBF provides a smoother response surface and a more uniform distribution of samples than I-SRBF. GP shows a significant clustering of samples at the corners of the domain, but also produces a very smooth response surface. GP shows the smoothest trend for the uncertainty, but also quite large uncertainty values. Finally, Fig.18 shows the optimised NACA hydrofoil shapes. These are similar; small differences are present in the leading edge of the hydrofoil that are deemed to be the origin of the numerical noise.

Using adaptivity to noise represents a further improvement for the model training efficiency. Model overfitting is avoided, and the overall accuracy is increased. Optimisation solutions are achieved with a reduced number of high-fidelity calls and therefore reduced computational cost.

7. Conclusions and future directions

Some recent trends in multi-fidelity surrogate modelling for CFD applications have been presented. The thread that connects these trends is the adaptivity. Adaptivity allows to efficiently manage the available resources (*e.g.* computational cost and time for CFD simulations) in the training of fast digital models for design optimisation and uncertainty quantification. The following conclusions may be drawn:

1. The use of an adaptive grid refinement method allows the realizations of meshes able to capture the flow features up a user-desired accuracy, while keeping the simulation computational cost lower than using uniformly refined mesh. Adaptive grid refinement also provides an arbitrary number of fidelity levels for use in the MF context.

2. The use of adaptive sampling methods allows to train the models adding new samples only where it is most informative. This increase the model accuracy and reduce the computational cost associated to training the model.

3. The adaptive sampling can be extended to the use of more than one fidelity (physical model, grid, etc.). This allows the selection of the most convenient fidelity level, balancing the model accuracy with the computational cost. As overall results, the training procedure is further optimised, achieving more accurate models with reduced computational cost.

4. Auto-tuning capabilities of surrogate models allow to cope with numerical noise in the training data, which is generally unavoidable when CFD outputs are use. Auto-tuning avoids model overfitting and the risk of oversampling. It produces smoother models and represents a further improvement for model accuracy and training efficiency.

Future research will address the extension of adaptive methods discussed here to the more general context of multi-information source applications, where defining hierarchy of training sets is not straightforward (e.g. when RANS computations with pretty coarse grids are used together with non-linear PF computations with well resolved panel meshes). The possibility of integrating sources with heterogenous or missing data is also of interest, as well as the extension to time-varying outputs (e.g. time series).

Acknowledgements

CNR-INM is grateful to Dr. Woei-Min Lin, Dr. Elena McCarthy, and Dr. Salahuddin Ahmed of the Office of Naval Research and Office of Naval Research Global, for their support through NICOP grant N62909-18-1-2033. The HOLISHIP project (HOLIstic optimization of SHIP design and operation for life cycle, www.holiship.eu) is also acknowledged, funded by the European Union's Horizon 2020 program under grant agreement N. 689074. Part of the work is conducted in collaboration with NATO STO AVT task group on "Goal-driven, multi-fidelity approaches for military vehicle system-level design" (AVT-331). Dr. Riccardo Pellegrini is partially supported through CNR-INM project OPTIMAE.

References

BROGLIA, R.; DURANTE, D. (2018), *Accurate prediction of complex free surface flow around a high speed craft using a single-phase level set method*, Computational Mechanics 62/3, pp. 421-437

CLARK, D. L.; BAE, H.-R.; GOBAL, K.; PENMETSA, R. (2016), *Engineering Design Exploration Utilizing Locally-Optimized Covariance Kriging*, 18th AIAA Non-Deterministic Approaches Conference, p. 0428

LIU, H.; ONG, Y.-S.; CAI, J. (2018), *A Survey of Adaptive Sampling for Global Metamodeling in Support of Simulation-based Complex Engineering Design*, Structural and Multidisciplinary Optimization 57/1, pp.393-416

QUEUTEY, P.; VISONNEAU, M. (2007), *An Interface Capturing Method for Free-Surface Hydrodynamic Flows*, Computers & Fluids 36/9, pp.1481-1510

PEHERSTORFER, B.; WILLCOX, K.; GUNZBURGER, M., (2018) *Survey of multifidelity methods in uncertainty propagation, inference, and optimization*, Siam Review, Vol. 60, No. 3, pp. 550-591

SERANI, A.; PELLEGRINI, R.; WACKERS, J.; JEANSON, C.-E.; QUEUTEY, P.; VISONNEAU, M.; DIEZ, M. (2019) *Adaptive multi-fidelity sampling for CFD-based optimisation via radial basis function metamodels*, Int. J. Computational Fluid Dynamics, 33/6-7, pp.237-255

VOLPI, S.; DIEZ, M.; GAUL, N.J.; SONG, H.; IEMMA, U.; CHOI, K.K.; CAMPANA, E.F.; STERN, F. (2015), *Development and validation of a dynamic metamodel based on stochastic radial basis functions and uncertainty quantification*, Structural and Multidisciplinary Optimization 51/2, pp.347-368

WACKERS, J.; DENG, G.; GUILMINEAU, E.; LEROYER, A.; QUEUTEY, P.; VISONNEAU, M.; PALMIERI, A.; LIVERANI, A. (2017), *Can adaptive grid refinement produce grid-independent solutions for incompressible flows?*. J. Computational Physics 344, pp.364-380

WACKERS, J.; VISONNEAU, M.; FICINI, S.; PELLEGRINI, R.; SERANI, A.; DIEZ, M. (2020a), *Adaptive N-Fidelity Metamodels for Noisy CFD Data*, 21th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conf., Virtual Event

WACKERS, J.; VISONNEAU, M.; SERANI, A.; PELLEGRINI, R.; BROGLIA, R.; DIEZ, M. (2020b), *Multi-Fidelity Machine Learning from Adaptive- and Multi-Grid RANS Simulations*, 33rd Symp. Naval Hydrodynamics, Osaka

Measuring the Effect of Augmented Reality on Situational Awareness

Stephan Procee, Clark Borst, René van Paassen, Max Mulder, NHL Hogeschool, West-Terschelling/The Netherlands, procee@nhl.nl

Abstract

In order to measure the effect of Augmented Reality (AR) on Situational Awareness (SA), an experiment has been designed and performed on a ship bridge simulator with experienced participants with more than six months sea time. The AR interface is based on Ecological Design principles, i.e. information on the environment and the ship's possibilities, i.e. manoeuvring, is shown. The measurements show that there is a positive effect on all three levels of SA.

1. Executive summary

The main features of the work conducted are as follows:

- Objective: This experiment tests whether displaying operational information in an Augmented Reality interface has an effect on Situational Awareness of the user, i.e. the watchkeeper.
- Background: Keeping a safe navigational watch involves both looking outside and observing many displays and instruments. The hypothesis is that showing the relevant information in a video overlay the watchkeeper's construction of system information and the outside reality is faster and better, the workload is lower and lookout will be more effective.
- Method: Participants run two short scenarios on a bridge simulator in which they only observe their situation and report to a number of probes. The independent variable is the use of Augmented Reality during the scenarios. Each participant starts randomly to either one of the scenarios and is also randomly assigned to use Augmented Reality. From the measured reaction time to the probes and the rated quality of the given answers the degree of Situation Awareness is determined.
- Analysis: The hypothesized effect of Augmented Reality on the degree of Situational Awareness and workload can be tested with a t-test on reaction time, and by a Mann Whitney U test on rated answers.
- Results: 22 professional candidates participated in the experiment. Due to the COVID-19 outbreak, the school and its simulator center had to be closed which led to an abrupt halt of the experiment. The reaction times of the limited group ($n = 22$) were analyzed. Each of the reaction times on the total of 18 probes during the two scenarios did not significantly differ ($p > 0{:}05$) when Augmented Reality was used.
- Conclusion: The candidates showed great variation in age and professional experience. Although maximum care was taken to balance both groups, the lack of homogeneity in either group resulted in too much variance in the test results. Also, the number of candidates was too small to produce a meaningful average of these test results.
- Application: Questionnaires taken before and after the experiment show that candidates see great potential in using Augmented Reality in their work as navigators. Hardware related issues will resolve in time. The working principle of enriching an interface by overlaying system information on the outside visual world is expected to support the task of the navigator.

2. Situational Awareness

The main goal of this first experiment on the use of Augmented Reality (AR) Ecological Interface (EI), a.k.a. Marine Navigation Augmented Reality System, MNARS, in marine navigation is to measure the hypothesized effect on the Watch Officer's (WO) ability to build and maintain Situation Awareness (SA). In this paper the term MNARS will be used referring to the collection of hardware, i.e. HMD, Tracker and processors, and software providing the interface.

In the professional application AR is about enriching the observed environment with the aim to amplify cues in this environment. By observing this enriched environment, the WO is potentially supported in his assessment of the situation, hence his SA improves (SA as defined by any of the 26 definitions of SA mentioned by *Breton and Rousseau (2003)*).

This infers that in general the hypothesized effect of MNARS, can be measured by measuring SA.

However, *Sarter and Woods (1991)* have pointed out that SA is not a clear-cut, strictly defined, quality. Despite its generally accepted importance in several professional fields, there is not one generally accepted interpretation or method to ascertain the worker's awareness of his situation. Much depends on the professional context and application. *Breton and Rousseau (2003)* analyzed 26 definitions of SA that were present in literature. From this it can be inferred that there is no such thing as the SA of a WO. However, there seems to be an understanding that the measurement method, called SAGAT, designed and described by *Endsley (2012)* and *Endsley and Jones (2012)*, provides an accepted measure of SA. Some elements of SAGAT, however, are also subject to critique, e.g. freezing the situation in order to interview the subject might focus on the subject's ability to memorize specific elements instead of measuring the subject's understanding of his situation during action.

Without pretending to overcome the problem of ill-defined SA, this experiment aims at designing an experiment to test a number of aspects related to SA. From these results it can be inferred that there might be a positive effect of AR on a selection of tasks that the WO is responsible for. A complete description of these tasks, i.e. the Navigator's Work Constraints, is given by *Procee et al. (2017)*.

3. Experimental design

The experimental design is based on two simulated scenarios comprising of typical traffic situations in which an individual participant uses AR, condition one, or does not use AR, being condition two. The planned number of participants is about fifty from which each individual is randomly assigned one of the two test conditions and randomly either one of the two scenarios to start with. Each scenario lasts about 12minutes. After a 15-minute break the participant is assigned the other scenario. This is known as an independent measures experimental design; the results of the control group are compared with those of the experimental group. Beforehand, participants are introduced in the experiment and trained in wearing an HMD and using the experimental MNARS interface.

The motivation for this independent-measures experiment design lies mainly in its simplicity and its insensitivity to potential unbalance in the relatively small subject groups. Testing each candidate in two consecutive scenarios will introduce a learning effect, i.e. the candidate uses experience from the previous run. Hence, the problem of comparing the results of e.g. a subject using AR in his first run with the results of a subject without AR in his second run (and in the same scenario), must be dealt with.

Ideally, the homogeneity of the subject group is high in order to reduce the variation in measurements. Therefore, a population of students with at least six months of sea-going experience and three years of professional training and education is targeted at.

3.1. Description of simulated scenario

The test scenario is designed in such a way that there is no need for the participant to maneuver, i.e. the Own Ship (OS) sails at pre-set autopilot and fixed engine power. Hence, the sailed route and the observed situation is constant for each subject. The WO observes the progress of the OS and the developing traffic situation, which conforms to the normal working practice for a ship underway on a coastal trip. During the first three minutes of the scenario no probing is done in order to allow the participant to build SA. During the remaining 9 minutes the subject is probed verbally with nine questions that have to be answered in a loud and clear manner. These probes are recorded, spoken by a neutral voice unknown to the participants, in an audio-file in order to keep the timing and intonation of the spoken questions constant for each test run.

Each subject's test run is recorded on video. The camera is positioned outside the subject's vision at the rear of the simulator bridge thus providing a recording of the subject's physical motion and attitude. The subject is alone in the simulator room.

In order to eliminate the confounding limited field of view of the HMD, the test condition two, i.e. not using AR, is executed with the subject wearing the HMD yet without the overlaying interface. The HMD's oculars provide about 70% transparency, thus, in each condition the subject observes the outside reality with the same field of view and light intensity.

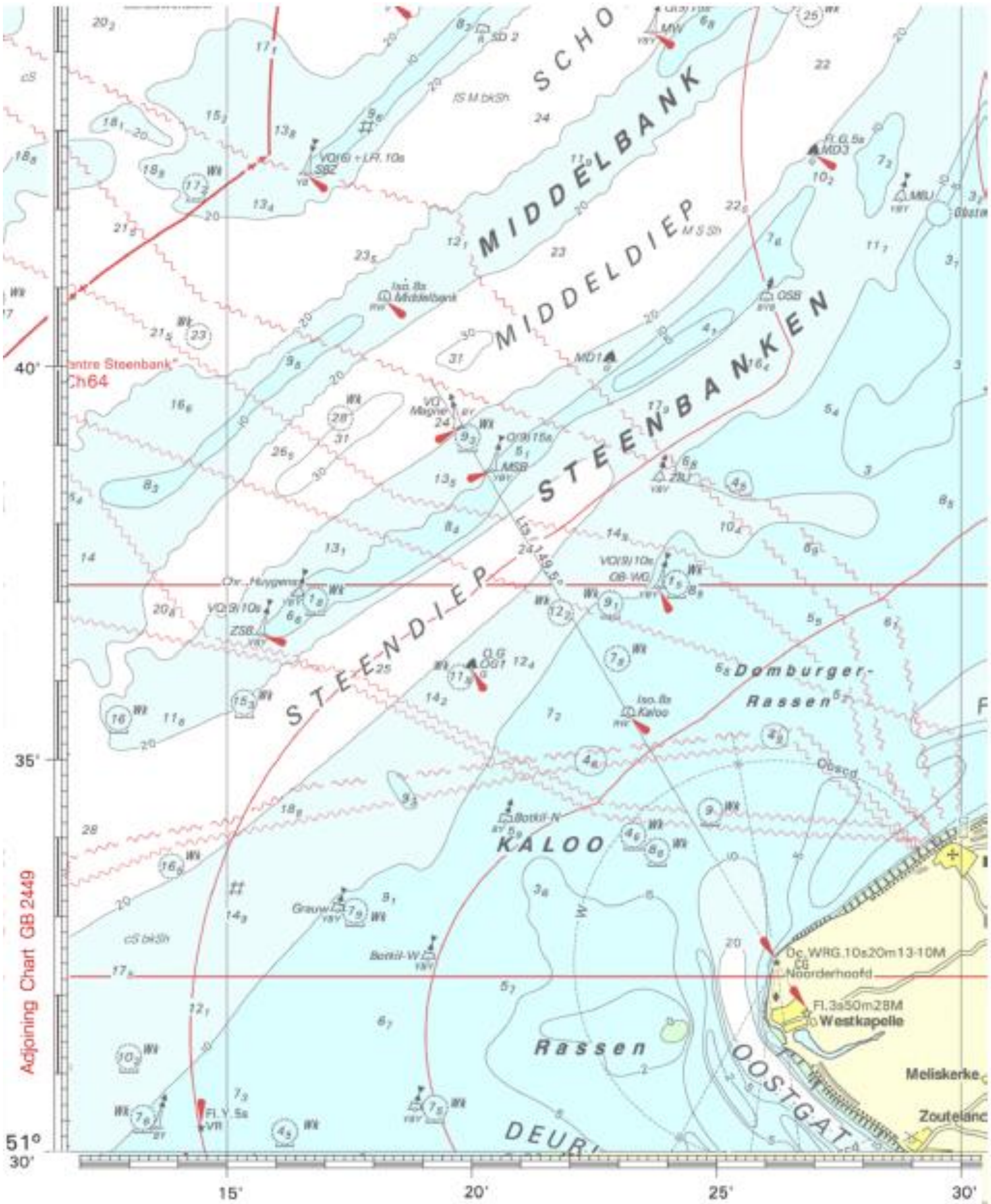


Fig.1: Steenbank area, leading line 149.5°/329.5° for both scenarios

Because the goal of this experiment is to empirically test the effect of MNARS in a realistic work environment, the simulated scenario should resemble as much elements affecting human operation as possible. The simulated visual sensation of a vessel's sea keeping movement while progressing in a dense traffic situation is usually well taken care of by contemporary ship's bridge simulators. However, the other sensory elements, e.g. sound, smell or feel, need some additional care to increase the scenario's fidelity. From these sensory elements, sound has the greatest effect on the perceived reality, and is also known to affect the worker's stress level, *Wickens and Holland (2000)*, pp.480-492. Therefore, a noisy environment is created in order to evoke a moderate level of fatigue and stress. Also, the runs are held during late afternoon until evening in order to evoke additional fatigue, i.e. additional to the subject's normal working day.

Individual variation from subject to subject in both test groups, e.g. level of experience, verbal fluency, a spontaneous or a precise attitude, will lead to variation in the measured response time and score of correct answers. This variation can be canceled out by using a large number of subjects in each group, carefully balanced, i.e. each group has an equal distribution of age, experience and sex.

Although scenarios are different from each other, though equally complicated, there will be a learning effect from one scenario to the other. By randomly assigning either one of the scenarios to the subject as his first run it is assumed that the influence of this learning effect is minimal when the means of both groups are compared with each other.

One scenario comprises a run over the Steenbank in the Northwest direction, starting near the Kaloo fairway buoy, the other is the opposite direction, inland bound, starting near the Middelbank fairway buoy, Fig.1. The area is chosen for its relatively long constant course along a buoyed fairway, with a strong tidal cross current and busy traffic with fishing vessels and commercial traders entering and leaving the Westerschelde estuary from the north. Working with two short scenarios, instead of one, has the advantage that environmental conditions, i.e. traffic, wind, current and visibility, can be varied. This enables making more measurements in slightly different conditions. The side effect is that subjects have a short break in between the runs and can do a new start with more experience.

3.2. Measuring

The deployed measurement method is known as 'Real Time Probing'. This measurement is chosen for a number of reasons. One reason is that it mimics a real-life situation, e.g. the captain entering the bridge and asking the WO some details about the present status; this is common practice and is often an implicit check on the WO's SA, a.k.a. 'challenge and response'. Another motivation for probing is that it better suits the question whether the subject 'knows' what is going on in his continuous work-process rather than recollects what 'was' going on after the run or at a freeze. This, again, resembles the realistic situation of watch-keeping.

One of the WO's tasks is track keeping, i.e. stay within a narrow margin of the planned route or course line. In both scenarios the participant is not expected to manoeuvre the ship, i.e. change course and or speed. Instead, the participant is asked to observe, and report his assessment at specific moments during the scenario. From the time it took to answer and from the quality of the given assessment it can be inferred whether the WO understands his position and movement relative to the intended track.

Another task is to participate safely in traffic, i.e. can the WO discriminate between the targets around and potentially act. This complicated task relates to observation, i.e. level 1 SA, understanding, i.e. level 2 SA, and predicting, i.e. level 3 SA. Some probes, questions, relate to observation, some to understanding and one question relates to prediction. The timing, i.e. latency, of the given answer as well as the quality of the answer is recorded as measure for the degree of SA.

It is normal practice that the WO is confronted with many constraints, i.e. tasks at hand, that have to be prioritized. From this the effect of AR can be inferred. It is assumed that measuring the WO's ability to work out a task, e.g. calculating the Estimated Time of Arrival (ETA) to the pilot station based on speed and distance, is directly related to the degree of awareness. The WO is only able to pay attention to alternative tasks, i.e. other than the previously mentioned primary tasks, if he feels comfortable with the situation, i.e. situationally aware.

3.3. Environment

To reduce confounding variables, a number of measures have been taken. Each scenario replicates itself exactly in the simulator. Because the simulator room lacks windows there is no influence of daylight. Lighting conditions are kept the same with the same number of operating monitors in the bridge console. The environmental audio noise is initiated by a 'ghetto blaster' playing a CD with engine noise, spoken probes, and an occasional telephone ring. In the background a radio is playing light music. This is all to replicate the 'normal' working environment on board which is necessary to reliably measure an effect of AR. Utmost care is taken to start the simulator at the precise timing, i.e. within a second, of the audio file. To avoid voice recognition the probes are recorded by an unknown female, speaking Oxford English in a clear and audible manner.

During the test run the candidate was alone on the bridge. The colleague candidate was outside the simulator premises to avoid getting knowledge of the probes and getting acquainted with the audible scenario.

3.4. Briefing

The response to a request for candidates on social media, i.e. Willem Barentsz Facebook, was overwhelming. A selection of potential candidates was made on the basis of relevant experience. These professionals were invited to pick their preference from a number of slots, where after they were invited to take part in the experiment. Some were willing to travel to Terschelling from as far as Antwerp! Prior to the planned time slot, an explanation of the test and its procedures, the simulator environment and a chart fragment of the Steenbank was sent to the invited candidate.

Each time slot lasted 105 minutes, divided into 45 minutes of training and briefing and four runs of 15 minutes each. The two candidates did an alternating run, resulting in a 15-minutes break in between.

The training consisted of a short explanation of the working principle of Augmented Reality, and the explanation of the Velocity Obstacles diagram and the colors and symbols that are shown in the interface. It was stressed that the Velocity Obstacles diagram was not relevant for this experiment. Candidates got the time to get acquainted with using a HMD and were trained in observing traffic and fairway markers.

3.5. Balancing the test group

In order to balance the groups, a scheme was made beforehand comprising two varieties, one was the dependent variable AR, and the second was the sequence of the scenarios, i.a. South bound first. At the start of the time slot the candidates were randomly assigned either of the two varieties. Because potential candidates were actively employed and sometimes had to adjust their plans on short notice, it happened more than once that slots had to be reassigned, therefore it was decided to fill the scheme in clusters of four at the time.

This resulted in a less than optimal balance between both groups which was partly due to the pandemic induced abrupt halt of the experiment.

4. Analysis of results

The audio tracks of the video were used for timing the response. These were loaded in a video editor that displayed sound level in combination with a time bar. This visualization allowed for more precise and reliable timing. Reaction time was measured in seconds, rounded to a half second. Although it seems straightforward to measure time between the ending of a question and the start of an answer, it appeared that subjects use all kinds of verbal pause expressions, e.g. “the course is .... ehhhh.... ehhhh... three two seven”. Arbitrarily the response timing was chosen to stop at the actual answer, i.e. “three two seven” in this case. The resolution of time was chosen to be a half second. The start time of the answer and the end time of the probe was rounded to a half second. The reason for this is that individuals vary in their vocal use, which makes their start time vague. Also, for questions ending with a consonant e.g. heading, the end is impossible to time in a precision better than half a second.

The quality of the given answers was, arbitrarily, rated into four categories. The first qualified as being precise, clear and convincingly pronounced and correctly answering the probe. These are the responses that would be expected from a WO with perfectly developed SA. The second category is a less secure answer, there might be an audible question mark at the end or there is an overall hesitation in the answer. This category would point to a somewhat lesser developed SA. The third category is pointing to a lack of SA, i.e. there is not a clear response, the answer is undecided, or not answered at all. The fourth category points to a wrong yet convincingly given answer. This might be the result of a misinterpreted probe or a blunder, it might show a complete lack of SA as well.

4.1. Test group homogeneity

From the 45 selected participants only 22 could be invited for the experiment before the school and simulator were closed down on 13.3.2020. Unfortunately, most our school’s potential 45 students that were targeted felt they were too busy and too stressed for exams to participate in the experiment.

The 22 tested candidates varied in age between 20 years and 65 years, distributed in five younger than the age of thirty, seven younger than forty, one younger than fifty, five younger than sixty and the remaining three younger than 65. One candidate didn’t mention his age. Their years of experience corresponded to their age since all were still actively involved in their profession. They all had over three years’ experience with ship’s bridge simulators. Their motivation to participate in the experiment could be expressed in six possible ways. Six candidates chose, “I always want to work with new technological developments”, six chose, “I am curious about the experiment (how it is done)”, four chose, “I want to test if augmented reality really works”, one candidate chose, “ I always help when being asked for”, and three candidates chose, “I like to express my opinion on technical developments whenever I have the chance”. Two candidates chose all five motivations.

Apart from asking their motivation to participate in this experiment, their opinion on using new technology in their work was asked. Two candidates chose. “I am not happy with the introduction of new technology that I have to work with”, one candidate chose “I am neutral”, twelve chose “I am curious, this might help” and seven chose “I am very happy with the new techniques, it sure will help in our job”.

Eight candidates used glasses for far-field to correct for Myopia, fourteen did not use glasses. The HMD can accommodate for users wearing glasses.

The candidates were asked to rate their present competence in observing and analyzing the traffic situation. Two felt very competent without doubt whatsoever, twelve felt competent to observe and analyze without error, six felt neutral rather secure but not insecure, one felt slightly less competent (a bit insecure) and one felt a little incompetent (not very secure on my ability to observe and analyze).

Also were candidates asked to rate their present knowledge about collision regulations. No one answered “a bit poor, I feel quite insecure about my knowledge”, one chose, “just enough for daily life at sea, some improvement would be better”, thirteen chose, “I know most of the rules, but I’m not certain about some” and the remaining five chose “perfect knowledge, let’s go for it!”.

Finally, candidates were asked to rate their present fitness for work. One candidate chose, "very tired, I had a long day and this has to be my last activity", three candidates chose, "tired or sleepy, but not exhausted", also three chose, "halfway my energy, not too tired, not too energetic", ten candidates chose, "it's OK this is my usual state and I can manage the rest of a long day easily" and five chose, "Let's go for it! Can't wait".

From the results of this pre-test questionnaire it can be concluded that the majority of candidates were highly motivated, fit for work, and competent in their knowledge and skill. Opinions were frankly given. Hence, there was no reason to assume that the present state of an individual candidate had a determining influence on his test results.

4.2. Test group balance

Candidates were assigned to either the control group or the experimental group. Ideally these two groups are balanced. Both groups had an equal number of candidates wearing glasses, four at each of them.

There was a slight imbalance between groups and the assigned scenarios. Five candidates in the control group were assigned the northbound scenario first against four candidates in the experimental group. Six candidates in the control group were assigned the southbound scenario first against seven candidates in the experimental group.

Also, the fitness for work showed a slight imbalance, three people in the control group felt tired sleepy or very tired against one candidate who felt tired or sleepy. Also, the attitude towards new technology showed a slight imbalance, three candidates in the control group were not happy to neutral against nil with that attitude in the experimental group.

The rated competence in observing showed a balance. The rated knowledge of COLREG showed a minor imbalance, the one candidate who rated 'just enough for daily work' was assigned, by chance, to the control group. The motivation for participating the experiment happened to be nicely balanced.

The age of candidates, however, showed a more serious imbalance. Three candidates in the experimental group were younger than 30 against two in the control group, five candidates in the control group were younger than 40 against two candidates in that range and of the eldest, 60 and above, three were assigned the experimental condition and one the control condition. This imbalance is illustrated in Fig.2 where the lowest bar corresponds to the group younger than 30 years and the top bar corresponds with the group older than 60. Column C, in blue, indicates the number of candidates in the control group, and Column D, in red, indicates the number of candidates in the experimental group.

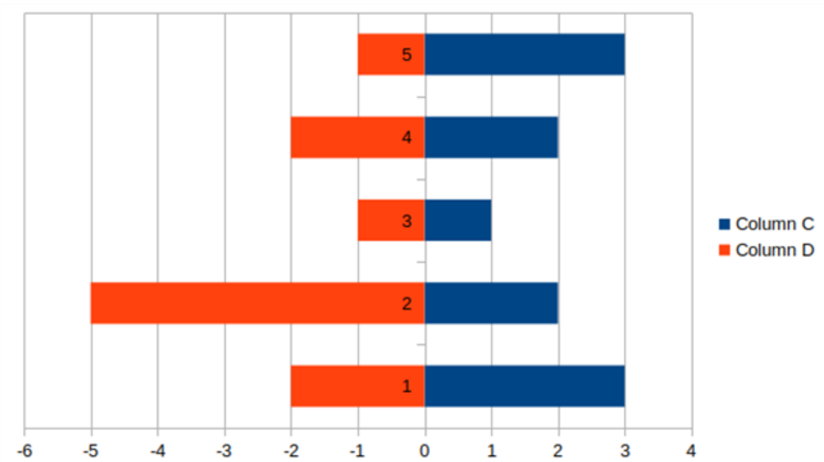


Fig.2: Distribution of age, control group Column C vs. experimental group Column D

4.3. Final questionnaire

In a questionnaire directly after the last run, the candidates were asked to reflect on the following statements:

This example of Augmented Reality interface... (multiple answers are possible)

1 improves watch keeping because there's more time for lookout;
2 improves watch keeping because there's a better understanding of track keeping;
3 improves watch keeping because it's easier to detect and understand the Aids to Navigation around;
4 has no effect on watch keeping because it does not provide relevant information;
5 has no effect on watch keeping because the RADAR/ECDIS display provides sufficient information;
6 has a bad effect on watch keeping because it makes you complacent (i.e. a false impression of safe situation);
7 has a bad effect on watch keeping because the synthetic information distracts you from reality.
0 no answer

The benefit in detecting NavAids was mentioned the most, 18 times, Fig.3. The importance of a careful and visually well-balanced interface design is shown by the number of times (seven) that candidates expressed fear of distraction.

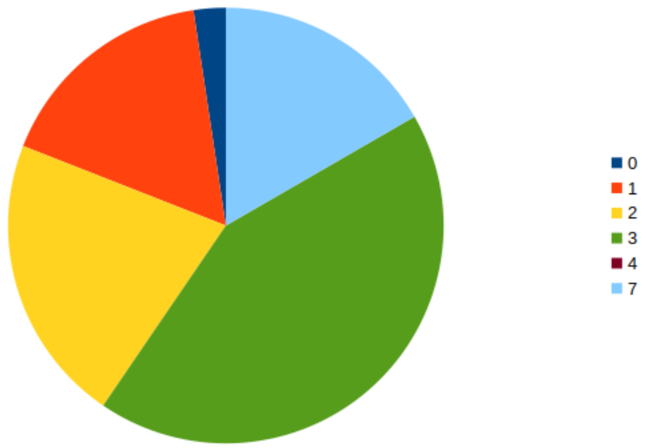


Fig.3: General opinion on this example of Augmented Reality interface

Which of the following statements reflects your opinion about Situational Awareness that you experienced during the experiment? (give one answer)

1 I understood the traffic situation faster;
2 I noticed no difference in the time to comprehend what was going on;
3 It took me longer to understand the situation;
4 None of the above.

The remarkable outcome is that candidates mentioned 'faster understanding' the most, (13 times), Fig.4. The alternatives 'No difference' and 'longer' were each 3 times mentioned.

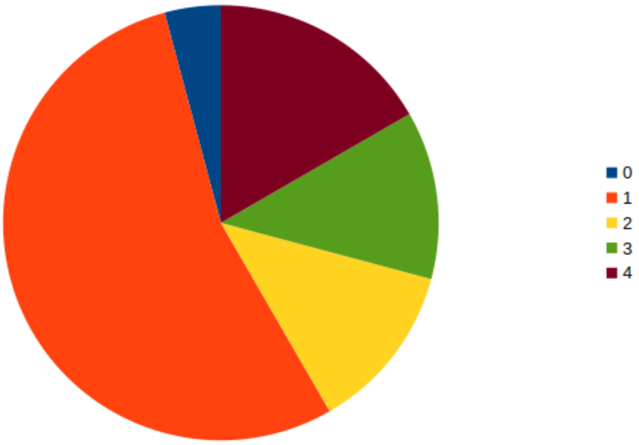


Fig.4: Opinion on the time to observe and analyze traffic

Which of the following statements reflects your opinion about Situational Awareness that you experienced during the experiment? (give one answer)

1 I had a better understanding of my own ship in relation to fairway and traffic;
2 I don't feel it made any difference in understanding what was going on;
3 I found it hard to understand where I was and how the traffic situation developed;
4 None of the above.

Most candidates mentioned a better understanding of their position, (9 times), Fig.5. A remarkable outcome is that six candidates mentioned option three. Some of these might have been in the control group. This is not further analysed.

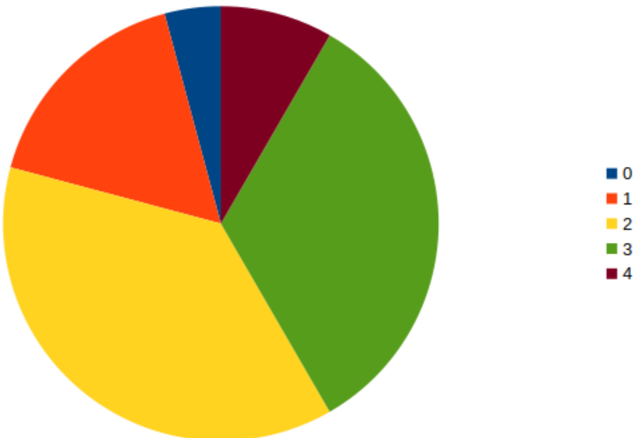


Fig.5: Opinion on understanding the general situation (position and traffic)

Which of the following statements reflects your opinion about the duration of the experiment? (give one answer)

1 This is too short to detect an effect on Situation Awareness;
2 In twelve minutes I can build a satisfying degree of Situation Awareness;
3 I was completely aware of the situation in less than 8 minutes.
4 None of the above

The majority felt able to build situational awareness during each test run, some were fast (9 times), some were very fast (8 times), Fig.6.

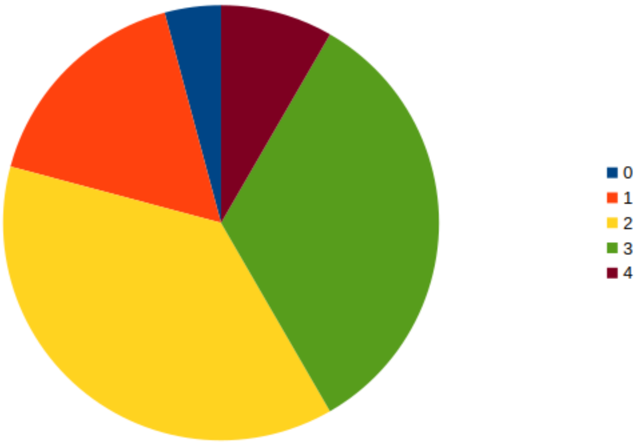


Fig.6: Opinion on the duration of the test runs

Which of the following statements reflects your opinion about the scenarios of the experiment? (give one answer)

1 The used scenarios were too complicated to understand fully what was going on;
2 The used scenarios were too simple to detect an effect on Situation Awareness;
3 The used scenarios were realistic and challenged my skills as navigator;
4 None of the above.

The majority mentioned realistic and challenging scenarios, (17 times). Some (3 times) found the scenarios too simple.

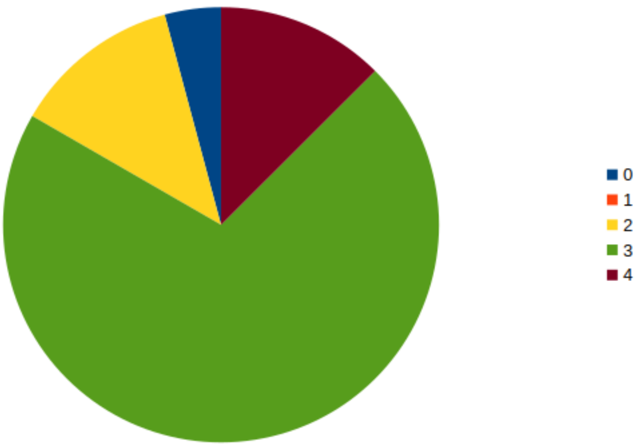


Fig.7: Opinion on the fidelity of the test runs

4.4. Statistical evidence for the experimental effect

One hypothesized effect was that the response time to probes will decrease with the effect of Augmented Reality. It is assumed that a higher degree of SA enables a quicker answer to probes that refer to the momentary situation, hazards, future and possibilities. Because of the experiment design, an independent samples t-test was used to analyze the difference between the mean response time of the control group and the experimental group for each probe. These 18 tests showed not only great variation in their means, some responses from the control group were even faster than those of the experimental group, but also in their sample distribution, hence a large standard deviation. For each test the calculated difference between the means could not be explained by the test effect because of the large standard deviation. Hence, for this group of professional navigators ($n = 22$) the effect of Augmented Reality on reaction times was not significantly improved ($p > 0{:}05$). This was shown for each individual probe. An additional t-test was done for probes during the same initial or the same second run in order to compare candidates with the same learning effect. This could neither show a significant effect, on the contrary, due to the reduced amount of cases, i.e. roughly half of the former group, the reliability of this test deteriorated even further. Because a significant test effect could not be shown, further details are not mentioned here.

The quality of the answers was rated and for each probe a Mann-Whitney U test was used to analyze the hypothesized effect, i.a. better answers are given when AR is used. Only one probe, 6a, tested significantly different. This referred to the expected course change of the vessel Antares; nine candidates from the experimental group provided a correct answer, rated 1, against two candidates in the control group. All other probes tested insignificant, i.a. the differences could not be explained by the test effect, or didn't show any difference at all (one probe only).

5. Discussion

The small number of candidates in combination with a lack of homogeneity distorted the group wise comparison of means. Trying to select candidates with assumingly more coherence, e.g. from the same age, was not successful. This would result in marginally small numbers of candidates per selection rendering statistical analysis useless.

The, only, significant difference in the rated quality of the answers for probe 6a can be explained by the lack of using ARPA/RADAR by the control group. Due to the length of cable attached to the HMD and the position of the tracking emitter most candidates felt uneasy moving around at the bridge. This hampered their ability to control the ARPA/RADAR with which they usually would identify other vessels.

6. Conclusion

For this limited number of participants, the test was not effective in showing the hypothesized effect of AR on SA. Most of the professional participants felt afterwards that they could get a better and faster understanding of their situation when using AR. Nearly all candidates mentioned that they had developed a satisfactory SA, i.a. both the experimental group and the control group.

Candidates also had the opinion that the duration of both runs was long enough and the scenario realistic enough to have confidence in the validity of the experiment. In general candidates felt positive about the development of AR where they felt that it could be supportive in their watchkeeping tasks.

All candidates used more time to get adapted to wearing the HMD and interpreting the synthetic overlay than they expected beforehand.

Careful planning in the distribution of candidates and getting sufficient candidates are of paramount importance for an effective experiment.

References

BRETON, R.; ROUSSEAU, R. (2003), *Situation Awareness: A Review of the Concept and its Measurement*. Technical report, Defence Research and Development Canada, Valcartier

ENDSLEY, M.R. (2000), *Situation Awareness Analysis and Measurement*, Taylor & Francis Inc.

ENDSLEY, M.R.; JONES, D.G. (2012), Designing for Situation Awareness, 2nd ed., CRC Press

PROCEE, S.; BORST, C.; PAASSEN, R.v.; MULDER, M. (2017), *Toward Functional Augmented Reality in Marine Navigation: A Cognitive Work Analysis*, 16th COMPIT Conf., Cardiff

SARTER, N.B.; WOODS, D.D. (1991), *Situation Awareness: A Critical but Ill-defined Phenomenon*, Int. J. Aviation Psychology 1/1, pp.45-57

WICKENS, C.D.; HOLLANDS, J.G. (2000), *Engineering Psychology and Human Performance*,3rd ed., Psychology Press

Digital Maritime Training in Covid-19 Times

Tracy Plowman, Volker Bertram, DNV GL, Hamburg/Germany, tracy.plowman@dnvgl.com

Abstract

The paper discusses digital postgraduate and professional training in maritime technologies, which experienced significantly increased demand since the outbreak of the Covid-19 pandemic. After a description of key options for live training and self-paced training, some experience and lessons learnt are shared. Options range from "simple" videoconferencing to sophisticated 3d Virtual Reality based training.

1. Introduction

A simulation of the 3d flow around a ship with Flettner rotors is a safe topic for a conference for naval architects and marine engineers. An estimated three colleagues may be interested to start with. A suitable mixture of complex equations, daunting diagrams and colorful displays will evoke admiration, little interest and no aggression.

In comparison, education in engineering is a dangerous topic. All engineers have been exposed to the topic (as students), and quite a few in the audience are involved in teaching. It is a bit like soccer:

- It used to be better in the past.
- The players (students or trainees) today are spoilt and not like they used to be
- The coaches (professors or trainers) are incompetent. We could do a better job.
- It is still great fun to talk about…

Teaching environments and techniques have changed since the Covid-19 pandemic, Fig.1. Like it or not, the training community had to embrace digital, remote forms of training rapidly. Post-graduate training in industry and academic training were affected likewise.

Fig.1: Before Covid-19 (left) and during Covid-19 (right)

The first response to the lock-down of physical classrooms has been employing ad-hoc measures such as reading PowerPoint lectures in videoconferences. But at the same time, a more fundamental discussion has started on

- how to provide quality training if the pandemic stays with us for longer;
- whether we should return to traditional classroom training if we could; or whether the experience with currently deployed remote methods indicates that a change towards more digital training methods is called for anyway.

2. Bringing management expectations and technology developments together

Classification Societies are big on rules and regulations. Training is no exception. Plan approval engineers and surveyors need proof of appropriate qualification to perform certain tasks, e.g. verifying the proper installation of an exhaust gas cleaning system ("scrubber"). Without record of such a qualification (= participation in a training), a surveyor may not perform the task and a customer may have to wait or go to a different port to get his required survey in turn.

This system had been challenging already in pre-Covid-19 times, as surveyors around the globe need to be trained, with highly qualified trainers usually located in the headquarters of Hamburg and Høvik, and training demand often coming in at short notice and by very few trainees. Covid-19 brought essentially a shut-down of travel outside your own country, and temporarily also a ban on meetings and classroom training.

The latent management strategy for a digital transformation also for training turned in this situation into a pressing necessity for action. The key requirement was a rapid solution of delivering training remotely.

We had already a wide portfolio of options employed in digital training, *Bertram and Plowman (2017,2019)*. Especially the software for live online teleconferencing has taken off since the Covid-19 outbreak. Tools like Skype and Zoom boomed massively, as they addressed the need for face-to-face meetings without physical meetings, in private applications, in schools and universities, and in the business world. The existing infrastructure for 1-to-1 chats and business meetings just needed to be enhanced by a few features to address specific training needs, such as enabling small-group in (digitally) secluded environments, so-called break-out rooms. We reviewed a variety of competing option for live online training, including Microsoft Teams, Zoom, GoToTraining, Webex, and Adobe Connect. Zoom was best in functionality, but had initial cyber-security issues. Teams has initially inferior functionality, but has been adding features similar to Zoom to become a comparably useful tool. We now use Teams and Zoom in our trainings, where the choice often depends on what the trainer is familiar with.

Looking at the required development effort as a short-term priority and the saving potential compared to traditional classroom training, Fig.2, it became quickly apparent that live online training was the most obvious choice. Software solutions for developing self-paced online training had been improved as well since 2017, bringing required effort down significantly. Namely the move from Articulate's Storyline to Rise brought down development times and costs significantly, while at the same time improving trainee satisfaction with the "look and feel" of the training products.

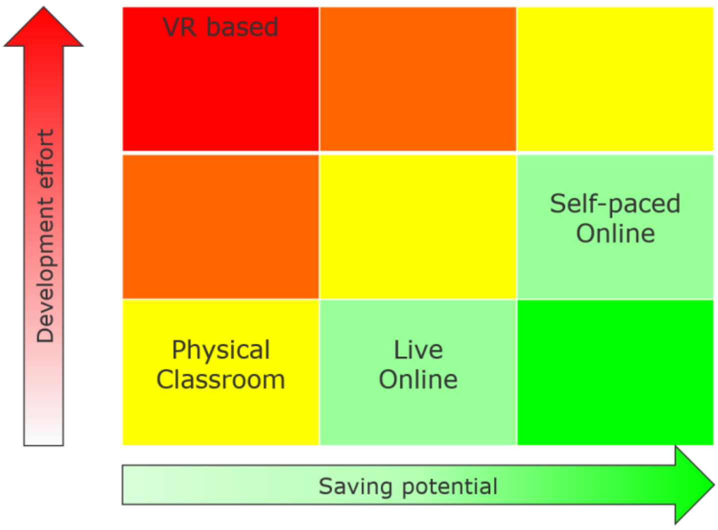


Fig.2: Live online training is an attractive option – at least as short-term option

Most trainings employ a blend of different training options. We used self-paced online training based on Rise predominantly for the following applications:

- resource libraries with reading material (pdf files or hyperlinks) and videos, often referencing to publicly available sources, such as IMO websites
- assessment tests, mainly in the form of multiple-choice tests, for self-assessment as well as formal assessment for compliance purposes
- secondary topics, such as background knowledge or historical developments

We used live online training (using Teams or Zoom) for the following purposes:

- Kick-off and closure of trainings
- Question & Answer sessions
- Short group activities
- Material that needed live commentary of trainers and was most likely to spark interactivity such as questions from trainees with fast response from trainer

The next chapter will discuss the considered choices for digital training solutions in more detail, always within the context of the Covid-19 constraints.

3. (Realistic) Options for digital training at Maritime Competence Learning & Academy

3.1. Traditional vs. Digital training

Until fairly recently, DNV GL's training experience was based on classroom courses, where frontal teaching is interspersed with various tasks to actively involve and engage the learners. Class size is usually limited to 15-20 people to allow small-group interaction, but training groups may be as small as 3-5 persons.

Over the past few years, we have responded to the increasing demand for "e-learning", which is a frequently used term by laymen expressing "something on the computer where my employees don't have to travel and sit in your classroom". The demand for such digital distance learning has skyrocketed with the outbreak of the Covid-19 pandemic.

In our work experience, we have used Lectora, Storyline Articulate and Rise as e-learning software. There is no neutral interface to export-import e-learnings from one to the other, except from Storyline to Rise. If need be, completely new programming is required to convert older Lectora trainings to the more modern Rise environment.

Table I compares traditional training with digital options. Although real world and digital world are never quite the same, we have largely equivalent options for traditional methods.

Table I: Traditional training elements with their closest digital equivalents

	Traditional	**Digital**
Reading	book, lecture notes	e-book, pdf, online reading
Frontal training	Lecture, presentation	Online lecture, webinar
Exercises	(group) exercises	(virtual) breakaway rooms
Assessment	written exam	online quiz
Social networking	Coffee breaks, etc.	Social media

Of course, virtually all training will combine elements of frontal training, exercises and assessment. And virtually all "traditional trainings" will have most of the training material in digital format, such as presentation material in PowerPoint, often with videos, handouts (reading, group work instructions, assessment) at least originally in Word or pdf.

Correspondingly, most digital training will combine various techniques to arrive at the learning objectives, e.g. self-paced studies and live online lectures and teleconferencing for face-to-face discussions.

If a traditional training is well designed using visually stimulating material with interspersed activities for the trainees, the conversion to digital equivalents is straightforward. Only the coffee breaks with real coffee and initial social bonding are vastly better in the real world…

We discuss our experience with various options in the following subchapters, updating *Plowman (2017)* with our experience of the past three years.

3.2. Self-paced online learning

3.2.1. Self-paced online courses (classical e-learning)

These days, most people think of self-paced, click-through self-paced online learning when they hear terms like "e-learning", "digital training solutions" or "digital transformation of classroom training". Alas, it is just one of our tools, albeit a powerful and useful one if properly employed.

A key risk with self-paced learning is that the trainee does not study, whether it is an old-fashioned book or an e-learning. Purely self-paced online courses generally have less impact than classroom training where individual feedback is possible and where learners generally have a higher attention rate.

Longer courses are generally subdivided into modules of typically 20-60 minutes' duration. Due to their longer duration, web courses generally employ a wider range of techniques to avoid fatigue. The training material employs techniques akin to PowerPoint presentations – text (sometimes animated), images, embedded videos.

Short courses with typical durations of 5-10 minutes are called "nano-learning". They are often employed for quick once-off instructions, e.g. when a new software is rolled out inside the company or a short safety instruction is needed and a pdf one-page instruction is ruled out, e.g. because some short video clip is needed.

In principle, all good advice for designing PowerPoint presentations for classroom training also applies to designing e-learnings. In addition, some self-paced online training software allows information on demand (e.g. mouse-over pop-up explanations, magnifying of images, links to websites or pdf documents). Information on demand allows decluttering slides with faster progress for those who don't need that information detail.

3.2.2. Videos

Videos are frequently used in e-learning, varying from small add-on to complete lecture recording. For training purposes, it is advisable to split longer videos into shorter chunks. Beyond 30-60 s, most brains start to wander.

Basic options for videos are:

- Recorded lecture: This option gives high focus on the trainer, making the perception much closer to classroom training. Blue-screen recordings of the speaker may be overlaid with slides (PowerPoint) while the trainer talks and advances slides as in traditional training, Fig.3 (left). While blue or green screens are quite cheap, you need a quiet room with proper lighting, a good camera on a tripod, good microphones, etc. for the recording. Alternatively, at significantly lower cost, the expert may self-recorded with a webcam, using standard Power-

Point features, clicking through the slides and running a natural narrative, Fig.3 (right). Typically, 2-3 takes are necessary to get a useful recording.

- Technical video: For special, usually promotional purposes, video production is outsourced. Prices depend always on content and length, but order of magnitude is 1000-3000 € per minute of video, *Bertram and Plowman (2017)*. For most training purposes, the production of such videos is prohibitively expensive. However, existing videos may be re-used, embedding or hyperlinking them. Videos on public websites (such as YouTube) are best hyperlinked. This avoids many legal issues and makes the own training material "light", i.e. reduces file size and required bandwidth for acceptable response.
- Animation video: Rather dry (technical or regulatory) material may be made more entertaining by using animated, cartoon-type videos. We use Vyond, Fig.4. As a rule of thumb, 1 minute of an animated videos costs 200-400 € to produce.

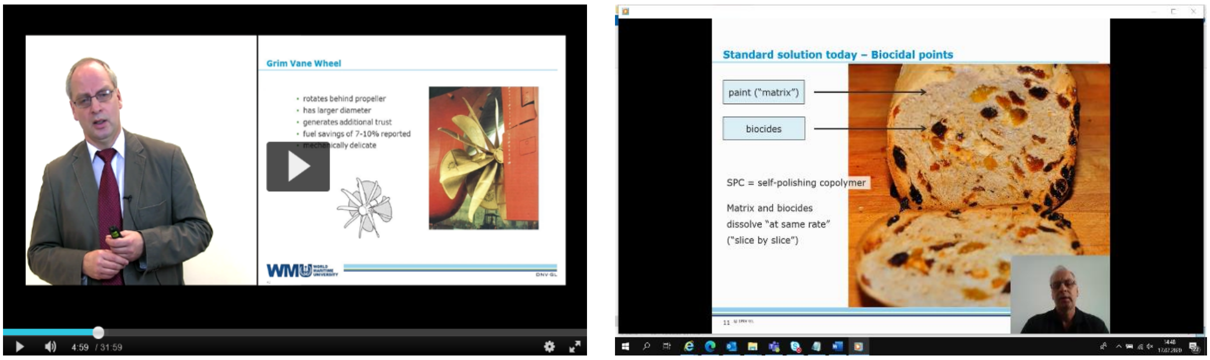


Fig.3: Video recording using blue-screen technique (left) and using ppt & webcam (right)

Fig.4: Stills from e-learning videos merging cartoon-like animations with tailored image elements

Producing new videos adds significant costs. It should thus be considered in each case whether a video is "nice to have", "important" or "essential" in the context of the learning goals. On the other hand, it is recommended to re-use existing videos (in full or in part) wherever this supports the learning goals: pay once (for the development), use many times. For video formats, wmv and mp4 seem to give the least technical problems. There are free online tools to convert older video formats to mp4.

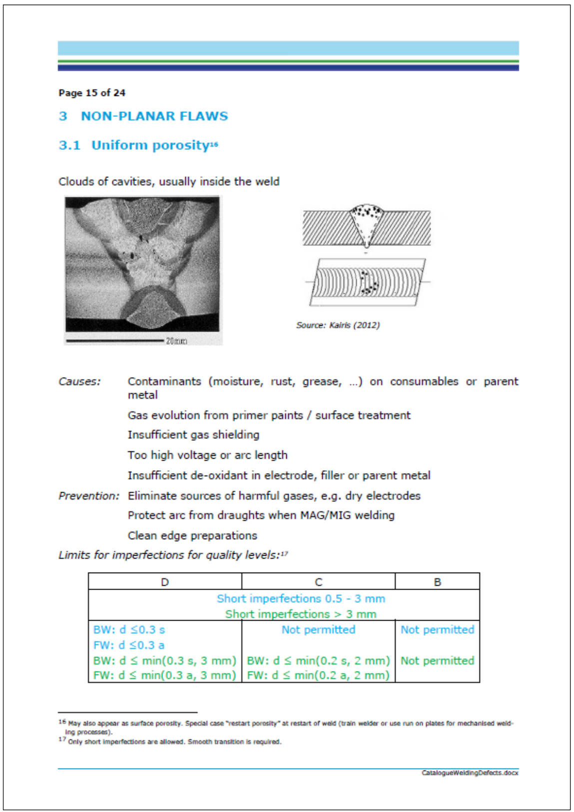
3.2.3. Online reading

While e-learnings should be strongly visual, there is usually text information that needs to be transmitted. Various options exist:

- Full text: All text information is given as full text, as pdf attachment, embedded or as links to websites. This is the easiest and cheapest way to produce; trainees do not need audio, i.e. they can use the training also in crowded areas (commuting, open-plan office, etc.) without headphones. For short trainings, like nano-learnings, this is a good option.

- Keywords on slides + audio narrative: This is akin to classroom training. However, the trainees then need headphones. Due to narrower audio bandwidth and unavoidable accent challenges in multinational settings, trainees need to concentrate harder and fatigue faster.

Our preference has moved towards "silent" reading options, based on the feedback from our multinational, mostly non-native speaker trainee customer-base.



Page 15 of 24

3 NON-PLANAR FLAWS

3.1 Uniform porosity[16]

Clouds of cavities, usually inside the weld

Causes: Contaminants (moisture, rust, grease, …) on consumables or parent metal
Gas evolution from primer paints / surface treatment
Insufficient gas shielding
Too high voltage or arc length
Insufficient de-oxidant in electrode, filler or parent metal

Prevention: Eliminate sources of harmful gases, e.g. dry electrodes
Protect arc from draughts when MAG/MIG welding
Clean edge preparations

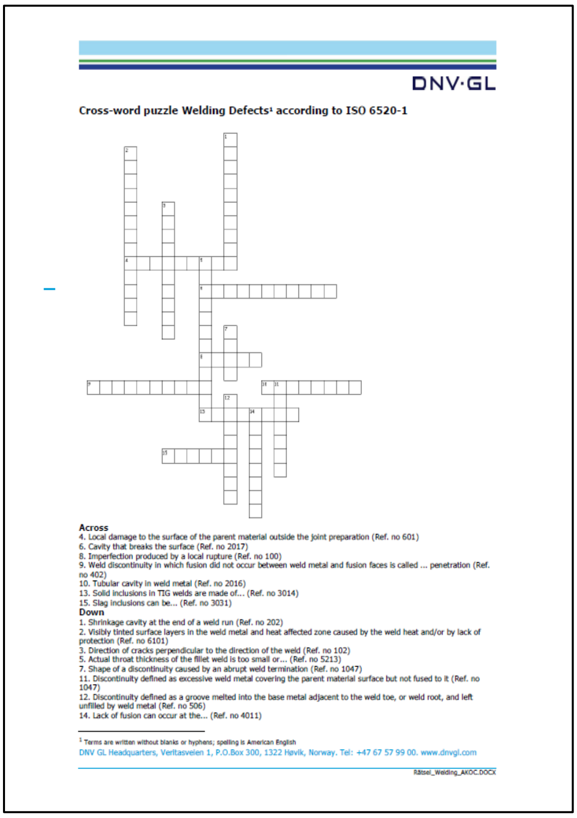
Limits for imperfections for quality levels:[17]

D	C	B
Short imperfections 0.5 - 3 mm Short imperfections > 3 mm		
BW: d ≤0.3 s FW: d ≤0.3 a	Not permitted	Not permitted
BW: d ≤ min(0.3 s, 3 mm) FW: d ≤ min(0.3 a, 3 mm)	BW: d ≤ min(0.2 s, 2 mm) FW: d ≤ min(0.2 a, 2 mm)	Not permitted

[16] May also appear as surface porosity. Special case "restart porosity" at restart of weld (train welder or use run on plates for mechanised welding processes).
[17] Only short imperfections are allowed. Smooth transition is required.

CatalogueWeldingDefects.docx

Fig.5: Typical reference knowledge in pdf



DNV·GL

Cross-word puzzle Welding Defects[1] according to ISO 6520-1

Across
4. Local damage to the surface of the parent material outside the joint preparation (Ref. no 601)
6. Cavity that breaks the surface (Ref. no 2017)
8. Imperfection produced by a local rupture (Ref. no 100)
9. Weld discontinuity in which fusion did not occur between weld metal and fusion faces is called ... penetration (Ref. no 402)
10. Tubular cavity in weld metal (Ref. no 2016)
13. Solid inclusions in TIG welds are made of... (Ref. no 3014)
15. Slag inclusions can be... (Ref. no 3031)

Down
1. Shrinkage cavity at the end of a weld run (Ref. no 202)
2. Visibly tinted surface layers in the weld metal and heat affected zone caused by the weld heat and/or by lack of protection (Ref. no 6101)
3. Direction of cracks perpendicular to the direction of the weld (Ref. no 102)
5. Actual throat thickness of the fillet weld is too small or... (Ref. no 5213)
7. Shape of a discontinuity caused by an abrupt weld termination (Ref. no 1047)
11. Discontinuity defined as excessive weld metal covering the parent material surface but not fused to it (Ref. no 1047)
12. Discontinuity defined as a groove melted into the base metal adjacent to the weld toe, or weld root, and left unfilled by weld metal (Ref. no 506)
14. Lack of fusion can occur at the... (Ref. no 4011)

[1] Terms are written without blanks or hyphens; spelling is American English
DNV GL Headquarters, Veritasveien 1, P.O.Box 300, 1322 Høvik, Norway. Tel: +47 67 57 99 00. www.dnvgl.com

Rätsel_Welding_AKOC.DOCX

Fig.6: Crossword puzzle as media break

While not glamorous, pdf files and online reading text are often a good and cost-effective option:

- Short instructions: "Sometimes it makes more sense to deliver new training content in the form of a job aid. Don't stretch out a small amount of content in order to create an hour elearning course," *Ferriman (2013)*. If you have little to say, put it on one page.
- Reference knowledge as add-on: A cardinal sin in training is 'slidumentation', *Duarte (2008)*, the mixing of slides with documentation. Much of our traditional training material contains reference knowledge. Nobody can seriously expect trainees to retain this knowledge after brief exposure: Catalogues of welding defects, Fig.5, pages of regulations applicable if A exceed this threshold and B that. All the trainee should learn is where to find that documentation and how to work with it. Transferring classroom training to digital solutions, we often include links to pdf files or websites, where the reference knowledge is found, and focus on the learning goals "I know this resource exists", "I know where to find it" and "I know how to work with it". Referenced websites should not be short-lived; few things frustrate trainees more than clicking on hyperlinks and getting error messages. Websites under your own control and rather stable links (Wikipedia, IMO regulations, ISO standards, etc.) work well, though.
- Lecture notes: Online trainings cannot be printed as traditional PowerPoint training material. However, most participants feel a need for take-home reading material for later reference in their professional life. More or less extensive lecture notes as pdf downloads from online

training are popular and make pedagogical sense. The lecture notes can be updated frequently and cost effectively, much faster and cheaper than programming an e-learning. In one extreme case, we converted an older e-learning, which was text-heavy with many technical drawings, into a pdf-attachment of lecture notes (96 pages) and a lean e-learning consisting essentially of a page for the download of the lecture notes and a quiz to ensure that trainees had studied the material.

- Instructions for activities: Pdf files may also be used for interspersed tasks or case studies. For example, after an e-learning has presented material for 20 minutes, you may attach a pdf with a crossword puzzle reflecting the presented material, Fig.2. Time for a coffee, a pen and let's crack that crossword. Such media breaks work well and generally receive positive feedback from the trainees.

Pdf files come with some inherent advantages:

- They can be downloaded and printed. We get a lot of reading down during our commuting to and from work. And we often prefer reading a paper version, where we can work with a pen or a highlighter, and where the strain of reading seems less after hours spent in front of computer screens.
- They are standard software from a major supplier, in an open format based on ISO 32000. As such, it can be expected that in decades to come, we will still be able to open and read pdf files, with free and easily available software.
- The standard reader for pdf software comes with a search function, which is particularly useful in large documents.

3.2.4. Online assessment

Often the trainees are "motivated" by a test at the end required to get the formal qualification. It may not follow a feel-good modern view on pedagogy, but it has been proven to work - and our industry is used to it. For such an assessment, there are various options:

- Ungraded quiz: The softest option: Have a quiz (usually programmed in the e-learning software) with tasks, most often multiple-choice questions, and give immediate feedback to the trainee whether the answer was correct or incorrect, possibly giving additional explanations on the correct answer. This type of quiz is intended to give just voluntary assessment to the trainee how much or little has been learnt. It also breaks the habit of just clicking 'continue' again and again.
- Graded quiz: As above, but this time there is an overall grade at the end, most often without additional details. The trainee gets the final score and whether this was enough to pass. The assessment result is entered automatically in the learning platform and possibly an e-certificate is issued and emailed or offered for download. This is our standard option for courses. If the certificate is important (e.g. a university degree, a formal license, etc.), this approach is not suitable as it is difficult to ensure the identity of the person taking the test.
- Classroom quiz after e-learning: In cases where the identity of a candidate must be checked and ensured that no external help was received, we have not found an alternative to classroom testing under supervision. The knowledge acquisition may be based on digital solutions, but the knowledge assessment in classroom makes the approach "blended learning".
- Human evaluation of free text: In some cases, the assessment may be in the form of a free text (essay). We use this option in our joined post-graduate diploma courses with World Maritime University,. This approach comes with significant time required from the subject matter expert, but allows checking more advanced reasoning and documentation skills, as e.g. needed in auditing.

3.3. Live online training

In response to the Covid-19 pandemic, we needed to develop training solutions rapidly. Often lecture series of 8-12 modules (of 1.5 h each) had to be converted from face-to-face classroom teaching to digital solutions, Fig.7.

The first task was generally a critical review of what could be rapidly converted to self-paced study, e.g. quizzes and historical background information. Such a partial conversion gave well-received breaks from online face-to-face videoconferencing, for training purposes a.k.a. “virtual classrooms”. The remaining parts, which involved strongly discussion between expert and trainees, and “off-the-record” lively comments by the expert, were kept as live online training.

Unlike business videoconferencing, the training options of e.g. Microsoft Teams and Zoom offer some additional features, such as digital breakaway rooms allowing smaller subsets of the trainee group to discuss a task in private with possibility of the trainer entering virtually each sub-group. Overall, the experience is that this type of training is more tiring, possibly due to reduced audio and visual resolution. It is recommended to reduce conventional full days in classrooms to half-days online while doubling the course duration in calendar days.

- Subject matter experts (SMEs) are much more open to online live training. Webinars are often enthusiastically embraced, as they allow direct contact to customers.
- SMEs are generally neither communication nor webinar technology experts. Presentation material (PowerPoint) generally needs reworking for online delivery.
- Most often, a second person is needed to support periphery work, e.g. monitoring chats.
- Live online training may be combined with prior or follow-up reading material, e.g. trainer and trainee introduction.
- Teaching material should be strongly visual and cut down on reading text, Fig.8.
- After 5-10 minutes speaking time, an interactive element (e.g. a “poll”, Fig.9) should stimulate the audience to refocus on the topic. Otherwise the temptation to multi-task (i.e. read incoming emails, etc.) becomes overwhelming for most people.
- While recordings can be offered after a broadcast, in our experience virtually nobody downloads these recordings.

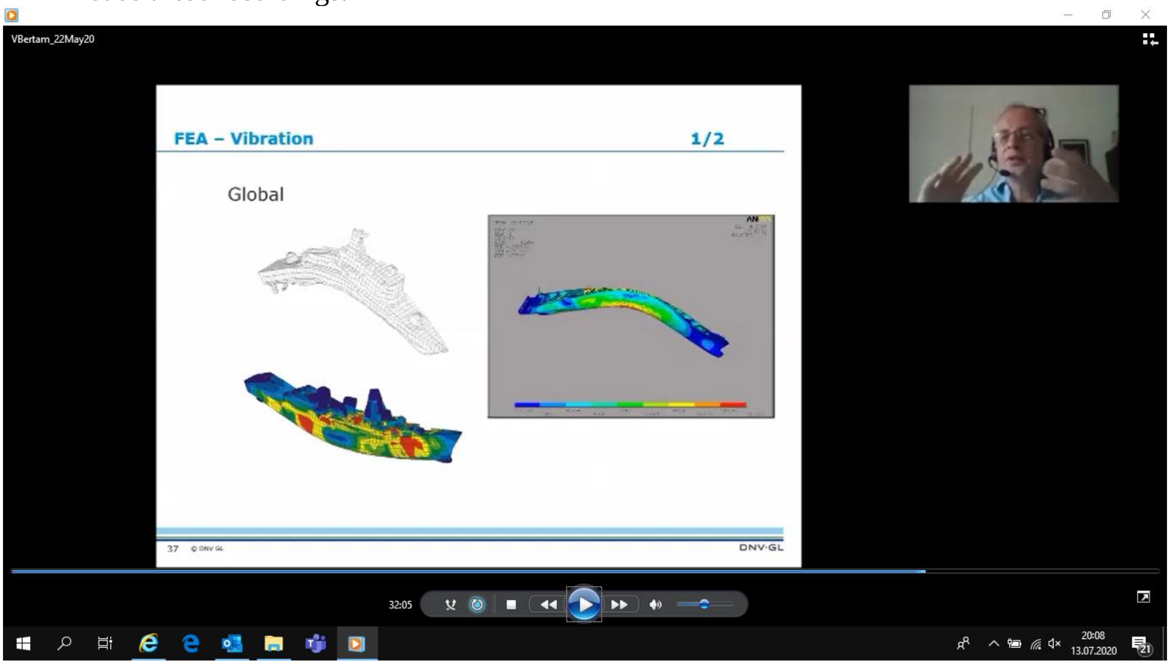


Fig.7: Volker Bertram (DNV GL) in live online lecture from Hamburg to students in Valdivia/Chile

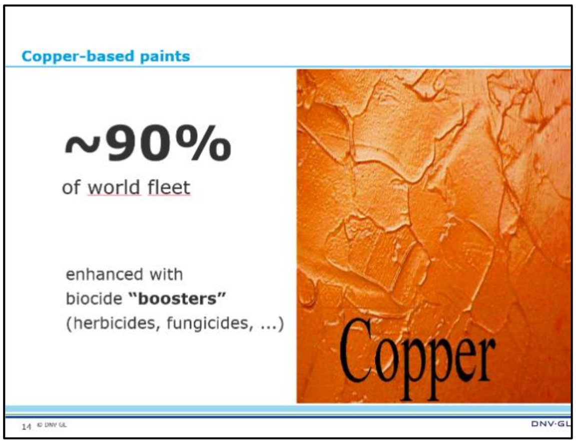


Fig.8: Typical webinar slide

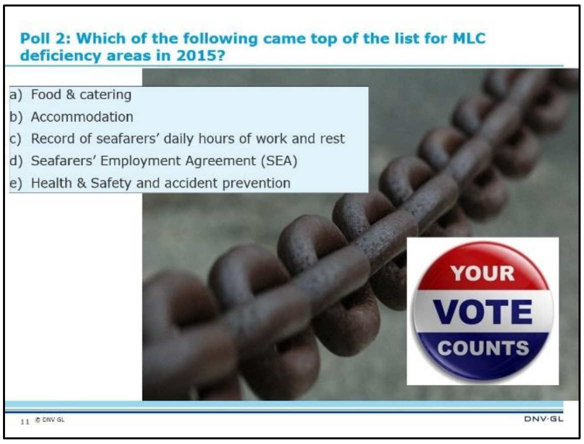


Fig.9: "Polls" stimulate audience to think

Short "conference-style" presentations of 20-30 minutes have often been converted into online webinars. As with classroom/conference presentations, there are good ones and bad ones. Bad ones are of the format "you look at PowerPoint slides while the expert drones on". Participants often zone out, doing other things like checking their emails, passively absorbing the audio and tuning back into the webinar occasionally. The good ones are relatively brief, focused on a single, tangible topic with a clear take-home message, and strong user interaction. In our line of training, we often must respond rapidly to new developments, e.g. new regulations coming into force. We have found that webinars are an attractive addition to our toolbox of training solutions in this respect. DNV GL's line of external webinars is called smart-ups.
Some lessons learnt about live online training include:

- Live online training always comes with live audio. That makes them easy to prepare and generally livelier that e-learning but introduces the accent challenge: Listening to a voice from another nationality requires more concentration that listening to someone from your own language. The first few minutes our brain tunes into to a different accent; native English speakers are often the hardest to understand for non-native speakers. Keywords should then appear on the screen to help the listeners. Also, simple vocabulary helps, where in engineering most engineering terms are considered as simple by an engineering audience.
- Presentations can be broadcast with or without webcam video of the speaker. Having an inserted window with a speaker makes a webinar more personal. Lip reading also helps with comprehension; it is good to have a close-up of the speaker for that purpose.
- Chats function are good for trainees to write down immediate questions, but disruptive for the trainer. Regular breaks need to be planned for the trainer to review questions and respond to them. Responses may be live or via a chat or email.
- Live online training software allows tracking user behavior and exporting statistics e.g. to Excel files. Information gathered includes trainee registration information, time of joining, time of leaving, attention rate (percentage of time when window with webinar was active), questions asked, what was answered in polls. Both for training and marketing purposes, such statistics can be very interesting.

Webinars have become a standard tool for many companies. The problem is that we all get flooded with emails, touting upcoming webinars. As a simple self-defense, webinar invitations often land in the spam folder. In order to avoid this fate, best use a specific invitation from a known colleague/ manager and find a title for the webinar that raises curiosity or motivation to join.

3.4. Gamification & Virtual Reality

Gamification of teaching using video game technology has attracted a lot of attention. Virtual Reality (VR) is seen as a key technology for (maritime) training, *Plowman (2017). Bertram and Plowman*

(2018). Virtual Reality is not only fascinating and fun; it is also a powerful tool for training, especially when it comes to visual assessment and human interaction, e.g. judging when to initiate action in maneuvering, crane operation, etc.

However, the price of developing VR-based training is high. Models need to have the right level of detail, balancing realism and response time. Import/export from CAD systems or other models (e.g. finite-element models) may save time, but is never as straightforward as hoped for or promised by vendors. VR-based training is generally well received by trainees and effective; but it comes at high costs. Having a ship modelled over several decks, along with equipment, interactivity, training tasks and solutions, assessment, etc. may take 100,000 to 1,000,000 €. Such an investment needs either subsidizing from R&D projects, opportunistic recycling of available, suitable models or a mass market willing to pay premium fees for training, such as firefighting.

DNV GL has developed SuSi (Survey Simulator), a VR-based training solution for ship inspections, *Bertram and Plowman (2020)*. SuSi provides realistic and cost-efficient 3D training software for survey inspections, using Virtual Reality technology and detailed models of ships and offshore structures, Fig.10. The virtual inspection gets trainees exposed to deficiencies that would take years for a surveyor to experience in real life. An inspection run can be recorded and discussed in a debriefing with an experienced supervisor/trainer, pointing out oversights and errors by the trainee.

Fig.10: Level of detail in DNV GL's SuSi (Virtual Reality based survey simulator)

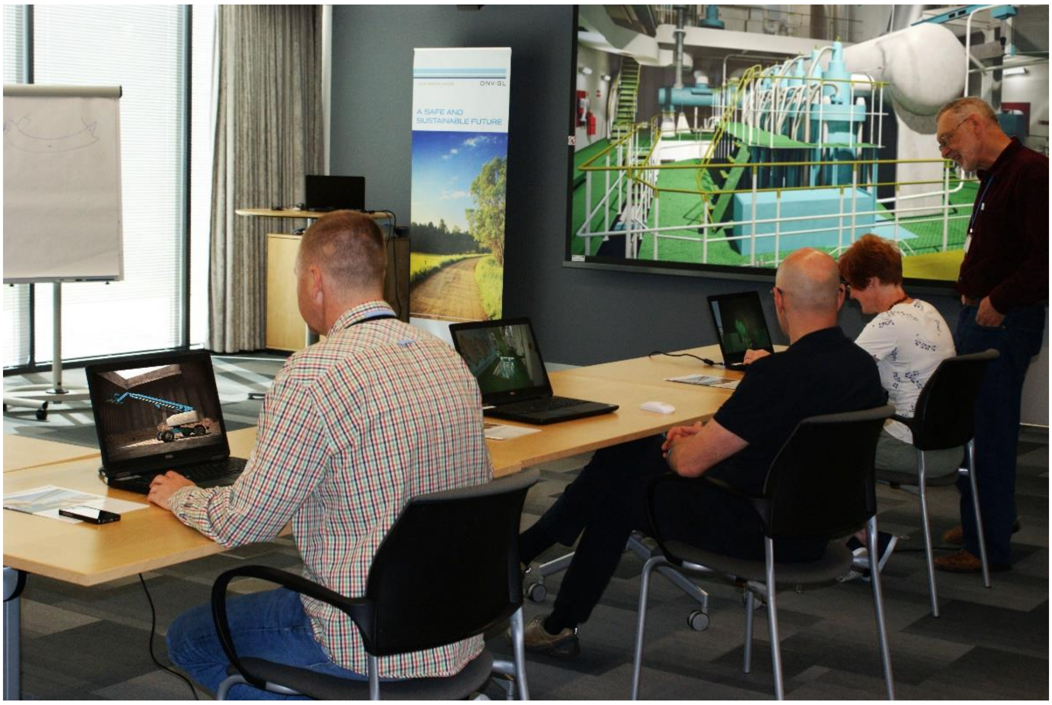


Fig.11: VR-based training works better with trainer supervision, *Bertram et al. (2020)*

Besides the cost issue, there are other aspects to consider with VR-based training:

- Cyber-sickness, with symptoms akin to motion sickness, may occur, especially if using head-mounted displays,
- Trainee group coherence may be lost due to varying IT savviness. Much is intuitive for video gamers, nothing for digital immigrants aged 50+.

VR-based training does not seem suitable for self-paced learning without support.

A pragmatic approach where the trainer guides the class collectively through the ship (e.g. with a single PC and a data projector, or a shared screen) and trainees interrupt when they spot a deficiency has been well received by participants from industry across a wide range of nationalities (cultures), educational backgrounds, management levels and age groups, Fig.11.

3.5. Social media

Training (from school over university to life-long training) always has a social aspect, making friends and meeting them again. Overlooking this aspect with a tunnel vision on "we them to answer these questions correctly" would be a mistake.

As response to an invitation to a webinar, we received the following reply: "I shall not register to the webinar, whatever the topic, for the simple reason that I do not see the point… no networking, no coffee, no time out of the office." This echoes a widely shared feeling: people miss the exchange of experience, the maritime gossip, the networking. Can social media step in and help? In DNV GL, we employ two technologies (besides video conferences which were covered above) to "reach out and stay in touch" within groups of people interested in a certain theme:

- Yammer, a social networking service for private communication within organizations. For training purposes, it is thus limited to internal training. There are mixed feelings about using a platform like Yammer as add-on in digital training solutions:
  - Use Yammer (like any other social media platform) selectively. As a posting platform for occasional nuggets of information ("I was recently at a conference on our theme and the proceedings can be found here") or specific questions ("Has anybody any experience with…?"), it works well in small and coherent groups.
  - It works better for younger groups than for older groups.
  - The more Yammer groups you subscribe to, the more messages pop up. Soon people react in mental self-defence and no longer open any of them.

  In some cases, only time will tell whether a social media channel like Yammer works for its intended training purpose or not.
- Email: Emails work well for pre- and post-training contact if they are concise and relevant. You can for example send short personal presentations of trainer and trainees before a training to establish a rapport and identify special areas of interest. Or a trainer may email a special publication in response to a question during the training. Email may also work as an electronic hotline; often specific questions come some time after a training, when trainees have to solve a specific problem in their line of work. If it so specific that it is of little interest to the rest of the trainees, individual emails work better than a Yammer.

4. Conclusions - Combine and conquer

You can achieve Death by Powerpoint, i.e. boring training based on PowerPoint. But trainees may die many deaths, i.e. irrespective of platform, you can be boring: in classroom presenting, using pinboards and flipcharts, talking in videoconferences or accumulating a multitude of videos.

Make it relevant, make it short, make it fun – in traditional training as in digital training. And most of the time, variety adds to making it fun and tailoring learning objectives to the training platform. No media is per se evil, and no media is per se perfect. Combine (training options) and conquer (the hearts of the trainees).

The Covid-19 situation has forced us to adopt digital training options, whether we wanted it or not. Due to time pressure, not all options worked well. And other things worked surprisingly well; one trainer perceived more audience focus on the training than in classroom training. We learnt some lessons and, for sure, we are not at the end of our exploration of the digital universe of teaching. But we are convinced the after Covid-19, we will not simply return to the pre Covid-19 modus operandi.

The virus has changed our private world, our business world, and also our training world. Deal with it!

Acknowledgements

We thank my colleagues of the developer team in the MCLA (Maritime Competence Learning & Academy) department for their support and help over the years in shaping my skills and my view on training in the specific maritime world, namely Ulrich, Bernie, Torsten, and Margit.

References

BERTRAM, V.; PLOWMAN, T. (2017), *Maritime training in the 21st century*, 16th Conf. Computer and IT Applications in the Maritime Industries (COMPIT), Cardiff, pp.8-17, http://data.hiper-conf.info/compit2017_cardiff.pdf

BERTRAM, V.; PLOWMAN, T. (2019), *A Hitchhiker's Guide to the Galaxy of Maritime e-Learning*, 18th Conf. Computer and IT Applications in the Maritime Industries (COMPIT), Tullamore, pp.7-23, http://data.hiper-conf.info/compit2019_tullamore.pdf

BERTRAM, V.; PLOWMAN, T. (2018), *Virtual Reality for maritime training – A survey*, 17th COMPIT Conf., Pavone, pp.7-21, http://data.hiper-conf.info/compit2018_pavone.pdf

BERTRAM, V.; PLOWMAN, T.; FEINER, P. (2020), *Survey simulator – A Virtual Reality training tool for ship surveys*, 19th Conf. Computer and IT Applications in the Maritime Industries (COMPIT), Pontignano, pp.7-15, http://data.hiper-conf.info/compit2020_pontignano.pdf

DUARTE, N. (2008), *Slide:ology: The Art and Science of Creating Great Presentations*, O'Reilly & Assoc.

FERRIMAN, J. (2013), *9 things people hate about elearning*, LearnDash, https://www.learndash.com/9-things-people-hate-about-elearning/

PLOWMAN, T. (2017), *Maritime e-Training – Matching Requirements to Solutions*, 11th Symp. High-Performance Marine Vehicles (HIPER), Zevenwacht, pp.55-65, http://data.hiper-conf.info/Hiper2017_Zevenwacht.pdf

An Integrated Platform for the Development of Autonomous and Remote-Control Ships

Xu You, Wuhan University of Technology, Wuhan/China, youxu2016@whut.edu.cn
Feng Ma, Wuhan University of Technology, Wuhan/China, martin7wind@whut.edu.cn
Suli Lu, Wuhan University of Technology, Wuhan/China, sulilu@whut.edu.cn
Jialun Liu, Wuhan University of Technology, Wuhan/China, jialunliu@whut.edu.cn
Xinping Yan, Wuhan University of Technology, Wuhan/China, xpyan@whut.edu.cn

Abstract

This paper describes an integrated platform for the development of autonomous and remote-control ships. The platform takes a 7 m KVLCC2 model, which is the well-known tanker form for ship design and research community, as the target ship. Ship hydrodynamics are studied through Computational Fluid Dynamics and publicly available towing tank tests. Accordingly, mathematical models are built and validated through field tests. With an understanding of maneuvering and resistance performance, an onboard autonomous system is implemented with docking, navigation, and undocking functions. A communication system is set up with mobile and public internet for remote control at a low cost.

1. Introduction

In recent years, considerations on economic expenditure and environment protection attract research direction to intelligent, remote-control, and autonomous ships. Universities, institutes, and companies have performed large quantities of research works through simulation studies and experiments with ship models. In 2018, the International Maritime Organization (IMO) put forward the concept and classification of Maritime Autonomous Surface Ships (MASS), which further promotes the development of autonomous ships.

To meet the rapid development of intelligent ships, China Classification Society has improved the framework of Rules for Intelligent Ships (2015) and puts forward Rules for Intelligent Ships, *CCS (2020)*. Based on the direction of the local intelligent application to the whole ship intelligent application, and auxiliary decision-making to fully autonomous decision-making, the Rules for Intelligent Ships add rules of the remote control and autonomous control to standardize the intelligent ship specification framework and corresponding functional/technical requirements.

To fulfill these requirements and follow the framework of MASS and Rules for Intelligent Ships (2020), we have built up an integrated platform for the development of autonomous and remote-control ships under the concept of "Navigation Brain System", *Yan et. al. (2019,2020)*. This platform uses a 7 m KVLCC2 ship model as a research object to verify remote control and autonomous control for cargo transportation tasks. To simulate the real scenarios as much as possible, the testbed is chosen in Tangxun Lake and Qinhuai River. Apart from traditional devices like GPS. etc., this platform equips Lidar, communication unit, and PC with software and algorithms to perceive more accurate environment information, make prompt decisions, and give specific control orders to implement navigation tasks autonomously. According to the performance in field experiments, this platform shows the stability and safety in remote control and autonomous control experiments.

2. Ship hydrodynamics and mathematical modeling

2.1. Ship particulars and test environment

The target ship, **Error! Reference source not found.**, is a 7 m KVLCC2 model, which is illustrious and is easy to get a comparison with known data to check the fidelity of the maneuvering model. Table I shows the main particularities of this ship, *Yasukawa and Yoshimura (2015)*, http://www.simman2019.kr/. The hull of this ship is made up of glass fiber. The service speed is 5 knots according to its scale. As the main propulsion motor is cooled with water, the water of the stern cabin can form a free surface, which harms the maneuverability. To overcome this shortage, this ship carries about 2.4 tons to make itself in light load with the stern tilt to promote the performance in control. The fore draft is 19 cm and the stern draft is 32 cm. To meet the environment protection, this ship uses a pure electric motor for propulsion and step motor for the rudder. This model equips bow and stern thrusters, which are helpful for docking and undocking. With all devices for navigation purposes, including perception, propulsion, and steering, the total power is 700 Watt under the service speed on average.

Fig.1: 7 m KVLCC2 model in the field

Table I: Main particulars of the KVLCC2 model

Main Particulars	
L_{WL}	7.12 m
B	1.27 m
T	0.34 m
CB	0.81
Rudder	
Type	Spade with a NACA 0018 profile
Average height	0.35 m
Average chord	0.19 m
Total area	0.07 m^2
Propeller	
No. of blades	3
D	0.22 m
P/D at 0.7R	0.72
Service speed	WUT
U	1.00-1.08 m/s
Fr	0.12-0.13

2.2. Ship hydrodynamic analysis and motion modeling

Ship hydrodynamics on the KVLCC2 model has attracted broad interests. Moreover, a large amount of data is proposed to check the correctness and accuracy. Comparing to the data of Yasukawa, MARIN, and SIMMAN, *Yasukawa and Yoshimura (2015)*, we used CFD calculation on this ship model to analyze the ship hydrodynamic parameters, Fig.2. Combining with field experiments in Tangxun Lake, the ship maneuvering motion model was built to verify the accuracy comparing to the SIMMAN's experiment data.

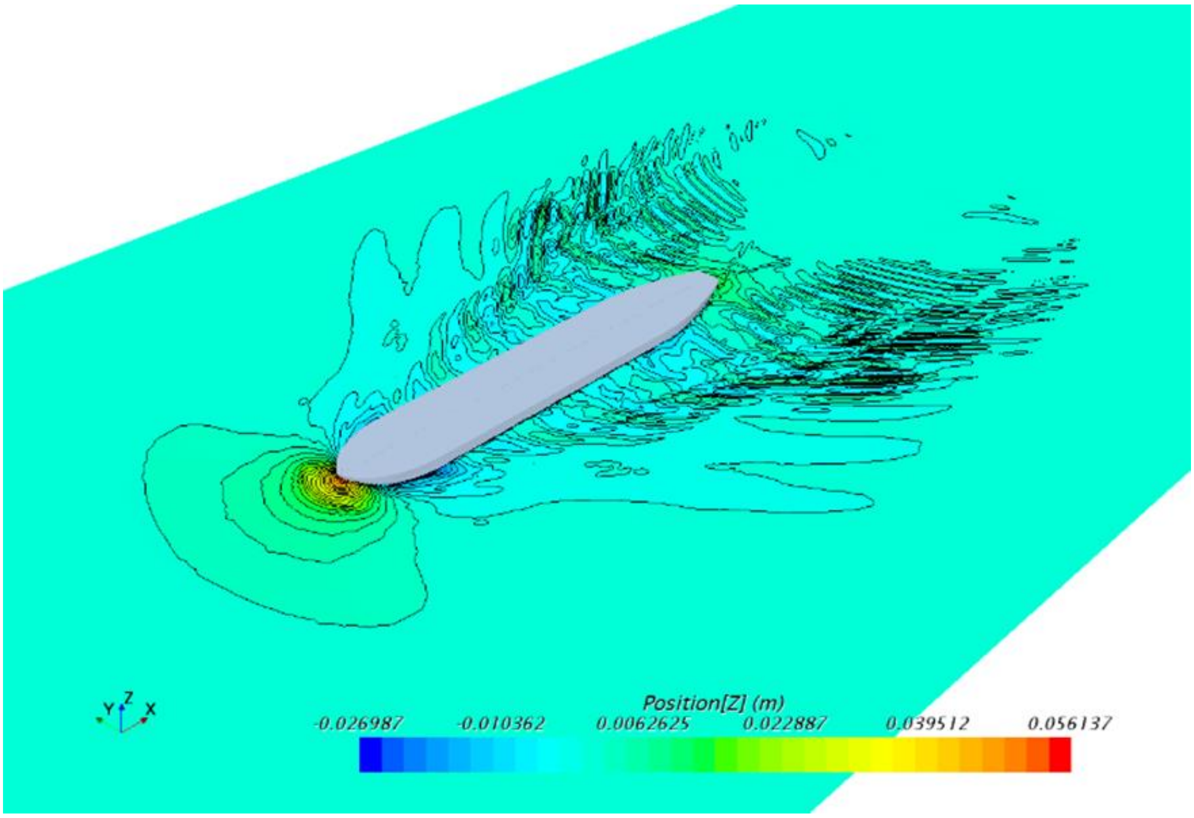


Fig.2: CFD calculation on the KVLCC2 model

Fig.3: The testbed in Tangxun lake

Tangxun Lake is one of the lakes in Wuhan, which is used as the first testbed is selected. As shown in **Error! Reference source not found.**, the area of the testbed is a square with a side length of 0.6 km, which is about 0.36 km^2. In this area, the average water depth is about 1.5 m, which is deep enough to ignore shallow water effects. To verify the maneuverability of this ship, the preliminary tests, including zig-zag tests and turning tests, were carried out in this lake.

Fig.4 presents the experimental results for the turning tests in right and left full rudder comparing to the simulation results. Fig.5 shows the results for the 20/20 zig-zag tests and simulation results comparing with validation using MARIN data from SIMMAN 2020 and field experiments carried out by WUT. The maneuvering model shows good agreement with the towing tank results from MARIN while it has a certain deviation from the actual field tests. The differences may cause by a lack of measurements and calculations of wind and wind-induced waves effects.

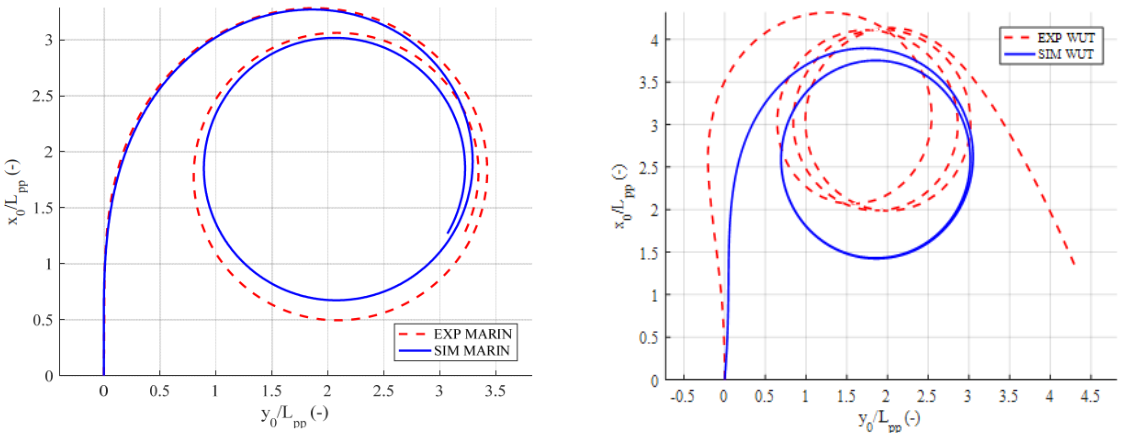


Fig.4: 35° turning maneuver with simulations comparing using MARIN data from SIMMAN 2020 and field experiments carried out by WUT

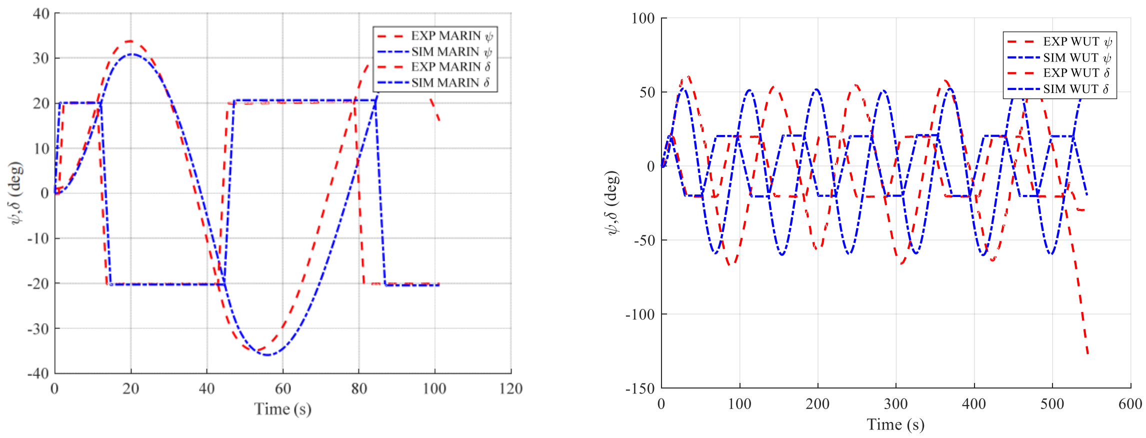


Fig.5: 20/-20 zig-zag maneuver with simulations using MARIN data from SIMMAN 2020 and field experiments carried out by WUT.

3. Design of the integrated platform

3.1. Hardware structure

As shown in Fig.6, the sensors and devices equipped in this ship can be divided into seven parts:

- Sensors
  Sensors include differential GPS, Lidar, Camera, IMU, and Bow sensor, which could perceive information from both ship and surroundings to construct a risk situation map during

navigation.

- Communication
  Communication includes AIS (160MHz antenna), Long Range (LoRa) (433MHz antenna), and 4G router. These communication methods can cover communication requirements at different distances.
- Power supply
  The power supply is based on lithium batteries and solar panels. In strong sunlight, the solar panels supply 800 W, enough to operate the ship without using the batteries.
- Rudder system
  The rudder system uses a step motor with a connecting rod structure to ensure the control accuracy of the rudder.
- Propulsion system
  The propulsion system has a brushless DC motor for main propulsion and two lateral thrusters for docking.
- Decision-making system
  The decision-making system consists of a mini-PC and programmable logic controller (PLC). It is in charge of collecting data from sensors and feedback from the rudder and motors, then making suitable decisions like humans to generate control orders.
- Control station.
  To improve safety and fault tolerance while navigating, this ship still keeps a control station for operators in case of any emergency.

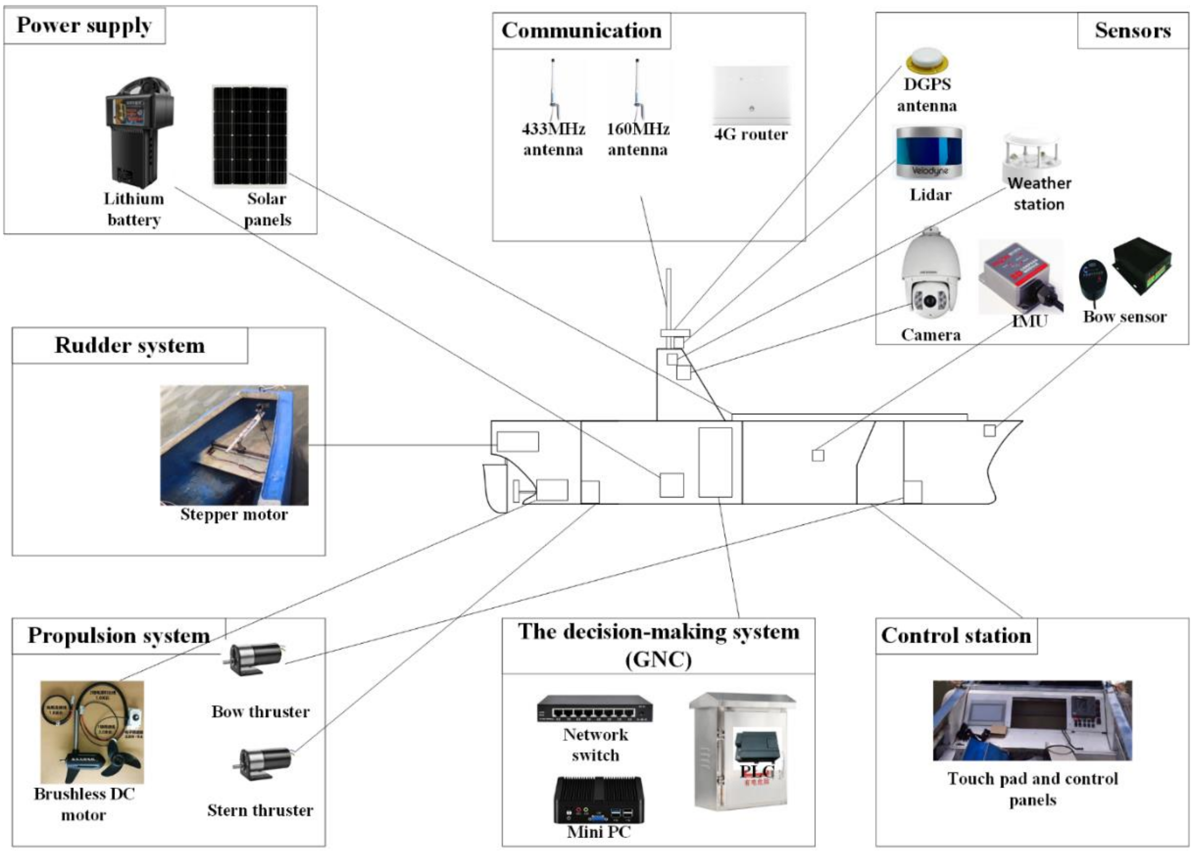


Fig.6: The hardware devices and structure of 7 m KVLCC2 ship in this platform

Besides the devices on ship, a ground station is constructed to remotely control the model. As shown in Fig.7, this station is equipped with a communication unit, PC, displays, and console with the control button and engine telegraph that combines the control function of rudder and speed. The data collected from the remote ship is delivered to the ground station through LoRa and 4G signals based on different distances and communication environment. LoRa can only transmit limited data, which is localized in core information like ship status and obstacle coordinate. The ground station can use this information to reconstruct the environment in maps. However, the information is limited, not reflecting the real environment completely. The communication unit shows the benefits of communication quality guaranteed by a local network in a limited distance, which is not disturbed by public communication signal range requirements. These features make it suitable for delivering information without public communication devices. Comparing to LoRa, the 4G signal has a wide bandwidth for video data and Lidar data, besides the core information mentioned above. The main disadvantage is the need for public communication equipment support.

Fig.7: Ground station structure for remote-control operations

Considering navigation in real situations, this platform mainly focuses on inland river navigation, where public communication cannot achieve full coverage. To guarantee communication capability, combining the two communication methods, and designing different strategies is advisable. Based on these two communication methods, the ship's operator type can be switched to remote manual control or remote order control. In the right part of the ground station, operators can use engine telegraph with buttons to remotely control the ship in manual mode.

3.2. Software structure

The software of this platform is mainly divided into two parts:

- ground station, and
- ship station.

The software on ship contains information collection, decision making, and order implementation. Considering the hardware structure and stability, the parts of information collection and order implementation use C++ and C language to negotiate with sensors and actuators (rudder and propulsion) through MCU and PLC. As for the decision-making part, Python is the primary language to build a data procession structure with advanced algorithms. To better show the ship status and navigation task, a display window is created with PyQt5 to help operators monitor and control test ship.

In this design, the navigation task and experiment are divided into several modes, including Local control, Remote control, Local auto, Remote auto, Simulation, Real-time control, etc. Fig.8 shows the display of Local control; the top of the window shows the battery on-off of different devices and sensors, which can prevent electronic trouble to some degree. The left sidebar shows the shipping status of GPS, ship speed, heading angle, ship attitude, and control order value. The right region is a basic map showing the edge of the shore, bridge, routes, and ship location. Fig. **1** shows the 2D situation of complete information comparing to the 3D situation in Fig.7, by which operators can both give their control commands for order control and manual control.

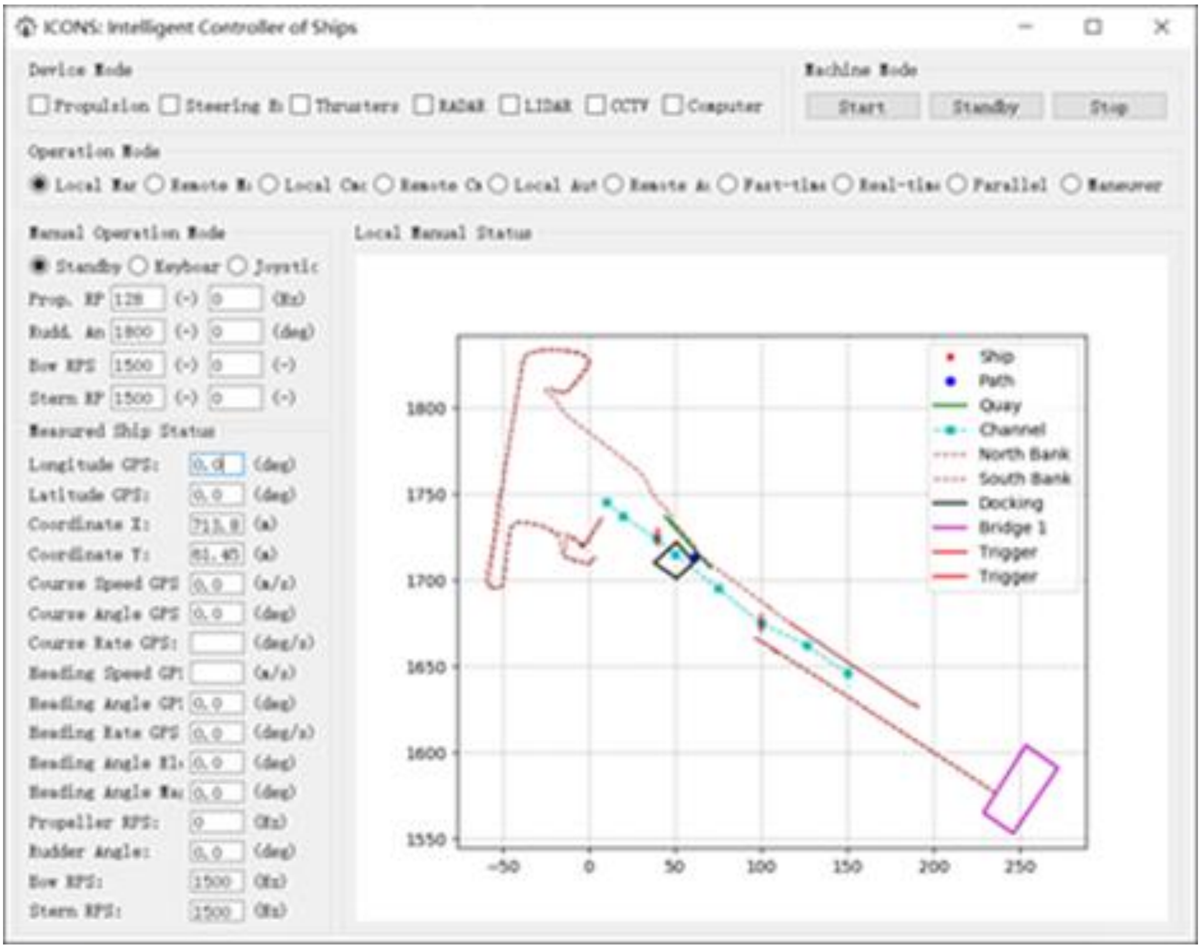


Fig. 1: The human-machine interaction face for ship status monitor and control

4. Demonstrations of remote-control and autonomous operations

4.1. Remote control experiments in Tangxun Lake

In Tangxun lake, the preliminary tests and adjustments are carried out to adjust this ship to a suitable performance. Then several trajectory tracking and path following experiments are implemented based on LOS and PID algorithms. From these tests, 7 m KVLCC2's performance shows the ability to follow design routes effectively and rapidly. Fig. **2** is a Google satellite map capturing the test moment in Tangxun lake. Fig.3 shows the interface of the remote control system. Considering the communication methods and previous experiments on ship control, remote control tests are implemented though remote manual control method, remote order control method, and switch process. We have verified the effectiveness of remote control in Beijing, Shanghai, and Nanjing with a distance of about 2500 km in 2018. In October 2019, with the help of Delft University of Technology and MARIN, a demonstration was performed in Wageningen, The Netherlands with more than 8500 km away.

Fig. 2: The experiment in Tangxun Lake on Google satellite map

As mentioned before, remote control modes can be divided into remote manual control and remote order control. These two control methods may need a different level of information. The remote manual control needs more detailed information leading to high pressure on communication bandwidth, which requires a 4G or more efficient transmission method. If the data transmission is limited, using a digital twin may help to construct a 3D environment to overcome these disadvantages. The remote order control uses navigation tasks to control remote ship, which means the detailed rudder angle and speed value is translated by the local computer in a remote ship when certain heading angles or path points are given. These two methods are both necessary and can complement in different situations.

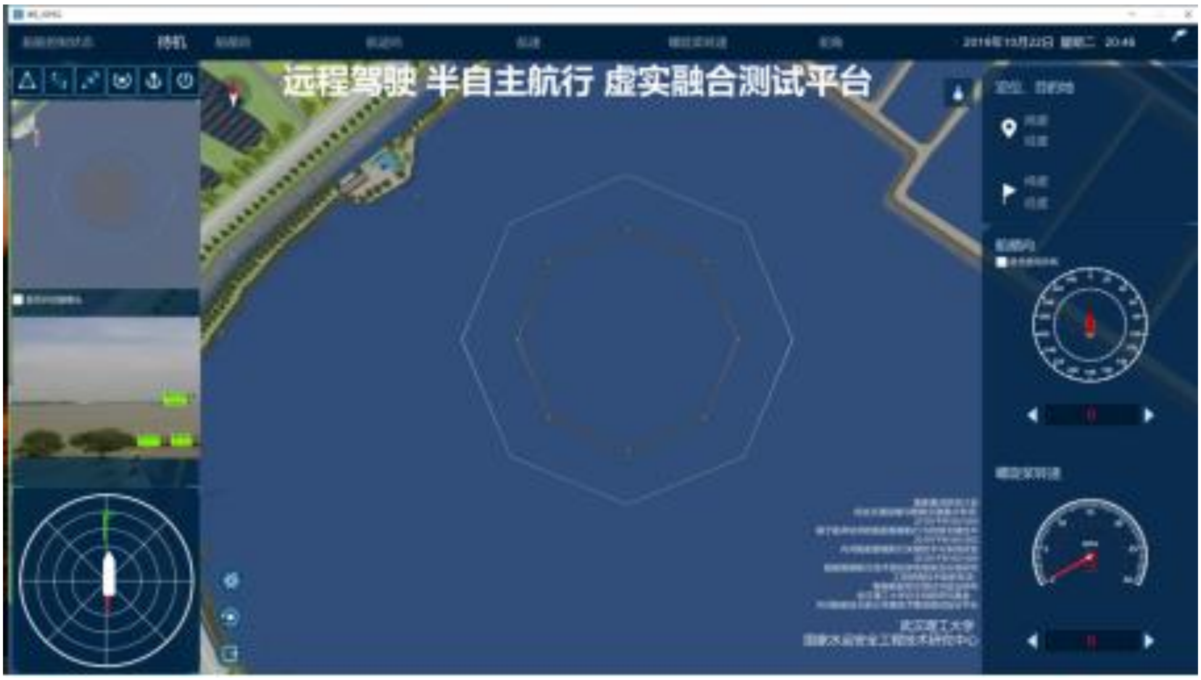


Fig.3: Remote control system display

Ship remote control is considered a necessary process from traditional ship to autonomous or smart ship. It can test the communication ability and intelligence level on ship control, which gives experience and technical support for later development on the autonomous ship. In terms of economic effect, the remote control may cut the number of crews and promote driving safety.

4.2. Perceptual information fusion and digital map construction

Besides the autonomous navigation tests in the Qinhua river, we use GPS data, Lidar, and CCTV to enhance the precision of river environments. Fusing this information, the construction of 3D and 2D digital maps decreases the demands of communication requirements. Fig.11 presents the Lidar and video acquisition process. Comparing to the real situation, Fig.12 shows the digital map in 3D, which contains riverbank coordinates, bridge coordinates, and river depth. It also shows the moving boat according to detective sensors, which can promote the precision ability and decision-making ability for collision avoidance.

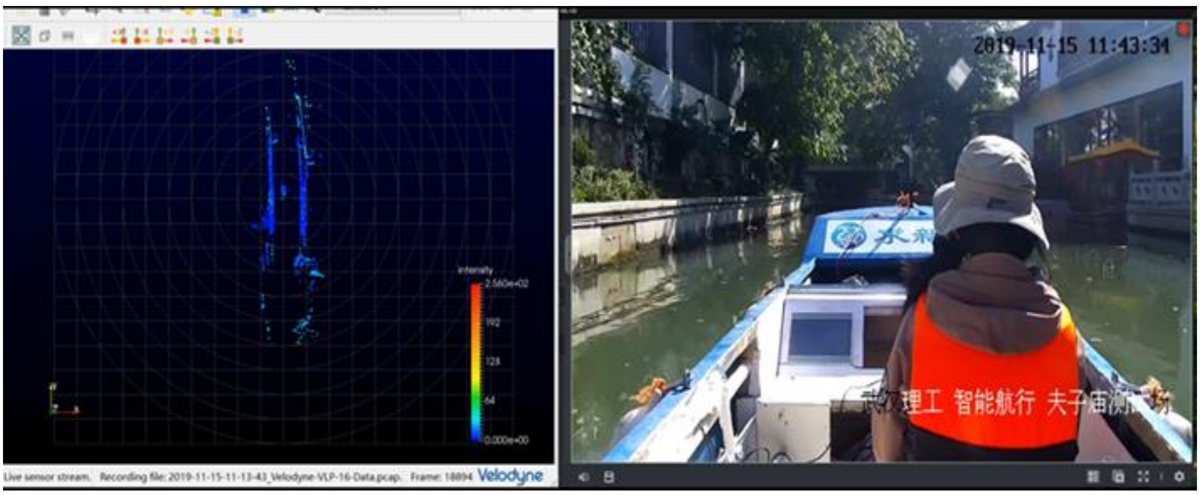


Fig.11: Environment information collecting by Lidar and video in real scenarios

Fig.12: Digital 3D scenarios in the Qinhuai River

4.3. Autonomous control experiments in Qinhuai River

To further research and verify the autonomous navigation ability of this platform, the testbed is changed into the Qinhuai River. In Fig. **4**, the white line presents the whole test route from Xishuiguan (left start) to Dongshuiguan (right end), which is about 5 km long. The main test area is located in Xishuiguan, as shown in the red region. After the GPS measurement and 3D reconstruction in a digital map, the river environment can be explained as shown in **Error! Reference source not found.**.

Fig. 4: The testbed in Qinhuai river

The whole autonomous navigation test is divided into three parts: autonomous docking and undocking, autonomous navigation, and autonomous turning around. Furthermore, docking and undocking area, navigation route, turning around the area are designed at the first step. Fig.5 presents the procedure of docking, and these five pictures demonstrate how this ship leaves the riverbank and starts to follow a designated route. The preparation work is adjusting the bow thruster and stern thruster to ensure them providing the same thrust so that ship can leave the riverside smoothly. The speed of docking and undocking is less than 0.2 m/s. After the model reaches the designated area, the heading angle is adjusted by the lateral thrusters, and then thrusters are off to keep the ship stationary. After that, the main motor and rudder begin to work. Autonomous navigation is implemented by following the path until reaching the turning area.

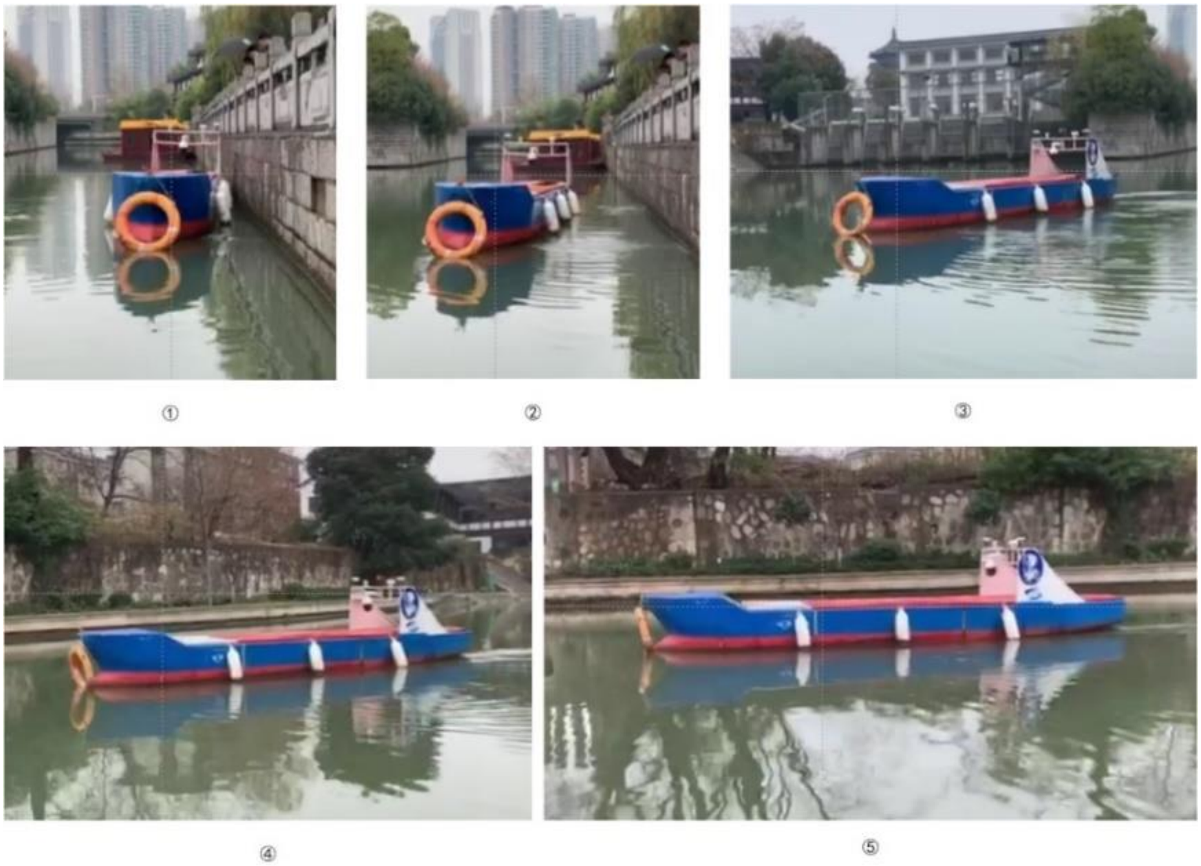


Fig.5: Undocking procedure of 7m KVLCC2 ship in autonomous navigation test

Fig. **6** presents the situation of how the ship turns around. Like the steps in undocking, KVLCC2 decreases the speed to almost stationary and turns on the thrusters in the opposite direction. This ship could have a small turning circle in the narrow water area. Pictures from one to nine show this procedure. On the way back, the control steps are reversed. This test shows the effectiveness and possibility of autonomous navigation in inland rivers and reveals how ships implement docking and undocking techniques.

Furthermore, it can be extended to other rivers and ships in an ideal situation. However, there are still some issues that need to be solved. This river has a stable environment for wind and current, which means the disturbances are small to influence this ship. For inland rivers, bridges and other shelters like high buildings can block GPS signals, which does harm locating ability and controlling ability. Future tests will focus on these issues.

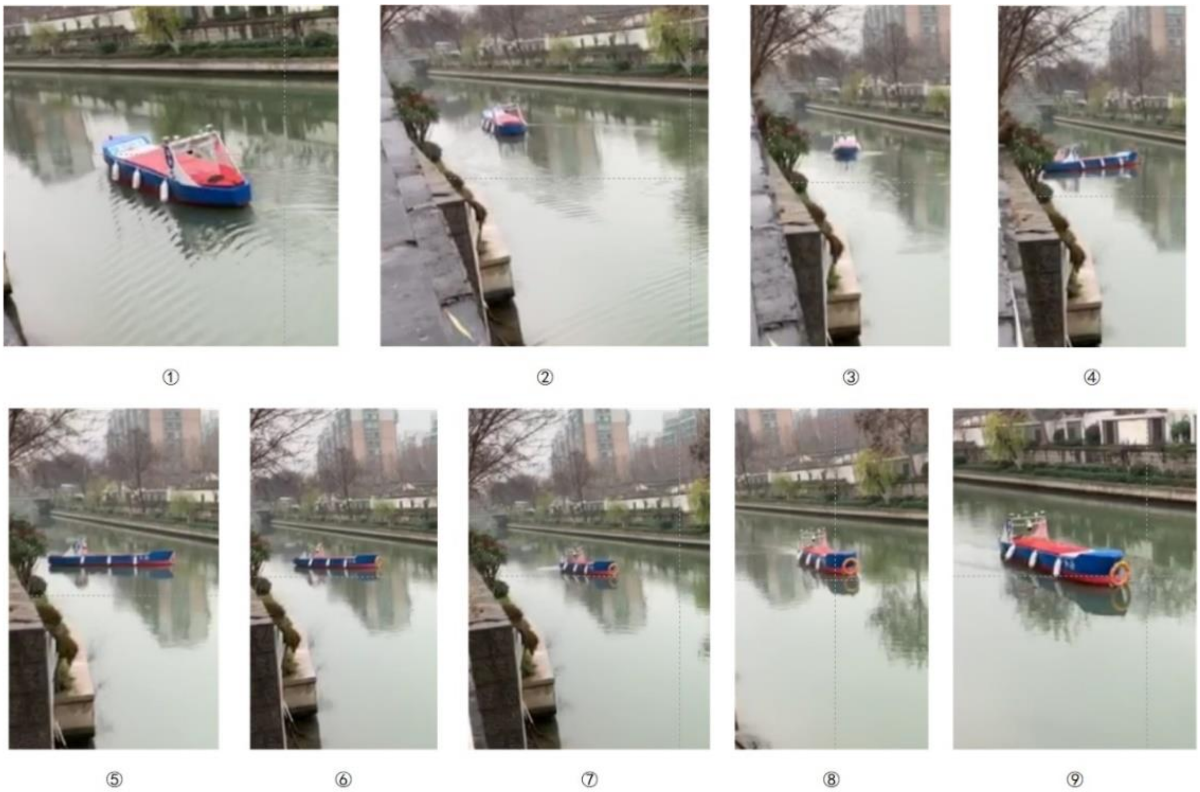


Fig. 6: Turning procedure of 7m KVLCC2 ship in autonomous navigation test.

5. Conclusion

This paper introduces an integrated platform for remote control and autonomous control in two real scenarios based on a KVLCC2 model. Depending on the combination of software and hardware equipment, the remote control is realized in Tangxun lake through 4G and LoRa communications with autonomous control in the Qinhuai River. Given the current technology condition, remote control is most likely to realize in real scenarios. Combined with the current needs for economic factors, remote control on ferries, and later inland ships may be one of the most accessible applications to implement. In the future, autonomous navigation considering disturbances and collision avoidance will be further studied. The intelligence level on the autonomous ship will be promoted step by step in this direction.

Acknowledgment

This work was supported in part by the National Key R&D Program of China (2018YFB1601503 & 2018YFB1601505), Research on Intelligent Ship Testing and Verification (2018473), National Natural Science Foundation of China (51709217).

References

CCS (2020), *Rules for Intelligent Ships,* China Classification Society

YAN, X.P.; LIU, J.L.; MA, F.; WANG, X.M.. (2019), *Applying the Navigation Brain System to inland ferries*, 18th COMPIT Conf., Tullamore, pp.156-162

YAN, X.P.; LIU, J.L.; FAN, A.L.; MA, F.; LI, C. (2020), *Development and future perspectives of intelligent ships,* Ship Engineering 42(03), pp.15-20

YASUKAWA, H.; YOSHIMURA, Y. (2015), *Introduction of MMG standard method for ship maneuvering predictions,* J. Marine Science and Technology 20/1, pp.37-52

Platform-of-Platforms and Symbiotic Digital Twins:
Real Life Practical Examples and a Vision for the Future

Nick Danese, NDAR, Antibes/France, nick@ndar.com
Denis Morais, SSI, Victoria/Canada, denis.morais@SSI-corporate.com

Abstract

The platform-of-platforms paradigm remains a genetic condition to start enacting the Digital Transformation by leveraging the multiple digital twins that compose "the" Digital Twin in the parallel-processing yet overall asynchronous process underlying automated data & information availability to all. This approach is virtuously inscribed in any PLM macro-process environment. Funded projects aimed at various tool integration goals are underway, and first "open architecture", commercially available solutions have become available in 2019. Some are based solely on existing software, others required development of "universal" digital twin connectivity (a blessing in disguise). Research in Platform-of-Platforms solutions available today will be reviewed and a vision of the near- and longer-term future presented.

1. Introduction

In fairness to those who unknowingly invented Digital Twins it is a duty to state: Computers, ergo Digital Twins. "Cogito, ergo sum" could have been tough of as a non-sequitur in the still theocratic enough days of Descartes, but that very real Digital Twins remained undefined as such for decades can be surprising at a time when quantum physics is already "old stuff".

Leaping ahead to 2020, we can take stock of 35 years of PLM and see how little exists today in way of a truly relational/associative work environment, how complex it is to set one up and, upstream, how little people understand what it would be useful for in the first place. (One of the first recorded applications of PLM was in 1985 by American Motors Corporation who were looking for a way to speed up the product development process of the Jeep Grand Cherokee. The first step was the using CAD tools with the primary objective of increasing the productivity level of the draughtsmen - see www.concurrent-engineering.co.uk.) Having said that, hats-off to those who have tried so hard to build such environments and, to varying extents, somewhat succeeded in the face of undeniable, objective hardware, software and cultural limitations.

On the good side, the evolution of software, industry and market conditions as well as the cultural relation to computing have fed a very different IT ecosystem than that present when computers were first really put to the task some 60 years ago: the agnostic combination of data sets have spurred the advent of the finally underway platform of platforms era.

2. Platform of platforms: more than just technology

Very loosely defined, a platform is a piece of a process: a computer, a program, a person, another process, etc. In this respect, attention is still today paid almost exclusively to technology but, looking beyond the "Product" in the Product Life Cycle one finds people. People create products, consume products: people are the life of products and, hence, the distinct strands composing the DNA of all the processes associated with products. It can be said then said that people have been the major cause of the birth of the platform of platforms environment and, symbiotically, the capitalist economy and consumerism becoming the way of life were a major vector thereof.

Without being too picky on semantics, both computers and software tools are platforms, although it can be argued that different cultures, disciplines, data exploitation scenarios, etc. are platforms, too.

A simple example of the above is, simply, the computer itself: in order for a software product to exist it must run on a computer, hence the computer manufacturer provides a platform that hosts other platforms and, arguably, vice versa.

Looking at the software tool itself, it generally serves platforms extending way beyond other software: for example, while an internet page source is the same for everyone, the pop-up will be tailored to the individual that has logged on to that PC, perhaps even contextually with the contents o the web page being displayed. In this case, the advertising engine simply exploits the existence of the browser, the former existing only thanks to the existence of the latter and the latter supporting the existence of the former.

Finally, the person looking at the screen will make a decision: buy something, move to another page, turn the PC off, etc. That action makes the person a platform, i.e. an active component of the process. In short, although still much unrecognised today, authors and consumers have forever been one and the same: each platform supports other platforms, all platforms together create the ensuing process.

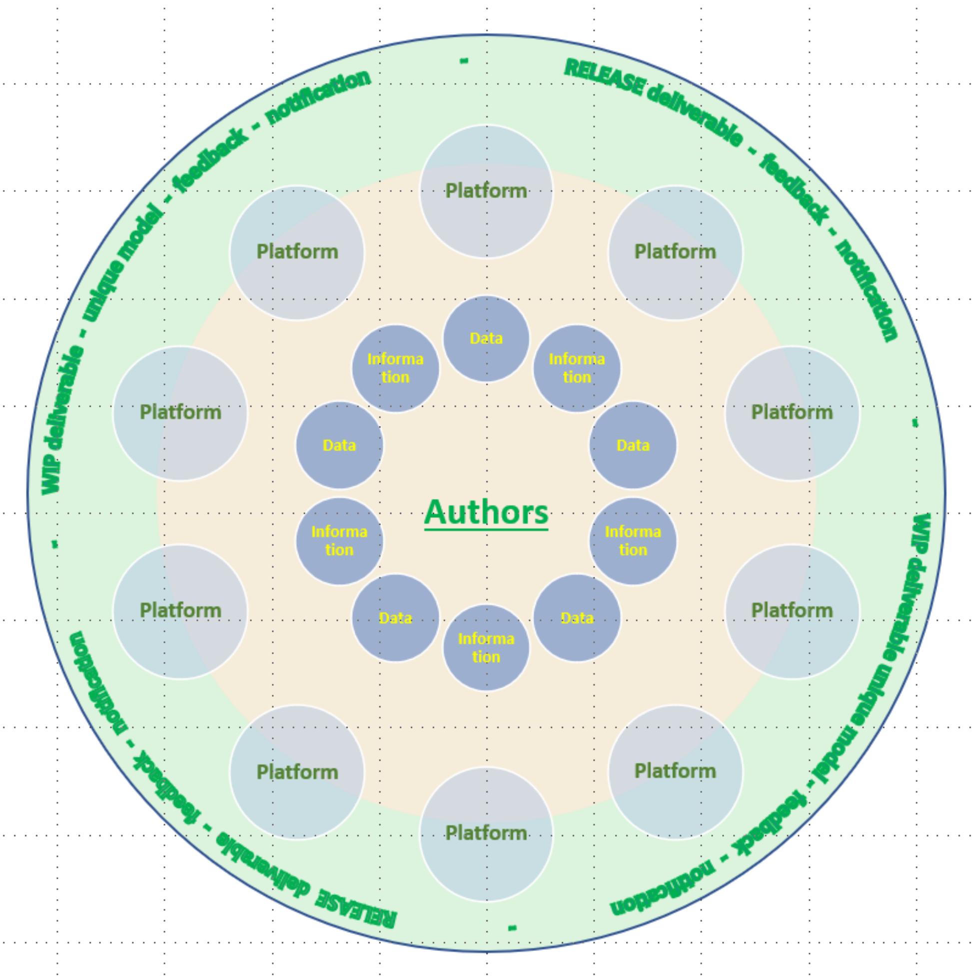


Fig.1: Authors are at the centre of the process space defined by the Platform of Platforms.

Not the topic of this paper but it is worth noting that the above is already seen in the nature of the good old subroutines present since the very early days of programming and some which working together despite being written in different languages.

On the philosophical side of technology, one could extrapolate and relate the deep meaning of the Ancient Greek meta-data and today's AI aim at making computers think: it may be wise to heed Descartes "cogito, ergo sum". Such relations lead to say that Digital Transformation, Platform of Platforms, Digital Twins are more than "just technology".

3. The distributed, collaborative environment at the source of the Digital Twin

A timeless reality, the fragmentation of the design and production processes into several time and scope limited segments constitute in themselves a significant communications problem, compounded by the fact that the many of the different software tools in use today rest on unyielding proprietary data structures, many of which are metadata-poor.

3.1. The distributed environment

The environment has always been one of multiple authors, individually and separately generating all sorts of interlaced data using multiple software programs, data formats and based on multiple platforms. This leads to multiple instances of data and information which can very easily become a genetic, self-immune disease of every process, and create noise and confusion throughout the framework of the design & production space. The unavoidable incompleteness, skew and subjectivity of each instance adds to the uncertainty, which people attempt to palliate with repeated data transmission effectively increasing confusion. Errors are the ensuing corollary of that and force re-work, which itself causes a domino-cascade effect of resource reassignment and compounded losses. The farther along the way an error occurs, the greater the damage, and all in the absence of a guarantee that all direct and indirect issues will eventually be resolved at all.

On the bright side, a distributed environment has the immense potential of being able to collect and enact all relevant contributions to a given process in virtually unlimited fashion. One such example is crowd engineering, where the final consumer is the author of the product itself. Many large consumer product companies, but also many software companies, have experimented with crowd engineering in recent years, and several commercial web platforms have been set up for this.

The on-going challenge is to coordinate the inherent fragmentation of the distributed environment by managing the platforms, their work, data and information and to make everything available, not just accessible, a job that PLM platforms attempt to carry out since 1985, *Danese and Pagliuca (2019).*

3.2. The collaborative environment

Another pillar of a successful process is constituted by the correct and timely use not just relevant but actually correct data, information and processes by the concerned authors. While much processing can already today be accomplished automatically by software programmed for the task, some processes or portions thereof are carried out by people. In all cases, a certain level of discernment must be applied at all levels in every process: some can be provided by AI and some must be contributed by correctly informed people. Despite the existence of as countless as ineffective if not detrimental "checking" strategies, it remains a challenge to be correctly informed. Still too often people don't even know whether they are informed correctly, or not, and have no means of knowing it, let alone improving their level of information.

In order to be of use, data and information have to be put at the disposal of the concerned authors in a timely fashion and in a format suitable for its consequent use. Then, where the data is, which format it is stored in, when was it stored, how was it generated, who provided it, etc. become fundamental parameters that define the usefulness of the data in question for the prescribed purpose.

That Digital Twins are created and maintained in asynchronous fashion only exacerbates the need to coordinate the team. Adaptive communications, intrinsic in the LEAN and AGILE paradigms, become paramount and a genetic component of any successful effort, first and foremost by reinjecting data and information produced by the consumption thereof - e.g. by the authors - into the process' information stream via, inter alia, the feedback loop that is key to managing change.

Implementation of the required information process indicating whether some data is available, or decision is made, or not, is another on-going challenge to be addressed by PLM.

3.3 The Digital Twin: A collection of interconnected Unique Models

Unique Models (essentially unique data sets) compound the power and effectiveness of making data, information and goods available. A Unique Model is composed of data being made available to a given recipient and specifically so in its contents, formatting, timing, etc. Unique Models are nothing else but Digital Twins themselves, made up of pieces of other Digital Twins. This elevates the concept of a Digital Twin above the notion of data, information, author, source, etc. In fact, and perhaps even more important:

- A true Digital Twin may even contain, for example, "equivalent" data sets from different authors required to perform similar tasks at different levels of sophistication: for example, a coarse mesh FE model for concept validation and definition of scantlings and a fine mesh FE model for subsequent local forced response vibration analysis. The two models are different, not interchangeable, but equivalent in that they represent the same ship.
- On a more meta-data yet digital level, the true Digital Twin may include in its in the "discernment" space opposite opinions bearing on, for example, the ship's performance required for the mission profile.

Another example of the absolute value of Digital Twins in what remains a "relative" reality is Computational Fluid Dynamics (CFD), a discipline that makes use of separate models to achieve the same goal and which sees different software and people use the same data set to arrive at very different results and conclusions.

What is relevant is therefore that the "master" Digital Twin includes all the useful data, information, metadata, etc. and that this is regrouped in coherent "subset" Digital Twins made available to the concerned authors in a timely fashion and appropriate format.

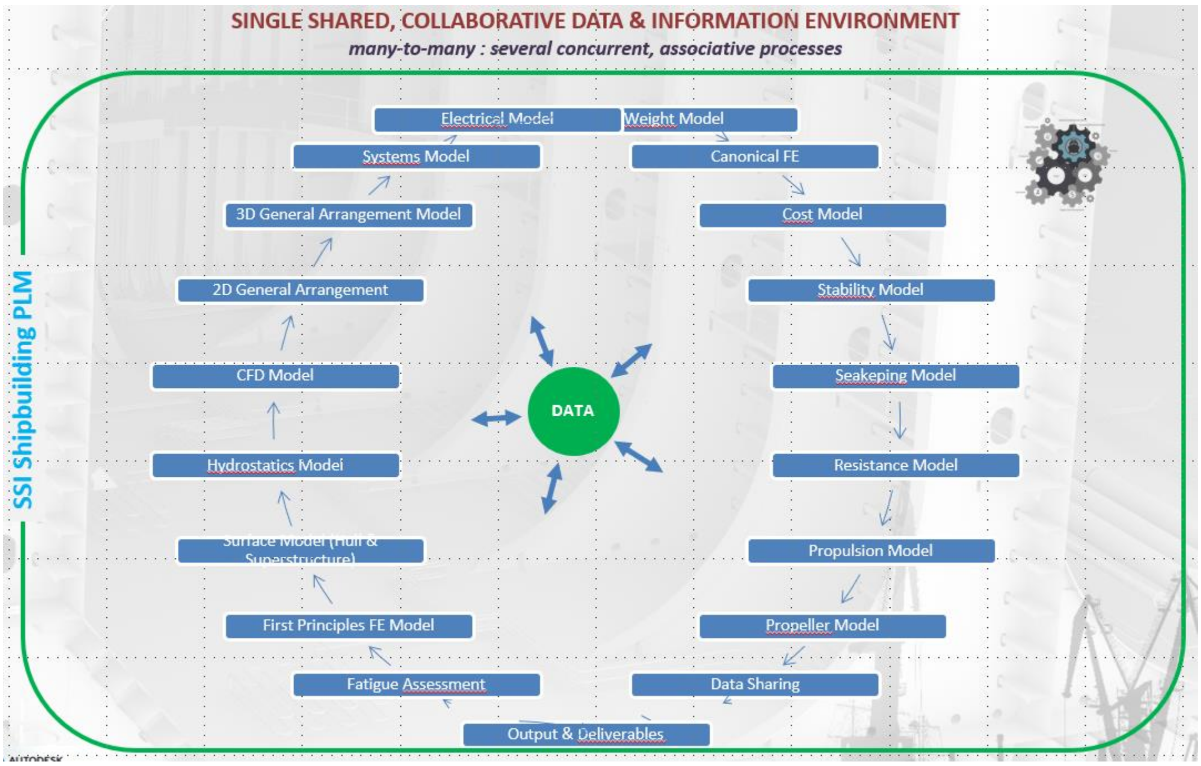


Fig.2: A very incomplete list of "source" Digital Twins making up the "master" Digital Twin

This is something that the original sequential PLM was not very strong at, but a definite possibility in the presence of today's improved understanding of many-to-many processes and the support provided by very performant hardware and adaptive relational/associative software. The quasi-science fiction neural networks of the 1980s now live through the still infant AI applications of today's advertising and will continually increase the ubiquity of the Digital Twin's value. It may be argued that AI is still rich in rational algorithms and not yet very creative: for example, think of the advertising that continues to pop-up for an item that was purchased once, months ago, never before, never since.

4. The authors of the Digital Twin

For simplicity, rather than listing authors specifically and attempting to identify individual people, processes, software, etc., the authors at the source of the ship design and shipbuilding Digital Twin space are represented here as "activities" and their life-period documented within the duration of a Product Life Cycle.

It is immediately evident that the scope of application of the Digital Twin is from t=0 to scrapping and that operational and refit data will enter the picture, too.

On the other hand, still today many show surprise at some activities being present during the complete life cycle, or just about, for example: Sales, Production Engineering, etc.

Interestingly and very instructive, dressing the same chart but for software, people and processes provides a very complementary vision of which of them are active when. The cross-connecting of activities, processes, software, people, etc. can go on and on in various directions and to finer and finer detail. The result is the expected, pervasive many-to-many, heavily parallel workspace. Although somewhat "proprietary", the software products taken as an example on this occasion illustrate well how some unsuspected ones are required and should be used at unexpected times during the life cycle of the product, exploiting the multi-author Digital Twin to the benefit of the overall process, that is the entwined financial bottom lines of shipyard and ship owner.

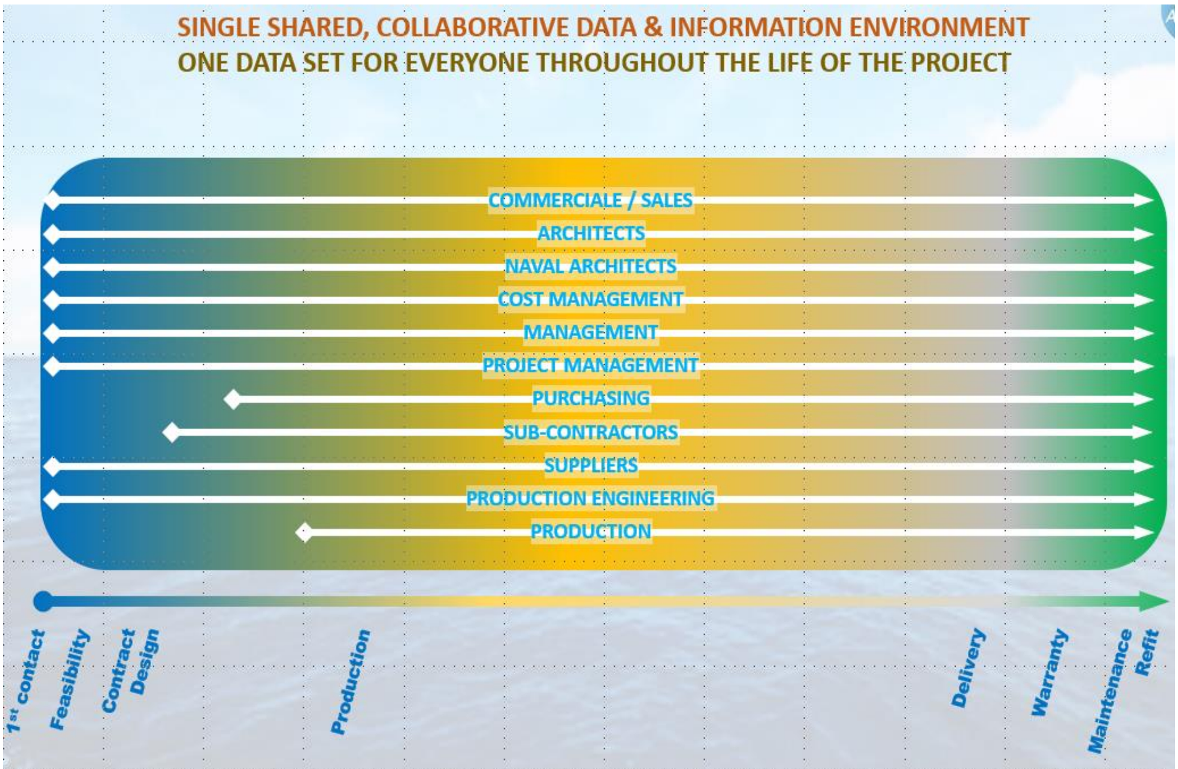


Fig.3: Mapping who does what, when, and requires exactly what input

Of course, any and every software tools present on the market is a potential author, must be considered and its used evaluated versus its applicability and measurable contribution to the Digital

Twin and its exploitation in the context of the open architecture Platform of Platforms ecosystem advocated here.

In addition to software being an author per-se - for example, let us consider the automatic re-run of a stability analysis that changes scope to include different Class criteria as a function of some non-geometric input such as mission profile, number of passengers, etc. - it will be vital that people-authors have available and know how to use the software that will allow them to contribute constructively to the Digital Twin.

This requirement could be perceived as difficult to satisfy when looking outside engineering. A telling example is that of a sales person who must be able to compose a "custom" vessel based on a "series" platform in the presence of the customer and document feasibility - is the VCG too high, or the weight in excess of the allowed maximum displacement, etc. - and price while reviewing configuration options, all without calling the office and typically finding that the whereabouts of the one person who could have answered the question are unknown.

In fact, the ubiquity not just of data but also of tools and processes, be they authors or communication engines, is probably the major driving factor of the successful Digital Twin and exploitation thereof.

The above leads to the notion that a Platform of Platforms is not, cannot and must not be an absolute notion. On the contrary, it is relative to and tailored by its purpose.

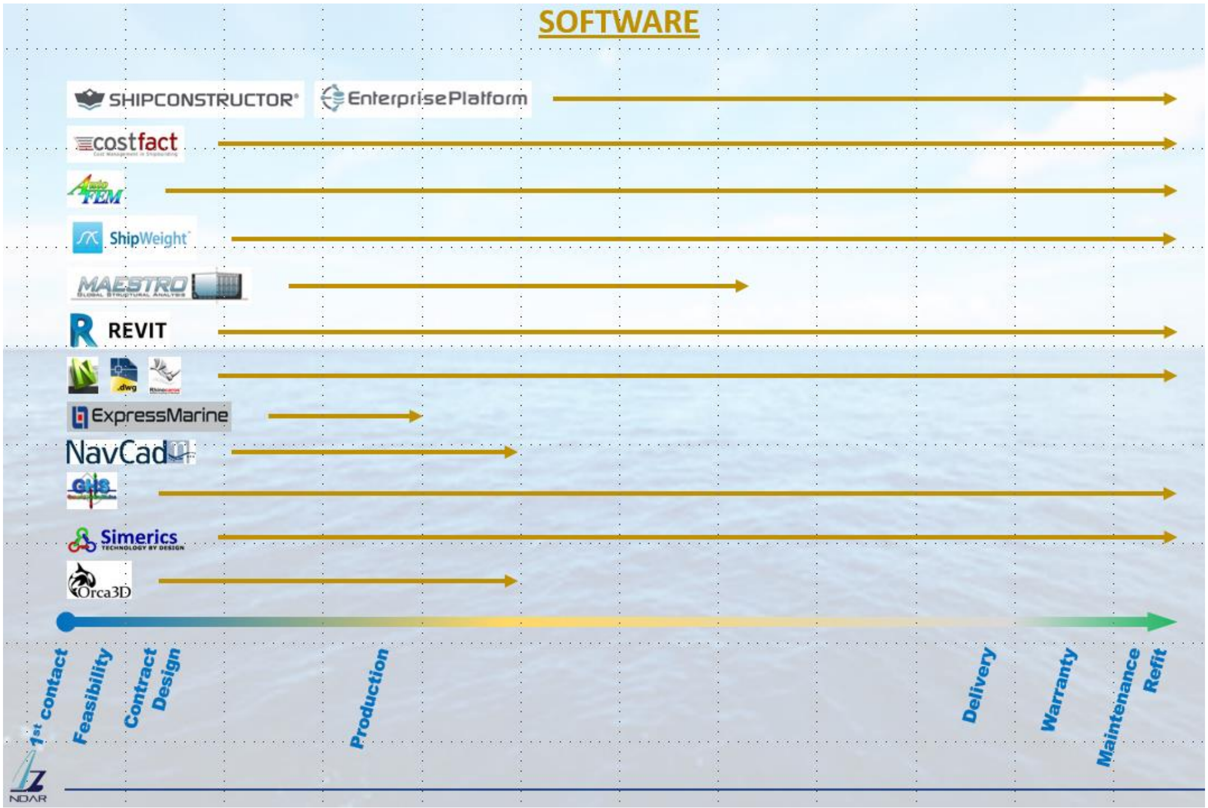


Fig.4: Mapping the applicability and use of specialist software tools during the Product Life Cycle

5. Digital Twins: The good, the bad and the ugly

From a functional standpoint Digital Twins are good or bad, there really is no in-between, and there is no such thing as a Good Universal Digital Twin Template in existence. It is a fact that every sizeable every Digital Twin around today is indeed incomplete, that is the ugly, and it could be argued that an impeccable but incomplete Digital Twin is possibly an acceptable not-so-ugly after all half-way. How ugly is ugly will be determined by the true level and amount of risk posed by the Digital Twin condition of incompletion: that will move the pointer towards good or bad, a very difficult situation when that pointer is in fact a subjective evaluation.

From a Taoist, Yin-Yang perspective and making do already with what one has, it will be constructive and productive to fall back to the individual "sub" Digital Twins that make up the Digital Twin: once the goal mapped clearly, it will be better to be 100% sure of 80% of something rather than 80% sure of 100% of the same. This will be further expanded upon in a later section, but another underlying sine-qua-non condition must be satisfied for a Digital Twin to be valid in the least, that is the existence of a managed information stream.

6. Information streams: Work-In-Progress-Real-Time vs Release

It is safe to say that there is no functional Digital Twin without some information flow, and there are at least two major information streams underlying a useful Digital Twin:

- Work-In-Progress-Real-Time
- Release

The Work-In-Progress-Real-Time information streams flows all the time and is just that: everyone feeds their respective WIP work to the Digital Twin space, concerned authors are notified, a feedback loop is provided to channel their reaction back to the Digital Twin space, etc. All those concerned see what others are doing in certain areas of the Digital Twin. The Work-In-Progress nature of the Digital Twin is instrumental in spurring notes, comments, etc. and in spawning checks, what-if studies, and the earliest possible troubleshooting. The inherent an asynchronous multi-platform Digital Twin can live, let alone thrive, only and only in the presence of an efficient Work-In-Progress-Real-Time information stream.

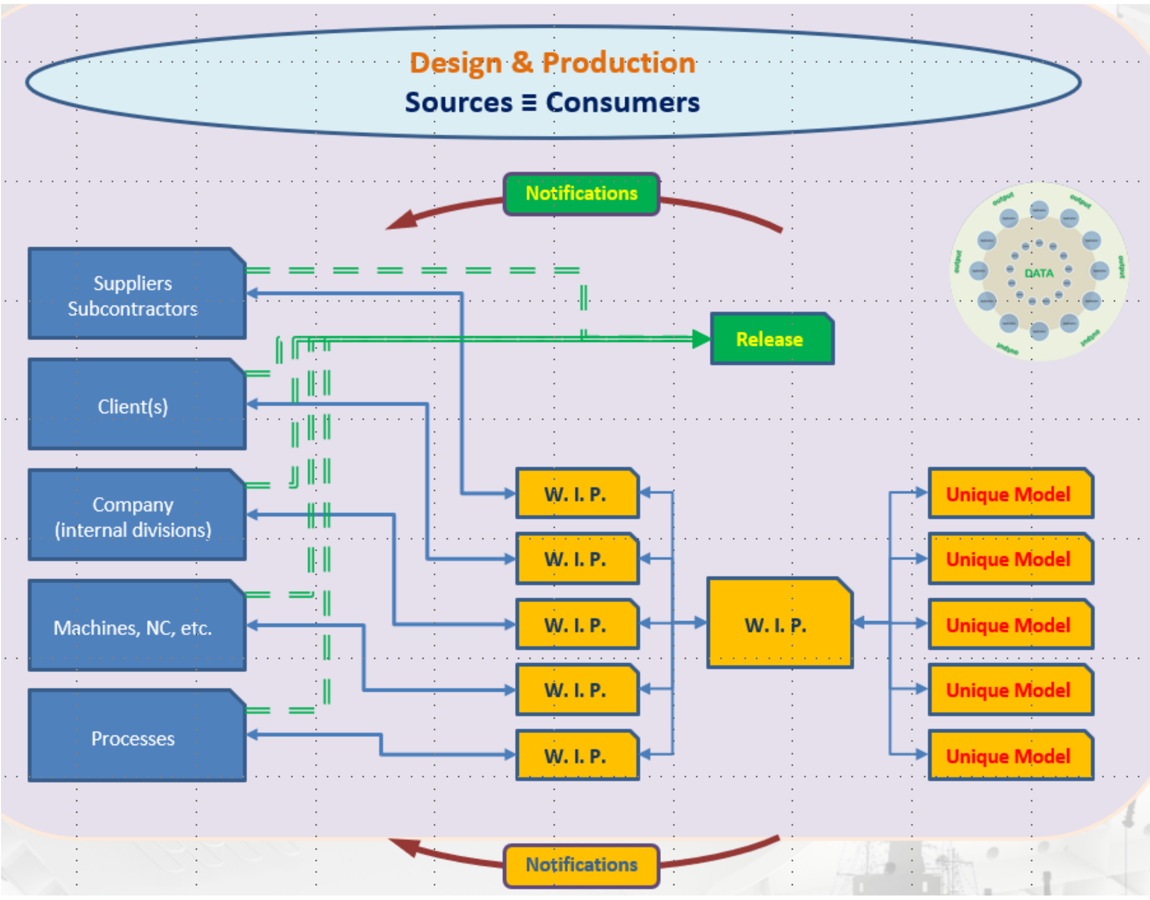


Fig.5: Work-In-Progress-Real-Time vs Release information streams

The Release information stream flows only when some portion of a Digital Twin becomes fixed by being "released". For example:

- an engine has been selected: all subsequent work related to the engine must refer to the selected engine
- the structural layout has been finalized: all subsequent work related to structure must refer to and make use of that structural layout
- the GA has been validated by the ship owner: the structural engineer is thereafter not allowed to change the location of pillars
- …

Once released, that part of the Digital Twin becomes immutable and the corresponding Work-In-Progress stream is stopped.

It is not the topic of this paper to discuss the implementation of information streams, but it will be noted that even "just useable" solutions are still very hard to come by, and even less so outside the highest echelons of the industry. On other hand, some level of success can be achieved by creatively using document management systems that allow that while respecting security and intellectual property requirements. In the author's experience there are several candidates on the market, ARAS PLM and Microsoft's Sharepoint based TEAMS being two incumbent favourites.

7. Practical Platforms of Platforms, Digital Twins, PLM, etc.

A universal Digital Twin is utopia: even if one were to collect absolutely all the data, information, etc. specific to shipbuilding, each company will implement different processes and follow different decision making strategies as a function of its specific requirements and constraints The scope of research leading to the present paper was intentionally focused towards the parts of a Digital Twin easiest to implement by any company and leading to the shortest-term, highest possible ROI within any generic, every-day context, that is, practical.

There are many seminal components of any given Digital Twin, and some have been mentioned already. One often unrecognised yet vital component of a good Digital Twin deserves a special mention: the ability to handle well change management. While information streams are crucial to convey the need for changes, the changes themselves, the results of changes and their consequences, the authors must be able to document changes in the first place, this consideration referring mostly to software and to the ability to identify parts of its own Digital Twin as a "change".

Otherwise, seminal components of Digital Twins include:

- People, Vision and Added Value
- Technology
- Software, Platforms and Tools
- Unique Models / Digital Twins
- Culture
- ...

This paper will focus on technology, software platforms and tools, considering that Unique Models Digital Twins are an achievable "given" within the company x context being considered and knowing that change management is already handled effectively in at least one of the out-of-the-box tool sets selected for the research initiative: The SSI DigitalHub and the SSI ShipbuildingPLM environments.

7.1. Practical

There are several daunting aspects in considering the overall concept of the Digital Twin and its effective use:

- the sheer amount of data
- the tens, hundreds and thousands of people involved
- the IT infrastructure required
- the difficulty in defining processes
- the inherent complexity of processes
- identifying which author will contribute what
- identifying the relative importance / gravity of an action or of its result
- the ever-changing requirements governing the above
- security
- the human handbrakes: fear of change, adversity to responsibility, etc.
- ...

Something is practical when it can be easily and effectively used by those involved to achieve the pursued goal. Now, while certainly not a substitute for a correct PLM implementation, the ensemble of many commonly available and already in use software tools can be federated into a Platform of Platforms, to build a not-so-ugly Digital Twin and exploited in PLM-style to generate ROI immediately, by virtually anyone.

In this respect, one crucial very early task when designing a Digital Twin is to establish the desired results, benefits and target ROI zones. If a Digital Twin cannot exist in a practical manner, no ROI will be generated, it will quickly become very ugly and, eventually, bad. On the contrary, losses will very likely ensue.

7.2. People, Vision and Added Value

Very directly related to "practical", people are too often the minuscule grain of sand that blocks the entire mechanism. Arguably, little has changed in data and information sharing techniques, processes and strategies in the marine industry since mainstream CAD and spreadsheets became available on the first PCs in the early-1980s, followed by user-friendlier databases in the early 1990s. AutoCAD® (by Autodesk®, USA), Microsoft® Excel® and Access® continue to be used and at least the former two remain the de-facto ubiquitous backbone of most companies. People have changed little when it comes to making good use of the Digital Twin begging them to from PC screens.

Vision is probably the most sorely missed driver of all industrial evolutions. The Kodak Moment becoming everyone else's with the advent of the digital camera is probably the most recent and resounding consequent disaster. (The Kodak moment was Kodak's commercial banner "the sentimental or charming moment worthy of capturing in a photograph" that became the epitomic moment when executives fail to realize how consumers are changing and how markets will ultimately evolve in new directions without them.) The lack of investment in implementing and exploiting existing tools to generate overall company ROI is not the sole area where short-sightedness is the cause of false economies. For example, man hours are often counted as such, e.g. a straight cost, as opposed to being valued and assigned as a function of their potential to generate added value. Current practice still includes reliance on:

- (subjective) human communications
- assumption that messages, duties, tasks, deliverables, etc. are evident, clear and known to everyone
- cost minimization takes precedence on ROI-producing choices

The lack of, if not the absence of a managed data & information sharing policy and strategy is accompanied by an under-exploitation of tools already in common use. Integration, cooperation, sharing, ERP and PLM are as commonly talked about and referred to as they are not implemented effectively, or at all. Added value at the company level is vital, that is how much was spent unnecessarily vs the sold-at price, an often-ignored metric. And, the potentially misleading measure of perceived “margin” at the microscopic level, typically the number of hours taken to produce something, takes precedence over measuring the impact of rework at company level, leading to potentially catastrophic false economies. Immediate opportunities to generate added value are the reduction, if not elimination, of errors, multiple instances of data and information and confusion: that is the mission of the Digital Twin.

7.3. Software, Platforms and Tools

Perhaps ironically, AutoCAD v.1 was a full-fledged 3D CAD program in 1982, Excel always offered multi-dimensional matrices and weakly relational environment, Access was released in 1992 as a fully relational database system and all include interfacing and integrating capabilities since the mid-1990s.

Another current world-wide de-facto standard is Rhinoceros3D®, by McNeel & Associates, better known as Rhino3D®. Its 1997 100000 users adopted Rhino3D's beta version within one year: it filled a technical void with user-friendly surface modelling. Victims of their own ease-of-use, these and other common software are used by virtually everyone to satisfy a limited, immediate need but are exploited only marginally and at a constantly diminishing fraction of their growing potential. This represents substantial company underperformance, and a considerable missed revenue opportunity and a weakening of the Digital Twin.

Today, many programs in addition to those mentioned here read and write a variety of proprietary data formats, but sharing meta-data remains another story altogether. Shared environments like DropBox®, OneDrive®, etc. are present on pretty much every PC and smart phone but used as mere repositories rather than the dynamic sharing platforms they are. This is another example of platforms which exist and are used separately rather than being harmonized to build a however small, good Digital Twin.

Connecting back to practical and recalling that communications are a sine-qua-non criterion and requirement, some software categories are defined accordingly:

- general purpose
- Read/Write different native formats
- multi-discipline
- integrated: run inside another software
- interfaced: shares data &information directly or via a native format with no degradation
- directly compliant: direct connection to Big Data, Digital Twin, AR, VR, IIIoT&S (Intelligent Industrial Internet of Things & Services)

Additionally, connected programs are considered only if they contribute exceptionally to the overall process despite their connectivity limitations, and can be categorized as:

- share data & information indirectly via a file format transformation/conversion, degradation is not ruled out

Other discriminating aspects were selected to include:

- size and industry-wide distribution of current user base
- options provided by the tool to communicate with all consumers: people, machines, other software, processes, etc.
- scope of application: design, production, both
- current and projected potential to serve the Integrated, Collaborative, Multi-Authoring, Managed Environment

A further level of categorizing is that of “purpose”, an admittedly anachronistic approach given the more recent evolutions in just about all software but nonetheless informative enough to justify being carried out. Some "purposes" could be:

- CAD
- Data & Information “management”
- file & document management systems
- rising technology: VR, AR, Big Data, Digital Twin, IIIoT&S

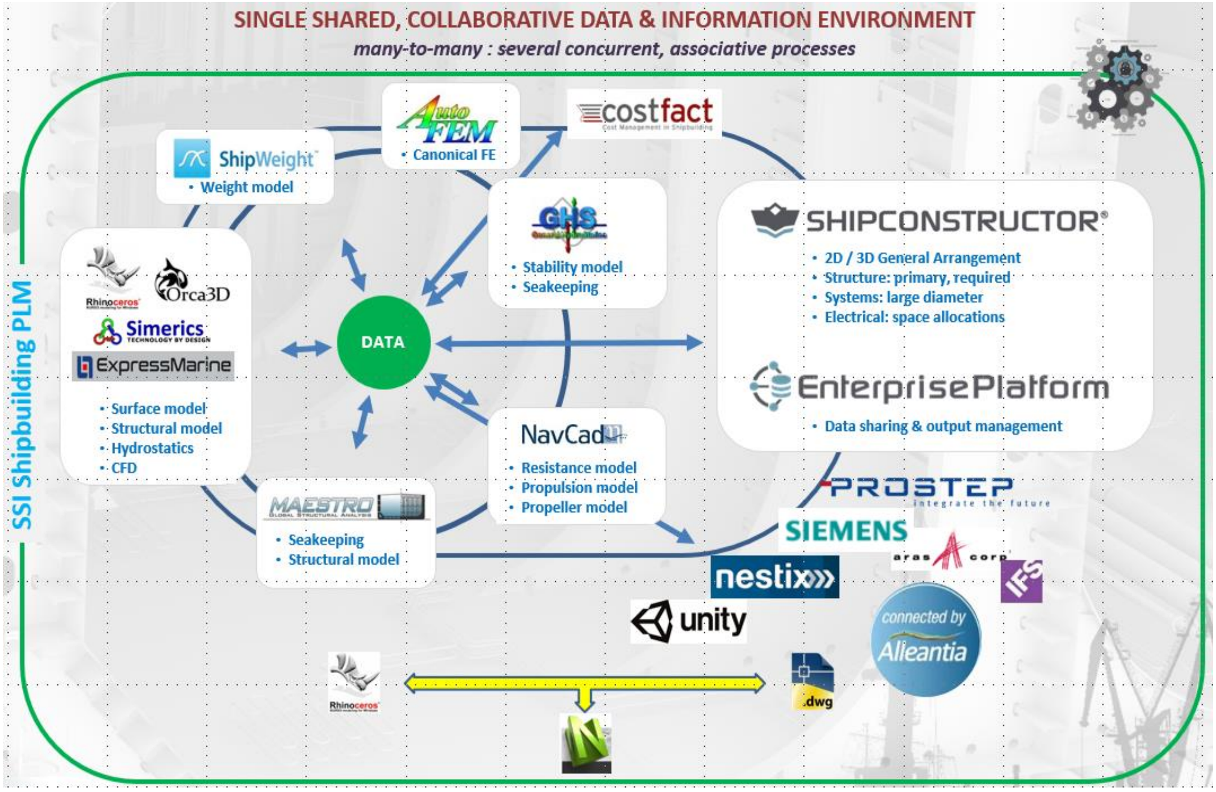


Fig.6: Schematic of representative software authors composing a practical Platform of Platforms

7.4. Out-of-the-box

The first goal of this research was the construction of an out-of-the-box Integrated, Collaborative, Multi-Authoring, Managed Environment made of tools that fulfil as many of the criteria listed here as possible, specifically being immediately available and productive for company x. The AutoCAD and Rhinoceros platforms where selected because they do so and additionally offer a number of unique business values, including:

- compatibility at the native data format level with each other and several other general purpose and specialist software
- easy to use, short time to proficiency
- significant untapped productivity potential
- several mainstream plug-ins and companion software based on these native formats and making use of the environment of both platforms

It is remarked that the sheer number of plausible candidate software products available in addition to the those selected for the study lend additional legitimacy to this study and its findings, *Danese and Dardel (2019)*.

8. A Paradigm Shift

The fundamental requirements, multiple platforms, rising technologies, varied formats specific to the candidate software components and the diversity of authors in nature, scope and purpose suggest a new approach to data & information use, storage and management, not dissimilar from the internet and processed Big Data models.

The functional paradigm that emerges from this study identifies a “space” in which multiple authors evolve in a collectively non-linear, individually linear, asynchronous, multi-relational, multi-platform, multi-format, collaborative fashion is that of the Digital Twin. (Somewhat ahead to this paper's conclusion, it will be noted that SSI’s EnterprisePlatform presents autonomous ubiquity of service and document handling automation characteristics which allow a vast data & information set extending far beyond what would be manageable with everyday manual procedures. The project space’s Digital Twin then grows to include crucial documents hitherto never considered, making EnterprisePlatform one of the few sine-qua-non parts of every performant Platform of Platforms.) The “space” collects a variety of data & information in different, heterogeneous formats:

- text files
- proprietary format files (3dm, dwg, etc.)
- xml files
- databases
- images, renderings, etc.
- Augmented and Virtual Reality
- ...

9. Practical Example

The number of practical examples multiplied over the course of the work, also due to on-going advances in some of the software selected for the R&D initiative. One of the first, and in many ways least impressive examples is illustrated here, its merits being relatively simple and near reaching. Not surprisingly, it was found in previous occasions that the full scope of the findings represents a challenge for those not open to considering, let alone accepting the value how much can be done with so little or those intimidated by comparing the scope of what is so very simply possible and immediately within grasp and their current practice. The more rewarding examples will be the subject of future reports on the present R&D initiative.

The example at hand consists in creating and maintaining a simple Digital Twin to serve some of the customary initial design activities, the subject being a 134 m L_{oa} motor yacht concept, extended to the advocated (virtual) production. The work was commissioned by a leading shipyard for a commercial customer and it will be referred to here as M/Y Digital Twin ("MYDT").

In the interest of consistency a few "qualifiers" are in order at this point:

- there is no convenient way to describe a non-linear collective process, therefore the author-to-author linear relations are presented in discrete format, grouped loosely in accordance with the somewhat overlapping software categorization by “purpose”
- referring to the "open architecture" mission of this R&D initiative, the reader is invited to replace the tools named here with her/his own if these were to be more productive.
- only few author-to-author relationship and interaction can be described here. The reader is therefore invited to construe a little and implicitly consider the shared, collaborative space.
- similarly, a sequential representation of a non-linear process inevitably violates some priorities and sequences of events, the reader should abstract from the formality of “lists” and construe the integrated, concurrent design and production space.
- crucial to the environment’s high-performance, most of the tools considered support scripting

and/or are automatable.
- unless explicitly noted otherwise, native file formats are used.

Moreover, a bland distinction is made between the overlapping phases of:

- design: from concept to design-for-production reviews, classification drawings and adjustments during production and warranty service, etc.
- production: starting with feasibility, early BOMs, weld estimations, what-if studies, design-for-production reviews, etc. through delivery and warranty, etc.

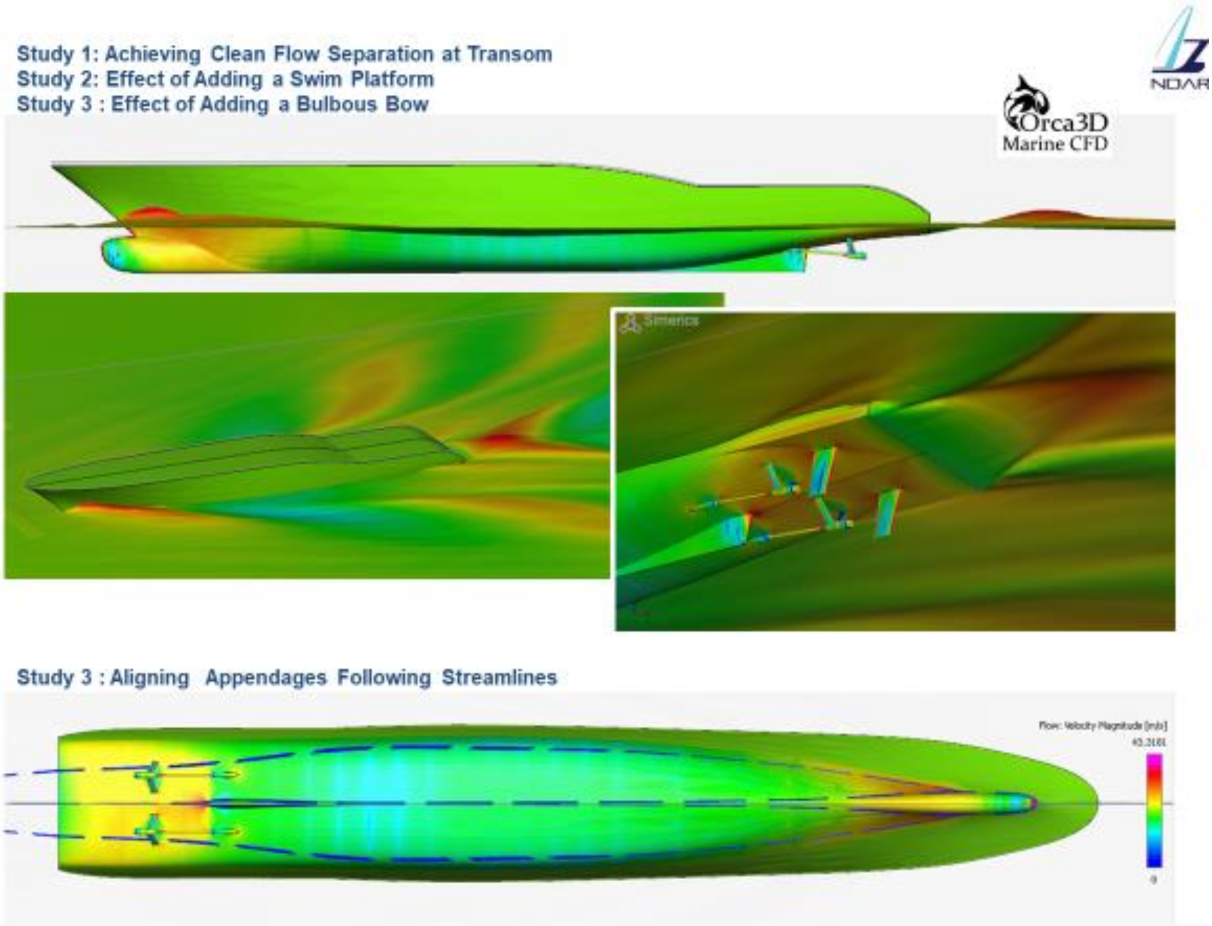


Fig.7: Visualisation and graphical post processing of flow analysis during the optmization of the combined hull, skeg, rudders and shaft brackets assembly for optimum flow into the propellers

10. M/Y Digital Twin

MYDT's principal characteristics are derived from initial GA plans, owner's requirements and initial Naval Architecture calculations:

L_{OA}	B_{OA}	D	Speed (cruising)
134 m	18.5 m	14 m	25 kn

LWL	DISPL	T	Gross Tonnage
125 m	6330 t	5.5 m	~4000 GT

Only the hull was addressed at this stage, it being considered that aside from its Weight and CG (that were estimated anyhow using ShipWeight®) the explicit modelling of the superstructure would have contributed little to drawing conclusions in this very first phase of the research. Of course, explicit modelling of the superstructure would allow air flow analysis using the same tool set employed in the CFD study, that was Orca3D CFD, production planning estimations, etc. The CFD engine used to analyse the Rhino3D surface models faired with the Orca3D plug-in was SimericsMP (marine).

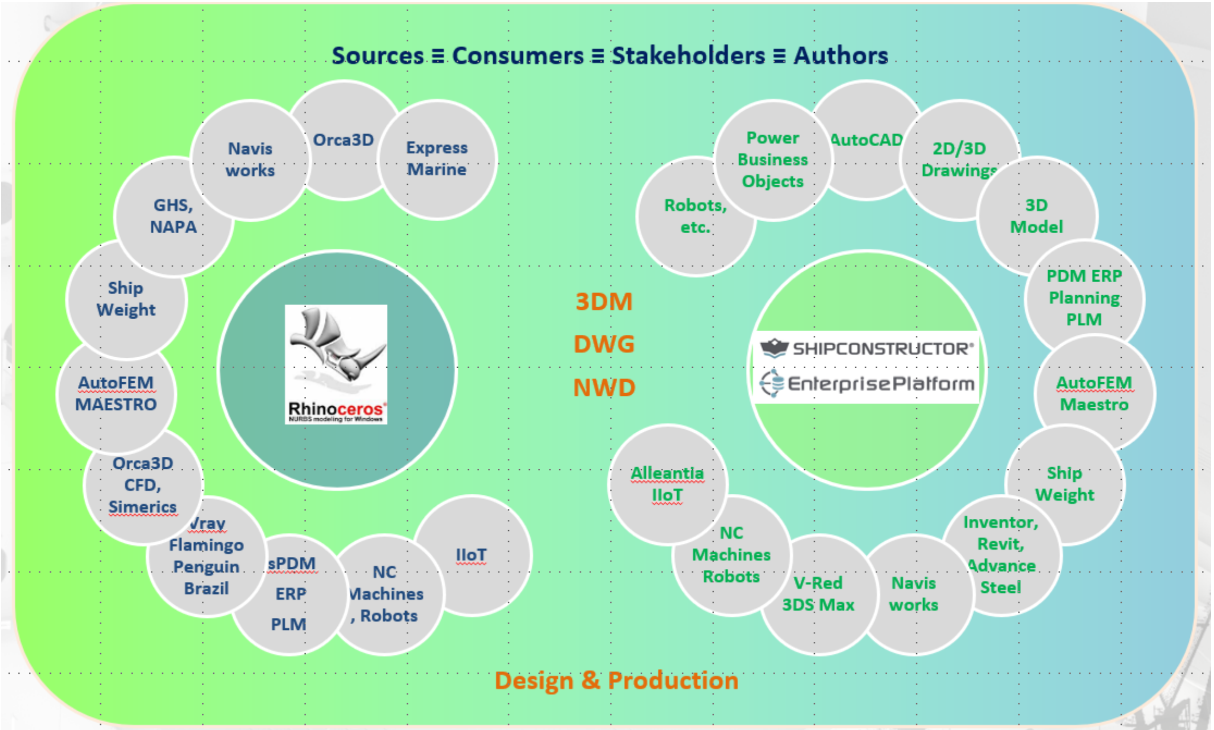


Fig.8: The Digital Twin common, shared data set transforms the hypothetical overlaps of yesterday into the symbiotic ecosystem of today and tomorrow

10.1. The Workspace

The common data set shared via the Digital Twin was exploited by several concurrent users, their actions and interactions based on the native Rhino3D and AutoCAD 3D data formats. Concurrent work is represented by a workspace rather than by a workflow, but using a more legacy approach a "list" of tasks / roles and actions will nonetheless be used to illustrate the workspace.

10.1.1. Tasks/Roles

Tasks and Roles are inherently one, and only software tasks/roles are reported here. A more complete report would also include those of processes and people. Tasks/roles included:

- ShipWeight: parametric weight and CG estimation, produce a full report for Naval Architect, Project Manager, etc., summaries in custom xml/text format for other authors.
- Cost Fact plus ShipWeight, ShipConstructor, ERP, etc.: parametric cost estimation, metadata merged with other authors'
- Rhino3D, Orca3D, Orca3D CFD plus V-Ray, Bongo, Penguin, Flamingo, Brazil etc.: create fair surface model to be the parametric base for other authors. Model compartments for stability calculations. Run early CFD experiments. Produce renderings, animations of doors, deck equipment, toys, etc.
- AutoCAD plus ShipConstructor, Revit, Inventor, AdvanceSteel, EnterprisePlatform, V-Ray, 3DS Studio Max, etc.: read Rhino + ExpressMarine model, compose 3D GA, automatically produce the corresponding 2D drawings, BOMs, discipline 2D/3D drawings, study/review/management 3D models (full and partial). Produce renderings.

- EnterprisePlatform plus Project, Navisworks: create and automatically maintain the project space's complete archive in accordance with planning. Use EP Operations to interact with other scriptable software, handle files generated by other authors (CAD, ERPs, DMSs, PLMs, MS-Excel, etc.) Merge with planning. Produce Unique Models

10.1.2. Actions

Only software actions are reported here. A more complete report would also include those of processes and people. Actions included:

- Project plus Navisworks, ShipConstructor, other authors: merged, interactive planning with aligned geometry and meta-data.
- Rhino3D plus ExpressMarine3D: parametrically develop first model of non-dimensional primary and secondary structure based on hull shape and GA.
- AutoFEM: read ExpressMarine model, evaluate local/medium size structural arrangements.
- GHS: read Rhino hull model (with/without compartments), read ShipWeight weight curves, define loading conditions, compute damaged and probabilistic stability, longitudinal bending moment, run seakeeping. Produce report for Naval Architect, Project Manager, text/xml summaries for other authors.
- NavCad and PropElements: read the Rhino model (stl format), compute resistance, wave drag, propulsion train (engine, power, propeller), converge design space extents for CFD runs. Refine propeller sized in NavCad.
- MAESTRO: compose global structure model, first principles scantling calculations, read ShipWeight's weight curve + GHS's loading conditions, compute time & frequency domain ship motions, full fatigue screening, early optimisation for production. Read Rhino model (iges). Read ExpressMarine model, apply scantlings for first structural estimation.
- Navisworks: collect 2D drawings, Rhino3D, AutoCAD, GHS, MAESTRO, ShipConstructor, ShipWeight rich 3D models for review, comparison, project management. Produce Unique Models.
- Several authors: seed the VR, AR, Digital Twin, IIIoT&S environments.
- Data is tracked and monitored by the presence / absence / changes of files identified in a user-defined, editable event list and by tracking data and changes to it within xml files. Data will change during the occurrence of the various events, author(s) are notified by MS-TEAMS®, etc.
- data & information are shared outside the company using any suitable shared environment, WAN, VPN, Cloud, etc. such as DropBox, OneDrive, FTP, MS-TEAMS, etc. Email, USB sticks and similar unmanaged data transmission means are formally proscribed.

As fundamental Naval Architecture parameters converge, some authors are called upon more than others, for example:

- AutoCAD plus ShipConstructor, EnterprisePlatform, Cost Fact: read the ExpressMarine model, create a formal structure model, advance main distributed systems model (pipe, HVAC, electrical), compose general reference model, produce Class drawings and advanced documentation for detailed cost analysis, produce renderings, VR, AR models (used with mock-ups).
- AutoCAD plus ShipConstructor, Revit, Inventor, V-Ray, 3DS Studio Max, etc.: advance 3D GA, produce 2D plans, renderings, feed the VR, AR, Digital Twin platforms.
- MS-TEAMS, other shared environment software: maintain data & information alignment amongst all authors.

This leads to further detailing the weight model and advancing final structural, stability and propulsion evaluations:

- ShipWeight: acquire advanced part lists with weights and CGs, continue comparison between weight budgets and weight of tracked items.
- AutoFEM, GHS, MAESTRO, NavCad, etc.: carry out advanced Naval Architecture checks, final life-cycle performance estimates.
- CostFact: increase detail level of cost prediction and analysis, accuracy of monitoring.
- EnteprisePlatform: compose initial work packages for production.

At some point during the events mentioned above, production engineering and production will have commenced:

- AutoCAD, Rhino3D, Cost Fact, Project, etc. plus ShipConstructor, ExpressMarine, AutoFEM, EnterprisePlatform, Navisworks: produce detailed design model and drawings, production engineering data and documents, planning, work packages, KPI trackers, specific documents to support individual departments and disciplines, etc.
- AutoCAD, Rhino plus 3DS StudioMax, V-Ray, Brazil, Navisworks: rendering, VR, AR.
- AutoCAD, MIM plus ShipConstructor, Navisworks, EnterprisePlatform: Digital Twin, operational PLM model of the vessel.

11. Platform of Platforms and Symbiotic Digital Twins

Despite its intentionally simplified scope, the R&D initiative unveiled the considerable extents of its practical reach, available here and now to company x. This is clearly revealed by the completion of every single further action carried out in the workspace. Key guidelines are found in the AGILE and LEAN definition of scheduled internal deliverables needed to track and manage the project's progress and identify situations requiring decisions and of ROI generating external deliverable (of which drawings remain a component, yet more and more minor).

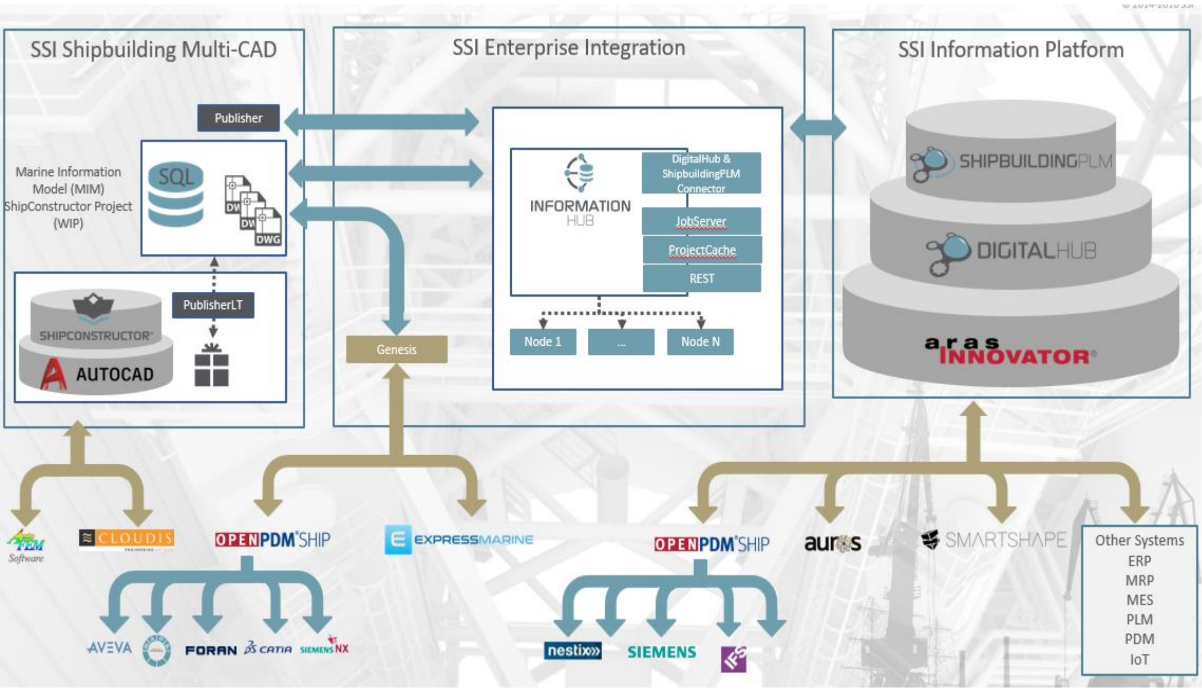


Fig.7: A schematic representation of a portion of the symbiotic Digital Twin ecosystem available today to Company X, here and now

Another important note to make is that despite the significant data & information handling and management abilities of some software, the out-of-the-box environment includes some unavoidably unmanaged components that require ad-hoc definition, instruction and supervision.

12. Vision of the future

Companies with vision can safely table significant expectations as long as the human factor is taken into account realistically. Technology, software, understanding of processes, etc. progress faster than they can be tested, reviewed and analysed for a rational implementation thereof. This must not be perceived as a hamper or as a cause of a self-damaging "always being behind". On the contrary, the definition of Platform of Platforms and very nature of the symbiotic Digital Twin lend themselves to being easily fertilized by new, future seminal authors. It is the responsibility of the Digital Twin designers to plan for organic growth of the Digital Twin under AGILE and LEAN guidelines. Progressive, ROI-driven evolution, not revolution, benefits from transition which in turn requires temporarily co-existing, possibly even overlapping "equivalent" data sets, processes, etc. Naturally, new tools and processes will favour different parts, portions and aspects of the symbiotic Digital Twin but this does not necessarily render obsolete the one fallen in disuse (think about a life-time, non-evolutive Digital Twin as that of in-service nuclear power plants).

The vision of positive, progressive future therefore rests on people having a vision that will define, design, maintain, evolve and implement Digital Twins, exploit them wisely and with clear purpose, taking full advantage of technological, software, process, market, industry and the wealth of alive and healthy, yet disregarded evolutions around us, everywhere and always. Then, it will be said Vision, ergo Digital Transformation.

12. Conclusion

The symbiotic Digital Twin is a reality that is enhanced and improves every day. One just needs to look away from the false safety of legacy and accept the reality of the cultural, liberating paradigm shift that will immediately enhance a company's quality, productivity and ROI.

In the near future it can so easily be common software, so much of it already in use in just about every office and shipyard around the world that will contribute so well to today's and tomorrow's good Digital Twin and launch a true Digital Transformation towards a good Platform of Platforms.

In the longer term, large, high-end systems will be the backbone of the evolving, a good Platform of Platforms and cease being the monolithic solutions they are purported to be today. To no longer be a monolithic do-everything platform and to become a part of the Platform of Platforms ecosystem will significantly help manage the symbiotic Digital Twin. It will also progressively eliminate the human factor overhead generated by the all-powerful yet impractical and elusive potential of large monolithic systems by creating ideal yet unrealistic expectations, let alone goals.

Tomorrow there will be tiny and very large Platforms of Platforms side by side and interacting. The unimaginably large Platforms of Platforms of tomorrow is already being built today, one Platform at the time, by people with vision.

The small feeding the big in LEAN and AGILE fashion and driven by the realistic and pragmatic visionary human is the ultimate answer of all industries, marine included.

Acknowledgements

Of the very many that over time have contributed knowingly and unknowingly to my work and personal evolution, I will name two who exemplify the immense scope of help I was blessed with: Arkadiy Zagorskiy, central pillar in troubleshooting the implementation of the sometimes crazy processes I cook up on an almost daily basis and Denis Morais, intellectual sparring partner and seminal source of so much of my work. I am here today thanks to you all, too .

References

DANESE, N.; PAGLIUCA, P. (2019), *Available vs Accessible data and information: The strategic role of adaptive communication in the Naval Architecture and Marine Engineering processes*, CNM, Naples

DANESE, N.; DARDEL, S. (2019), *Out-Of-The-Box Integrated, Collaborative, Multi-Authoring, Managed Environment for the Design and Construction of Large Yachts*, Design & Construction of Super & Mega Yachts, Genova

Proactive Control of Maritime Production Processes – A Sneak Preview

Carsten Zerbst, PROSTEP, Hamburg/Germany, Carsten.Zerbst@PROSTEP.com

Abstract

This paper presents an overview of the German R&D project ProProS, started in 2020, between shipyard Fr. Lürssen, university RWTH Aachen and software developer PROSTEP. The goal of the research project is to develop methodology and tools for efficient and near-time production planning with typical one-of-a-kind ship production.

1. Introduction

Maritime systems, especially in the field of special shipbuilding or offshore structures exceed the dimension of conventional products. Not only physically, but also by the number of parts, systems to integrate or assembly steps to perform. That is why the complexity of these large-scale products poses special challenges for shipyards — from the manufacture of parts to assembly and equipment. In addition to the technological complexity of the product itself, work planning is a challenge by itself. On the one hand, there is the sheer number of parts combined with the number of involved suppliers, on the other, hand most products are prototypes and thus do not support a planning effort comparable with those of mass production.

As a result, planning is usually done very roughly in weekly or monthly slices, the detailed planning only by foremen and foremen in the form of shift plans. Therefore the current production status is often not clearly visible to all participants. Acute problems often go undetected to those involved in project planning and control. Operational failures - caused by material delays or capacity bottlenecks - lead to production downtimes and extended lead times. Adding late time changes demanded by ship owners to this mix makes managing even more difficult.

Currently production planning methodologies and tools are targeting mass production like e.g. automotive industry. These are characterized by huge number of repetitions and support a certain amount of time spent on designing and tracking each assembly. Prototype production like in shipbuilding industry on the other side need to find much more efficient ways to design, plan and track assembly work.

The goal of this research project is, therefore, to enhance currently used methodology and provide tools, which enables the yard to achieve

- Significant reduction in turnaround time to design, plan and track assemblies for ship steel and outfitting
- Increasing the degree of digitization and the associated transparency in production
- Reduction in the proportion of missing parts in an assembly
- Efficient use of employees by eliminating waiting times
- Identification of cause-effect relationships of faults in the assembly of large-scale products

2. Project Overview

The project ProProS (from German "Proaktive Produktionssteuerung für die Produktion maritimer Systeme" = "Proactive production control for the production of maritime systems") combines industry and research partners. It is partially funded by the German Federal Ministry for Economic Affairs and Energy.

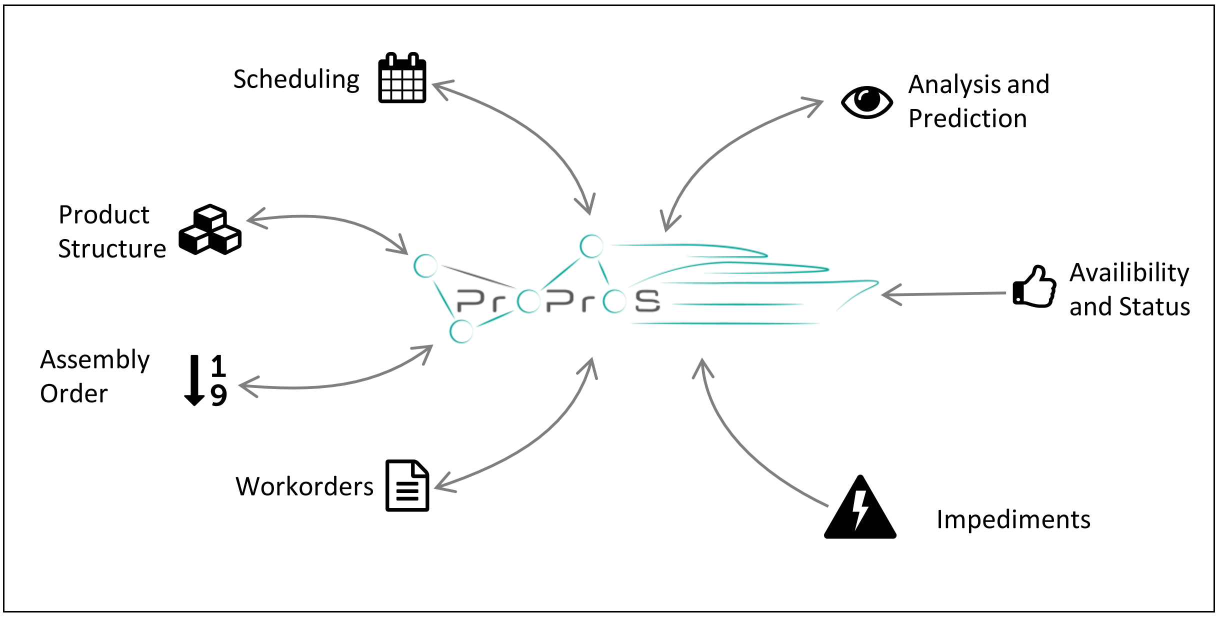


Fig.1: ProProS Overview

2.1 Project Goals

The most important project goals from the yard's perspective are:

- Increasing the degree of digitalization in production to achieve better transparency
- Forward planning, monitoring and control with key figures to achieve better adherence to delivery dates
- Reduction of the lead time for equipping a section
- Reduction of the proportion of missing parts in assembly
- Increase of infrastructure and resources utilization
- Efficient employee deployment (no search & sort times, no waiting times, minimum travel times, no alternative work)

The scientific objectives from research perspective are:

- Identification of cause-and-effect correlations of faults in the assembly of large-scale products
- Development of optimization algorithms for highly complex systems
- Research into the usability of real-time data for planning and controlling the production of large-scale products

2.1. Project Partners

The project is jointly run by three partners, bringing their respective strength into the project.

Project lead is the Fr. Lürssen Yard in Bremen-Vegesack. Founded in 1875 by Friedrich Lürßen, the yard is family hold and manged. They are producing most superyachts and naval vessels on several product sites in northern Germany with about 2800 employees. The yards role is to provide insight into their current methodologies and tools and later on to verify the applicability of the projects.

The academical project partner is the RWTH Aachen with their Laboratory for Machine Tools and Production Engineering (WZL). Their task is to development the planning and tracking methodology in the project.

PROSTEP is the third project partner. We bring in our shipbuilding industry related consultancy and software development experience. Our goal is to provide a prototypical software tool integrating existing data and implementing the new methodology.

2.2. Project Structure

The research project started in late 2019, the end is in late 2022. The complete workload is split into four work packages:

- WP1 Requirement Analysis - Analyse currently applied methodologies, available information, and requirements towards a better solution
- WP2 Enhance Methodology and Data Model - Identify better planning methodologies and supporting data model
- WP3 Implementation - Implement the solution as a proof of concept
- WP4 Validation - Validate the solution using real yard data.

The work packages 2 and 3 have an internal structure as found in Fig.2. This allows to run our project in an agile matter and better track project achievements. Implementing all five steps allows to run a full Demming cycle on shipbuilding production.

3. Outlook

When writing this overview paper, the WP1 is coming to an end. We have investigated the methodology, processes and software tools currently run at the yard for production preparation, planning, and feedback. This leads to a list of ca. 50 use-cases to be implemented by the future solution.

The WP2 will start in the weeks after COMPIT 2020, focussing on how to achieve better planning. The first results of WP2 will be presented at COMPIT 2021.

Acknowledgements

I thank my project partners Dr. Bernhard Urban, Heiko Buchholz, and Maja Rother (Fr. Lürssen Werft), Matthias Dannapfel, Richard Minderjahn (WZL of RWTH Aachen)

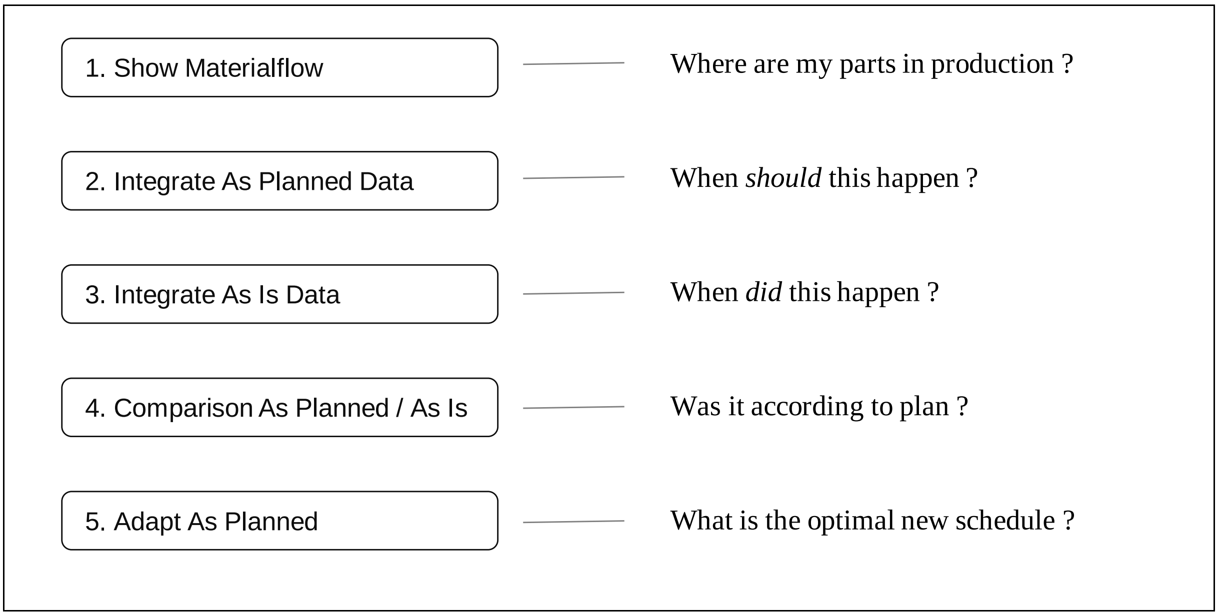


Fig.2: Agile steps in ProProS

Index by Authors

20th Conference on

Computer Applications and Information Technology in the Maritime Industries

COMPIT'21

Mülheim / Germany, 31.5-2.6.2021

Topics: Artificial Intelligence / Big Data / CAX Techniques / Digital Twin / Simulations / Virtual & Augmented Reality / Robotics / Autonomous Technology

In Design, Production and Operation of Maritime Systems

Organiser: Volker Bertram (volker@vb-conferences.com)

Advisory Committee:

Marcus Bole	AVEVA, UK	**Henrique Gaspar**	NTNU, Norway	**Rodrigo Perez**	SENER, Spain
Andrea Caiti	Univ Pisa, Italy	**Ken Goh**	Knud E. Hansen, Australia	**Ulf Siwe**	Swed. Mar. Adm., Sweden
Jean-David Caprace	COPPE, Brazil	**Stefan Harries**	Friendship Systems, Germany	**Myeong-Jo Son**	Korean Register, S. Korea
Nick Danese	NDAR, France	**Kohei Matsuo**	NMRI, Japan	**Julie Stark**	US Navy, USA
				Bastiaan Veelo	SARC, Netherlands

Venue: The conference will be held in the "Die Wolfsburg" in Mülheim an der Ruhr

Format: Papers to the above topics are invited and will be selected by a committee.

Deadlines:

anytime	Optional "early warning" of intent to submit paper
20.12.2020	First round of abstract selection (1/3 of available slots)
20.01.2021	**Final round of abstract selection** (remaining 2/3 of slots)
21.3.2021	Payment due for authors
21.3.2021	Final papers due (50 € surcharge for late submission)

Fees:

650 € / 350 €	regular / PhD student – early registration (by 21.3.2021)
750 € / 400 €	regular / PhD student – late registration

Fees are subject to 19% VAT
Fees include proceedings, lunches, coffee breaks and conference dinner

Sponsors: Aveva, Korean Register, Sener, Siemens (further sponsors to be announced)

Information: www.compit.info