Validation methodology for assessment of new eNavServices

Dag Atle Nesheim

Transport and Energy, SINTEF Ocean, Norway. E-mail: DagAtle.Nesheim@sintef.no

Marius Imset

Department of Maritime operations, University of South-Eastern Norway. E-mail: Marius.Imset@usn.no

Kay Endre Fjørtoft

Transport and Energy, SINTEF Ocean, Norway. E-mail: Kay.Fjortoft@sintef.no

The international maritime industry is on a current wave of digitalization and there is a dire need to enable an effective prediction of the different digital solutions, especially in terms of international implementation as well as regulatory requirements related to the same. When assessing new e-navigation (e-nav) solutions, these are not yet implemented, i.e., there is no possibility to validate technology on basis of lagging indicators - indicators which express actual performance within quality of the service to which the solution is targeted. One must instead turn to leading indicators - indicators which, based on a hypothesis, express expected performance within quality of the service to which the solution is targeted. The more valid the hypothesis, the more valid the leading indicators will be to express the new solutions' effectiveness. In this paper, we have validated KPIs from the Baltic and International Maritime Council's Shipping KPI Standard in light of their suitability in expressing actual performance, hereby validating the quality and suitability of a set of leading indicators from the port call scenario from the Research Council of Norway supported SESAME Solutions 2 project.

Keywords: Shipping KPI, Key Performance Indicators, E-navigation

1. Introduction

Several guidelines, such as the International Association of Marine Aids to Navigation and Lighthouse Authorities' guidelines for the development of testbeds (IALA, 2016) and the International Maritime Organization's guideline on software quality assurance and human-centred design for e-navigation (IMO, 2015) also rely on leading indicators. The same is true for vetting regimes such as the Ship Inspection Report programme from the Oil Companies International Marine Forum. Baldauf et al. (2016) estimate the effect of IMO’s e-navigation initiative. These all have in common an underlying hypothesis on which attributes constitute a future rise in quality once implemented. Finally, several international rules and regulations related to safe, secure and environmentally sustainable operations also rely on the notion of what will result in a positive effect.

So, how do we ensure that the hypothesis provide valid indicators of expected performance? The methodology described below, is created on basis of the relationship between leading and lagging indicators and how this relationship provides a continuous improvement loop between the two.

A leading indicator predicts performance while a lagging indicator expresses actual performance:

To assess whether the leading indicators were able to make high quality predictions, we use lagging indicators to express whether the predicted performance was met in real life.

Proceedings of the 32nd European Safety and Reliability Conference
Edited by Maria Chiara Leva, Edoardo Patelli, Luca Podofillini, and Simon Wilson

ISBN: 981-973-0000-00-0 :: doi: 10.3850/981-973-0000-00-0_esrel2022-paper

2. How are leading indicators created?

Leading indicators are created on basis of the notion that they are suited to foresee an increase in actual quality of the task at hand: An example is that a graduate student's grades can be used to foresee said student's ability to be an asset at a workplace. The higher the grades, the higher the expectations on the student becoming a valuable asset. A more relevant example would be that a certain (high) number of reported near misses (incidents which could have led to an accident but didn't) is a positive indicator of the quality of a ship manager's Safety Management System (SMS). Both of these hypotheses carry anecdotal evidence of being both right and wrong. A top-level student may be a disaster in the workplace and a high number of reported near misses may be due to an insufficient level of safety awareness. Nevertheless, both hypotheses are in use today, even though their underlying assumptions are far from unambiguous. How can we ensure a continuous validation and improvement of the hypotheses used to create leading indicators?

2.1. Why are leading indicators used?

If the validity of leading indicators is questionable, why are they being used? The obvious reason is that we cannot look into the future and see the actual effects of novel solutions and technology. We need to make qualified guesses and estimates based on obtainable information at the time of the assessment. If we assess the development of leading indicators and compare that to the process of planning practices, there are obvious similarities. We do not always know the status and what we are planning for, but we know that there will be a need for data to be used for the planning of an operation. The precision of planning will be more accurate the closer you are to the execution since the quality and data precision is getting better. You may not always know the leading indicators, which can change during the process, but you know the purposes of them.

3. Validating leading indicators through monitoring actual performance

As the use of leading indicators is required in cases where estimated or foreseen performance is needed to validate the benefits from novel solutions and technology, one must ensure that the leading indicators used, and the hypotheses on which they are based, are as valid as possible. To ensure this, a continuous loop of validation and improvements is suggested. In short, we use lagging KPIs, expressing actual performance, to validate whether the hypotheses and leading indicators were right in their assumptions.

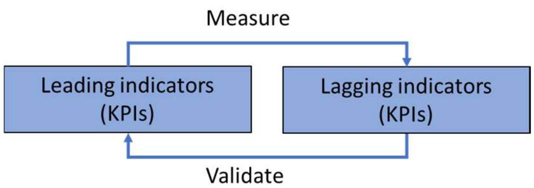

Figure 1 The relationship between leading and lagging KPIs

Leading indicators are initially created through hypotheses, stating that a certain set of these indicators is said to affect future performance in terms of quality, safety, costs or other parameters. The associated lagging indicators are derived from the hypotheses through identifying how this future performance should be expressed. These same lagging indicators are then used, when the solution or technology has been implemented, to monitor actual performance.

Through correlation analyses, we are when able to assess the leading indicators' validity. Should high values on lagging KPIs correspond with high values on leading KPIs, the hypothesis cannot be discarded. Through causation analyses, the hypothesis can be verified as valid. Further analyses may also identify spurious correlations and even the parameters actually affecting the performance - parameters which can then be used to form more valid hypotheses to enable the creation of more valid leading indicators for future assessments of new solutions or technologies.

3.1. The issue of timing

One obvious challenge related to the suggested methodology, is the issue of timing. If lagging indicators of actual performance indicate that the leading indicators used to predict performance are not valid, decisions have been made on invalid hypotheses. Nevertheless, the knowledge gained can be used to ensure that the same mistakes are not repeated. A proven invalid hypothesis may prove just as useful as a proven valid hypothesis, following the theory of Hegelian dialectic (Hegel, 1807) where a thesis is challenged by an antithesis resulting in a synthesis. This, however, requires that there is a similar set of solutions eligible for assessment further down the road, as in our scope related to E-navigation and Maritime Services.

4. A case study from the SESAME Solution 2 project

The SESAME Solution II Consortium consists of a group of Norwegian maritime technology suppliers, governmental authorities and universities/research institutions. The project is owned by Kongsberg Norcontrol, which provides Vessel Traffic Service (VTS) systems as part of their portfolio, and it is supported with funding from the Norwegian Research Council and has the following partners: NAVTOR, the Norwegian Coastal Directorate, the Norwegian Maritime Authority, Kongsberg Seatex, University of South East Norway, NTNU, SINTEF Ocean and Western Norway University of Applied Sciences.

The objective of the SESAME II project is to develop and validate a new method of ship traffic management, through digitization, that enables the bridge team and onshore authorities to:

- Increase shared situational awareness through digitization of ship to shore communications, focused on local ship traffic, port approach planning, and challenging weather conditions,
- Improve collaborative decision support through shared alerts and traffic predictions, focused on inappropriate ship behaviour, and
- Reduce the administrative burden of both bridge teams and shore-based operators through the digitization and automation of ship-to-shore reporting

The project shall accomplish these objectives by developing new functionality in both existing systems and prototype services, improving new communications strategies, and studying the effect of the new technology on operators. Development shall be on ship systems, shore systems, and communications equipment.

SESAME S2 seeks to be the first project to develop and demonstrate, on operational systems both onboard and ashore, a fully realized suite of e-navigation services. It aims to apply the Human-Centered Design (HCD) guidelines proposed in IMO MSC.1/Circ.1512, and as part of this, to quantify the effects of new systems and services on human operators and ship performance. These data will be part of the business case for end-users to buy into e-navigation solutions, especially ship owners and operators.

Traffic information and forecasts today are commonly provided by VTS operators and communicated to vessels by VHF. A basic assumption for SESAME II is that new digital technologies have the potential to improve coordination and information sharing between key stakeholders during port calls, thus can making the process more efficient and accurate.

4.1. The new SESAME system

The new SESAME system is composed of the C-Scope Management Information System, a maritime surveillance system provided by Kongsberg Norcontrol used for traffic management by VTS operators, with a new web-based module that allows port arrival service providers (pilots, tug operators and others) to collaborate with regards to stating readiness and availability, reporting starting and stopping a service, and showing service progress. When coupled with ship-to-shore route sharing functionality in the context of e-navigation, the intended operational benefit is first to enable an accurate Just-in-Time Arrival service that coordinates arrival time with berth and service provider availability. Such a service is expected to reduce the administrative burden of arrival planning, especially when plans suddenly require changing.

4.2. Assessing cost/benefit and operational impact

It was a key challenge here, as in most e-Nav projects, to perform a scientifically sound assessment of the actual operational benefits created by the technology solutions developed in the project. Two complimentary approaches were chosen: 1) A cost benefit analysis judging the economic advantages or disadvantages of an investment decision by assessing its costs and benefits in order to assess the welfare change attributable to it. WP6) and 2) A study addressing a specific, historic port call (WP2).

The reasons for looking into a specific port call was to work according to IMO MSC.1/Circ.1512, i.e. to get in-depth knowledge of how operations are actually done (context-of-use). The benefit of this approach is that it ensures a realistic representation of the process that is to be improved, rather than using more generic approximations that may not reflect how work is actually being done.

Having these two different approaches allows for comparisons of result, i.e., to ensure that data from a specific instance/sample of a port call aligns with data used for a more generalized analysis. The better the data align, the more valid are the generalization.

Here we meet the same issue related to timing: We are unable to assess the benefits of the solutions as they are not yet implemented and no actual improvement in performance can be proven. We therefore have to rely on leading indicators.

4.3. The historic port call study

The method used is a variation of A/B testing (https://usabilitygeek.com/introduction-a-b-testing/), where users are testing two variants of a system or website to compare, by means of usability indicators, which variant performs better. Our approach was to define a historic, real port call process ("as-is") along with measures of relevant indicators, and then recreate the same port call with use of the new system ("to-be"). The comparison is based on assumptions from the system developer (where we have data from SESAME S2).

In order to define a port call that both was real/historic and at the same time represented potential benefits of the new system, the project collaborated with Associated British Ports (ABP) in Humber, UK. It was also decided to narrow down onto two leading indicators, related to the project goal of reducing administrative burden:

1. Reduce the time spent on communication among key stakeholders (operators)
2. Reduce the number of communications between the stakeholders

The port call takes place on the Eastern coast of the UK, in the Humber. The identity of the two involved vessels is not relevant, so they have been anonymized. At about noon the arriving vessel (ARR) was passing Spurn Head, inbound for the Humber International Terminal 1 (HIT-1). HIT-1 is at this time occupied by another vessel (ARR), scheduled to depart at 13.00. As both of them were deep draft vessels, they needed to use the Sunk Dredged Channel (SDC). The SDC is only wide enough for one vessel at the time, and the depth is sufficient only within a limited tidal window. The two vessels were planned to pass one another and start/end their tug operations around at around 13.30, in the area between the Port of Immingham and Sunk Spit. However, during preparations for departure, the crew at DEP detected a critical equipment failure. As the situation problem onboard DEP was not sorted at the time ARR was approaching the port, she needed to turn back and head back for the SDC. Another affected vessel was the tug Svitzer

Victory, as she was scheduled to first assist DEP, and then meet and assist ARR on her way to HIT-1.

Figure 2 Map of the area and the two vessels.

The data for VHF communication was retrieved as recordings from the Kongsberg Norcontrol C-Scope system, used by the VTS. The recordings were analysed and transcribed in collaboration by a human factors researcher and experienced sea captain from University of South Eastern Norway, and a VTS specialist at ABP. Time for each communication started when one part called for the other, and ended when the communication was over. Time was measured in seconds.

The content of the communications, parties involved, and time spent on the communication was then summarized in an excel-table, allowing for direct comparison with the "to-be" scenario. The number of communications, as well as the total time spent, is summarized in the table.

The table below shows the details of the communication in the "as-is" (based on real data and measurements) and "to-be" (based on estimations from the system developer) scenarios. In some comms there is a dialogue. To make clear who is saying what, the dialogue is separated by a slash /, and the recipient is denoted by use of italics (this corresponds to the From and To column). Expected time savings are denoted in green text.

Table 1 Sample from data table

Time of event	Communication tasks/content	From	*To*	Duration (sec)	Using the Kongsberg Norcontrol C-Scope system	From	Time saved (sec)
12:58	Mandatory reporting - Passing papa 5 (i.e. buoy P5) / *Immingham 5.65, (tidal info gage reading), the Laura and Castle are on their way.*	Pilots on board	*Vessel Traffic Service Officer*	25	Pilot types in current position report, and request tidal info	Pilots on board	5
13:00	Can you give me an update on tugs? / *You'll be getting the tugs from the outer harbour, (VIKING DESTINY) is going to stations, the fourth one will be the Victory from HiT (DEP) if it sails /*	Pilots on board	*Vessel Traffic Service Officer*	40	Type question into Movement chat of ARR	Pilots on board	20

The table below provides an overview of the estimated effects of the new system on the port call process in the defined scenario.

Table 2 Overview of findings with comments

Parameter	"As-is" process	"To-be" process	Estimated effect of new system	Comment
Total time for communications	686 sec	440 sec	246 sec time reduction	Mostly related to pre-planning/system config allowing for "clicking" on predefined status situations, and that writing messages into the system takes shorter time than talking on VHF.
No of communications	31	28	3 comms tasks removed	These three comms are all related to pilots discussing status on the equipment problem (on VHF ch. 10). The new system allows to share this by writing instead.

5. Evaluating the actual effects of reduced time and number of communications related to the Port Call scenario

From the above, there are two leading indicators related to predicting the effects of the SESAME S2 solution related to a port call. The hypothesis is that a reduction in time spent and focus on administrative tasks, will result in a higher awareness at the Vessel Traffic Service (VTS), thereby reducing the risk for incidents, accidents and inefficient operations. This is not a very controversial hypothesis. Nevertheless, it is still a hypothesis and actions must be taken to assess its validity through introducing lagging indicators.

The basis for the lagging indicators in our case, is the Baltic and International Maritime Council' Shipping KPI Standard (BIMCO, 2022). This standard was developed to report, monitor and express Health, Safety, Environment and Quality performance (HSEQ) for ship management. The BIMCO database currently contains KPI-data for appr. 8000 ships, reported quarterly since 2011[a]. The KPIs are lagging by nature, expressing actual performance within the HSEQ-segment of shipping.

The figure below depicts how the KPIs from BIMCO are mapped to the foreseen solutions developed in SESAME S2.

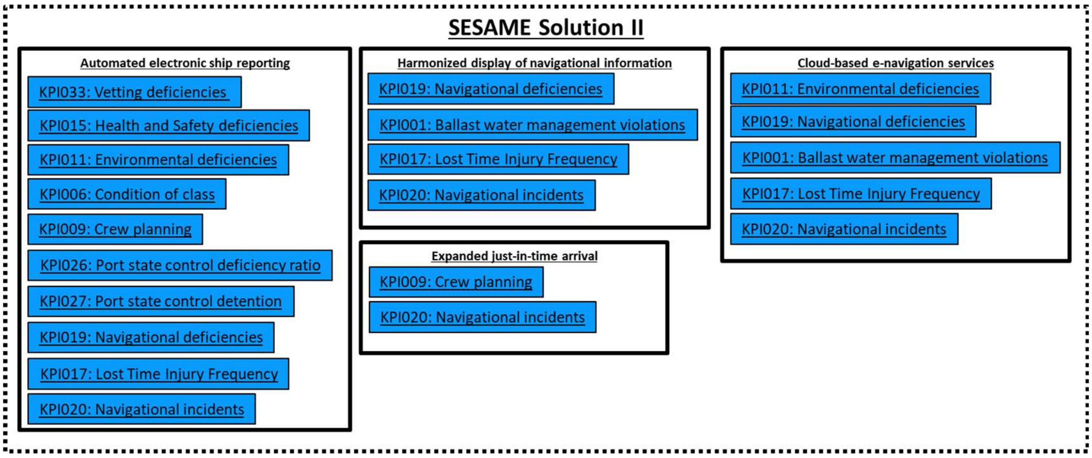

Figure 3 Mapping of BIMCO Shipping KPIs to new SESAME S2 solutions

[a] Not all 8000 ships have been reporting since 2011

The idea is that once a certain solution from SESAME S2 is actually implemented, the results should be visible through the performance-trends of KPI Values from the BIMCO database, hereby confirming (or indeed rejecting) the validity of the leading indicators related to time savings and number of communications. By mapping the KPIs from the BIMCO Shipping KPI standard to the solutions to be developed in SESAME S2, we can collect values on these KPIs both prior to and after the SESAME S2-implementation, enabling comparison of pre- and post-implementation quality of service. As the BIMCO Shipping KPI standard entails a set of well-defined, industry-wide KPIs, the quality of these KPIs and the potential to aggregate KPI values to industry averages, is high. The effects of the SESAME S2 solution will be visible through time, as one of three alternatives: 1) The HSEQ KPIs show improved performance, 2) The HSEQ KPIs show deteriorated performance or 3) The HSEQ KPIs show no change:

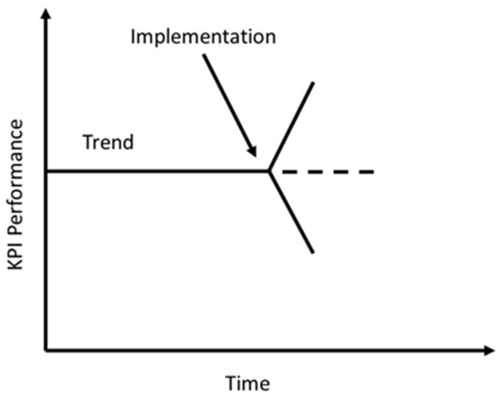

Figure 4 The relationship between new solutions in E-nav and actual HSEQ-performance in shipping

5.1. How to read and interpret the HSEQ KPIs (lagging indicators) to assess the validity of the leading indicators

There are three main alternatives when reading and interpreting the lagging indicators, namely as a snapshot, as a trend or as a benchmark.

A snapshot expresses the HSEQ-performance for a single quarter, either directly after the SESAME S2-solution has been implemented or after some time, allowing for the users to grow acquainted with the new solution. A trend expresses the gradual change of the HSEQ-performance over time. A trend expresses a more accurate picture of performance as it will disregard spurious peaks or falls in performance. One can also create trends for entities before and after a SESAME 2 solution has been implemented. A benchmark compares the ship or port in question, to comparable entities. A suitable set of selection criteria enables comparison in HSEQ-performance between entities subject to the SESAME 2 solution with entities not subject to the solution.

For our purpose, both trending and benchmarking seem most usable when assessing the validity of the leading indicators. If a digitalized VTS communication solution from SESAME S2 is implemented at the port of Humber, what are the actual consequences over time? Are the number of near misses reduced? How about navigational incidents? Do we see a clear trend of reduced number of incidents over time, once the SESAME 2 solution has been implemented? If so, the leading indicators related to reduced number of communications and reduced time spent on communicating orally can be said to be highly valid. This is however not a clear-cut case. What if the reduced number of incidents is due to a higher awareness between the stakeholders, which could be a normal consequence when new solutions are implemented? What happens when the stakeholders realize that the "extra spare time" may be used for other duties? Are the results valid for all types of ships and all types of crews? Non-native English speakers may benefit more from written communications than native English speakers, for example, as the potential for improved understanding is higher for non-native speakers. Some of these questions will be answerable once additional parameters are

Proceedings of the 32nd European Safety and Reliability Conference
Edited by Maria Chiara Leva, Edoardo Patelli, Luca Podofillini, and Simon Wilson

ISBN: 981-973-0000-00-0 :: doi: 10.3850/981-973-0000-00-0_esrel2022-paper

considered: Crew nationality, ship types, daytime versus night-time operations, etc. This in addition to continue monitoring of the performance trend to disregard initial effects as well as including later effects.

6. Concluding remarks

Assessing the potential value of e-navigation solutions will by nature, always have to be based on hypotheses where the features of said solutions are linked to a set of leading indicators. Typically, these hypotheses relate to predictions such as increased efficiency, effectiveness, safety, environmental sustainability and/or quality of operations. The actual results from implementing the e-navigation solutions will not be available before we are able to assess the (improved) quality of the corresponding operations. To break outdated or wrongful assumptions with actual proof in a continuous loop of improvement should be an integral part of the implementation of new solutions in the maritime industry.

Our case study shows that through experiments and observations, we are able to assess a highly relevant set of leading indicators related to reduced time consumption. This on basis of the hypothesis that time saved in communication between the ship and the VTS, will result in a lower safety incident rate. Nevertheless, these indicators are not expressions of actual performance in terms of quality of operations. We are yet to see whether the time saved will result in safer navigation as this is e.g., highly dependent on what the extra time is used for. A proper set of lagging indicators where hypotheses can be validated is crucial, both in terms of assessing whether the hypothesis is valid and whether adjustments are needed to fully reap the benefits. These indicators should express the long-term effects, on an industrial level and not represent a mere snapshot. Utilization of already internationally accepted and reported KPIs, such as the BIMCO Shipping KPI standard is therefore recommended.

Acknowledgement

This work was funded by the Sesame Solution II project (The Research Council of Norway, Project# 282331).

References

G.W.F. Hegel (1807), Phenomenology of Spirit, trans. A.V. Miller, London, Oxford, 1977.

International Association of Marine Aids to Navigation and Lighthouse Authorities (2016) Planning and reporting of e-navigation testbeds. https://www.iala-aism.org/content/uploads/2017/03/1107-Ed2-on-Planning-Testbeds-and-Reporting-of-Testbed-Results-Ed-2-June-2016-1.pdf (Accessed March 2022)

International Maritime Organisation (2015) Guideline on software quality assurance and human-centred design for e-navigation. https://wwwcdn.imo.org/localresources/en/OurWork/Safety/Documents/enavigation/MSC.1-Circ.1512%20-%20Guideline%20On%20Software%20Quality%20Assurance%20And%20Human-Centred%20Design%20For%20E-Navigation%20(Secretariat)%20(1).pdf (Accessed March 2022).

Baltic and International Maritime Council (2020) The Shipping KPI Standard V4.0. https://www.shipping-kpi.org/ (Accessed March 2022).

M. Baldauf, S.B. Hong (2016) Improving and Assessing the Impact of e-Navigation applications, International Journal of e-Navigation and Maritime Economy, Volume 4, Pages 1-12, ISSN 2405-5352.