NTIS#PB2024-

SSC – 481

DIGITAL TWIN METHODOLOGIES FOR THE INTEGRATION OF HULL MONITORING SYSTEMS WITH PHYSICS-BASED MODELS TO SUPPORT SHIP OPERATIONS AND SERVICE LIFE EXTENSIONS

Ship
Structure
Committee

2025

This page intentionally left blank

Member Agencies:

American Bureau of Shipping
Defence Research and Development Canada
Maritime Administration
Military Sealift Command
Naval Sea Systems Command
Office of Naval Research
Society of Naval Architects & Marine Engineers
Transport Canada
United States Coast Guard

Ship
Structure
Committee

Address Correspondence to:

COMMANDANT (CG-ENG-2/SSC)
ATTN (EXECUTIVE DIRECTOR/SHIP
STRUCTURE COMMITTEE)
US COAST GUARD
2703 MARTIN LUTHER KING JR. AVE SE
WASHINGTON DC 20593-7509
Website: http://www.shipstructure.org

SSC – 481
SR – 1482

DIGITAL TWIN METHODOLOGIES FOR THE INTEGRATION OF HULL MONITORING SYSTEMS WITH PHYSICS-BASED MODELS TO SUPPORT SHIP OPERATIONS AND SERVICE LIFE EXTENSIONS

Digital Twins have garnered significant interest due to their potential to revolutionize the testing of upgrades, maintenance practices, and the development of a comprehensive ship sustainment system that improves fleet performance and reduces operational costs. However, despite the promising capabilities of Digital Twins, challenges related to cost, scope, data integration, tools, and practical application complicate their development and broader adoption.

This report addresses these challenges by proposing a structured framework for the creation of Structural Digital Twins (SDT) that balances technical goals with considerations for resources, costs, and accuracy. It explores three practical SDT methods, using data from a hypothetical, full-scale naval vessel, to demonstrate how SDTs can be effectively implemented. The report also highlights improvements to digital twin methods through SDT frameworks, real-world SDT examples, and key findings. Additionally, recommendations for future studies are provided to address current limitations and maximize the potential of SDTs in the naval and maritime industries.

We thank the authors and Project Technical Committee for their dedication and research toward completing the objectives and tasks detailed throughout this report and continuing the Ship Structure Committee's mission to enhance the safety of life at sea.

W. R. ARGUIN
Rear Admiral, U.S. Coast Guard
Co-Chairman, Ship Structure Committee

for
P. D. SMALL
Rear Admiral, U.S. Navy
Co-Chairman, Ship Structure Committee

SHIP STRUCTURE COMMITTEE

RDML Wayne Arguin
U. S. Coast Guard Assistant Commandant
for Prevention Policy (CG-5P)
Co-Chair, Ship Structure Committee

RDML Peter Small
Chief Engineer and Deputy Commander
for Naval Systems Engineering (SEA05)
Co-Chair, Ship Structure Committee

Mr. Jeffrey Lantz
Director, Commercial Regulations and Standards (CG-5PS)
U.S. Coast Guard

Mr. Gareth Burton
Vice President, Technology
American Bureau of Shipping

Mr. H. Paul Cojeen
Society of Naval Architects and Marine Engineers

Dr. Malcolm Smith
Head, Warship Performance, DGSTCO
Defense Research & Development Canada - Atlantic

Mr. Kevin Kohlmann
Director, Office of Safety
Maritime Administration

Mr. Luc Tremblay
Executive Director, Domestic Vessel Regulatory Oversight
and Boating Safety, Transport Canada

Mr. Michael Haycock
Deputy Assistant Commandant for Engineering and
Logistics (CG-4D)
U.S. Coast Guard

Mr. Mark Lattner
Group Director, Ship Integrity and Performance Engineering
(SEA 05P)
Naval Sea Systems Command

Mr. Neil Lichtenstein
Deputy Director N7x, Engineering Directorate
Military Sealift Command

Dr. Thomas Fu
Director, Ship Systems and Engineering Research Division
Office of Naval Research

SHIP STRUCTURE EXECUTIVE GROUP & SUB-COMMITTEE

UNITED STATES COAST GUARD (CVE)
CAPT Jennifer Doherty
Mr. Charles Rawson

AMERICAN BUREAU OF SHIPPING
Dr. Qing Yu
Mr. Daniel LaMere
Ms. Christina Wang
Mr. Rich Delpizzo

SOCIETY OF NAVAL ARCHITECTS AND MARINE ENGINEERS
Mr. Frederick Ashcroft
Dr. Roger Basu
Dr. Robert Sielski
Dr. Paul Miller

DEFENSE RESEARCH & DEVELOPMENT CANADA ATLANTIC
Dr. Mahmud Sazidy
Mr. Cameron Munro

MARITIME ADMINISTRATION
Mr. Todd Hiller
Mr. Todd Ripley
Mr. Digvijay Singh

TRANSPORT CANADA
Ms. Dorianne Aubin
Mr. Bashir Ahmed Golam
Ms. Tayyebe Seif

UNITED STATES COAST GUARD (FLEET)
CAPT Doug Graul
Mr. Martin Hecker
Mr. Timothy McAllister
Mr. Debu Ghosh

NAVSEA/ NSWCCD
Mr. David Qualley
Mr. Dean Schleicher
Dr. Pradeep Sensharma
Mr. Michael McDonald

MILITARY SEALIFT COMMAND
Ms. Jeannette Viernes

OFFICE OF NAVAL RESEARCH
Dr. Jessica Dibelka

PROJECT TECHNICAL COMMITTEE MEMBERS

The Ship Structure Committee greatly appreciates the contributions of the individuals that volunteered their time to participate on the Project Technical Committee, listed below, and thanks them heartily. They were the subject matter expert representatives of the Ship Structure Committee to the contractor, performing technical oversight during contracting, advising the contractor in cognizant matters pertaining to the contract of which the agencies were aware, and performing technical peer review of the work in progress and upon completion.

Chair(s):
Dr. Paul Hess, StructRel Solutions
Dr. Jessica Dibelka, Office of Naval Research

Members:
Dr. Pradeep Sensharma, NAVSEA
Dr. Robert Sielski, Society of Naval Architecture and Marine Engineers
Dr. Pramod Kumar Dash, Nitte Meenakshi Institute of Technology
Mr. Chris Page, Noise Control Engineering
Mr. Neil Pegg, Defense Research & Development Canada
Dr. Suhas Vhanmane, Indian Register of Shipping
Ms. Amanda Blumenthal, NSWCCD
Mr. Matt Grassman, NSWCCD
Mr. Ian Thompson, Defense Research & Development Canada
Dr. John Mackay, Defense Research & Development Canada
Mr. Aimin Wang, Exmar Offshore Company
Mr. Jerzy Bogdziewicz, Marinus LLC.
Dr. Yikun Wang, University of Southampton
Dr. Geoffrey Cranch, Naval Research Laboratory
Mr. Dale Pederson, Herbert Engineering, UK
Mr. Gregory Grigoropoulos, Society of Naval Architecture and Marine Engineers
Mr. Rong Huang, Society of Naval Architecture and Marine Engineers
Mr. Boyden Williams, Viking Systems Inc.
Dr. Matthew Collette, University of Michigan
Dr. Haitao Liu, American Bureau of Shipping
Mr. George Wang, Society of Naval Architecture and Marine Engineers

Ship Structure Committee Executive Director:
LCDR Bryan Andrews, U.S. Coast Guard

Technical Report Documentation Page

1. Report No. SSC-481	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Digital Twin Methodologies for the Integration of Hull Monitoring Systems with Physics-Based Models to Support Ship Operations and Service Life Extensions		5. Report Date 3 June 2025 6. Performing Organization Code
7. Author(s) Alysson Mondoro, William Whitmore, Nathan Nelson, and Isaac DiNapoli		8. Performing Organization Report No. SR-1482
9.Performing Organization Name and Address Naval Surface Warfare Center Carderock Division (NSWCCD) 9500 MacArthur Blvd. West Bethesda, MD 20817-5700		10. Work Unit No. (TRAIS) 11. Contract or Grant No.
12. Sponsoring Agency Name and Address Ship Structure Committee C/O US Coast Guard 2703 Martin Luther King Jr. Ave. SE Washington, DC 20593-7509		13. Type of Report and Period Covered Final Report 14. Sponsoring Agency Code CG-5PS
15. Supplementary Notes		

16. Abstract

Digital Twins have garnered significant attention in the marine and naval sectors with ambitions to evolve into a comprehensive ship sustainment system that enhances fleet performance and reduces costs. Despite their praised versatility and potential, Digital Twins face challenges related to cost, scope, data, tools, and practical usefulness, complicating their development and adoption. This paper addresses these challenges by proposing an ontology for the design and development of Structural Digital Twins (SDT) that achieve specific objectives practically, considering resource, cost, and accuracy factors. It investigates three functional and practical SDT approaches using data from an instrumented, full-scale, notional naval vessel. The report is divided into three parts: advancements to digital twin methodologies through SDT ontology, practical examples of SDTs with real datasets, and a summary of key findings with suggestions for future research.

17. Key Words Digital Twins, Structural Digital Twins, Ontology	18. Distribution Statement Approved for public release; distribution is unlimited. Available from: National Technical Information Service Springfield, VA 22161 (703) 487-465		
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 135	22. Price

Conversion Factors
(Approximate conversion to metric measures)

To convert from	to	Function	Value
LENGTH			
inches	meters	divide	39.3701
inches	millimeters	multiply by	25.4000
feet	meters	divide by	3.2808
VOLUME			
cubic feet	cubic meters	divide by	35.3149
cubic inches	cubic meters	divide by	61,024
SECTION MODULUS			
inches^2 feet	centimeters^2 meters	multiply by	1.9665
inches^2 feet	centimeters^3	multiply by	196.6448
inches^3	centimeters^3	multiply by	16.3871
MOMENT OF INERTIA			
inches^2 feet^2	centimeters^2 meters^2	divide by	1.6684
inches^2 feet^2	centimeters^4	multiply by	5993.73
inches^4	centimeters^4	multiply by	41.623
FORCE OR MASS			
long tons	tonnes	multiply by	1.0160
long tons	kilograms	multiply by	1016.047
pounds	tonnes	divide by	2204.62
pounds	kilograms	divide by	2.2046
pounds	Newtons	multiply by	4.4482
PRESSURE OR STRESS			
pounds/inch^2	Newtons/meter^2 (Pascals)	multiply by	6894.757
kilo pounds/inch^2	mega Newtons/meter^2 (mega Pascals)	multiply by	6.8947
BENDING OR TORQUE			
foot tons	meter tons	divide by	3.2291
foot pounds	kilogram meters	divide by	7.23285
foot pounds	Newton meters	multiply by	1.35582
ENERGY			
foot pounds	Joules	multiply by	1.355826
STRESS INTENSITY			
kilo pound/inch^2 inch$^{1/2}$(ksi$\sqrt{}$in)	mega Newton MNm$^{3/2}$	multiply by	1.0998
J-INTEGRAL			
kilo pound/inch	Joules/mm^2	multiply by	0.1753
kilo pound/inch	kilo Joules/m^2	multiply by	175.3

SSC Report 1482

Digital twin methodologies for the integration of hull monitoring systems with physics-based models to support ship operations and service life extensions

Alysson Mondoro, William Whitmore, Nathan Nelson, and Isaac DiNapoli

2-12-2024

Table Of Contents

List of Figures

List of Tables

1. Executive Summary

Digital Twins have been investigated and used in the marine and naval sectors already to test out upgrades and maintenance options, with the hopes to expand to a more comprehensive ship sustainment system, improving fleet performance, and saving time and money. While Digital Twins are lauded for their vast applicability and potential capabilities, the practical concerns associated with cost, scope, data, tools, and usefulness make the path for development and adoption murkier. This report is intended to address these concerns by proposing ontology for the design and development of a structural digital twins (SDT) that meet a specific objective in a practical manner, considering all of the resource, cost, and accuracy considerations. By proposing a hierarchical approach to SDT objectives, the report helps clarify the complex and often ambiguous goals of SDTs, making them more manageable and focused on practical outcomes.

A key focus of the report is the identification and mitigation of uncertainties in SDTs, which arise from environmental variables, operational conditions, and the inherent limitations of models. The report highlights the need for improved definitions and more detailed guidelines on incorporating these uncertainties into SDT systems, ensuring they provide valid, reliable decision support for operators. It also emphasizes the importance of developing clear, performance-based assessment procedures tailored to different vessel types and their respective risk tolerances.

To address these challenges, the report proposes a logic tree framework for integrating performance-based assessments with uncertainty considerations, ensuring that SDTs can respond effectively to both steady and abrupt changes in operational conditions. It also underscores the need for robust validation processes, calling for the development of standardized testing methods to ensure the accuracy and reliability of SDTs.

Finally, the report presents practical examples of SDTs applied to real-world data from surface ships, providing different approaches, including surrogate modeling, advanced finite element modeling, and predictive methods. These examples highlight both the strengths and limitations of current SDT technologies and identify critical areas for further research and development.

2. Introduction

Since the early 2000’s, the concept of the digital twin (DT) has emerged as a critical technology for use in industries ranging from structural engineering to biomedical engineering and on. For naval and marine structures, digital twins are discussed in their potentials for use in safe and efficient operations of ships, optimal maintenance for ships and fleets, and optimal route planning for vessels. The goals of digital twin for ship structures distills out to objectives of optimizing performance, increasing availability, and reducing costs. The “how” is less straightforward.

The topic of digital twin is both so broad and so deep that the argument could be made that it is not a fundamental concept but the bridging across fundamental concepts enabled by the digital revolution. This characterization of digital twin as a ‘compilation of capabilities’ understates the challenges present in the topic. That each of the fundamental concepts almost always come with caveats, assumptions, limitations, future work, and that bringing them together is ripe for inconsistencies and self-contradictions, and un-tractable solutions chasing after perfection. However, this is exactly the role of the digital twin: it lives in reality; with all the uncertainties and errors, incorporating the best knowledge of the physical structure, physics, modeling and simulation, and prognostics to support decisions.

This report down scopes digital twin for surface ships further to focus on the structural components, herein referred to simply as ‘structural digital twin’ (SDT). The report starts with ontological questions regarding what digital twins are and proposes concepts to aid the designer and developer to work towards a viable, tractable solution. These concepts are covered in sections 3 through 7. Section 8 takes a brief aside to discuss the needs for validation for digital twins. And practical examples for SDTs are presented in section 9, along with their results and findings.

3. Structural Digital Twin Overview

This is the point where a paper would typically provide a simple, clear, one sentence description of the topic being discussed. For structural digital twin for naval and marine vessels, there is nothing simple nor clear about it. The section highlights the nuances of SDTs in an effort to enable the coherent design process for an SDT.

The following subsections touch on the aspects of “what” is a structural digital twin, including: “what” is the objective of an SDT and “why” are they relevant, “where” digital twins exist and in what form, “how” they work, and the “concerns” there are in the industry regarding the twin.

3.1. What and Why: Objectives of a Structural Digital Twin

The question “what is a digital twin” is almost always where people start their conversations, journal articles, or presentations about digital twin. But, invariably, the first talking point of the papers focus on the question “why”. Thus, this section addresses the what and the why.

Structural Digital Twins are at the forefront of development for naval and marine applications. For some, the interest lies in the philosophy of it: that a twin has the potential to be the convergence of data, knowledge, and reality to support the physical asset. For example, Glaessgen (2012) refers to the twin as:

> “A Digital Twin is an integrated Multiphysics, multiscale, probabilistic simulation of an as-built vehicle or system that uses the best available physical models, sensor updates, fleet history, etc., to mirror the life of its corresponding flying twin. The Digital Twin is ultra-realistic and may consider one or more important and interdependent vehicle systems, including airframe, propulsion and energy storage, life support, avionics, thermal protection, etc. The extreme requirements of the Digital Twin motivate the integration of design of materials and revolutionary approaches for material processing. Manufacturing anomalies that may affect the vehicle are also explicitly considered, evaluated, and monitored. In addition to the backbone of high-fidelity physical models of the as-built structure, the Digital Twin

integrates sensor data from the vehicle's on-board integrated vehicle health management (IVHM) system, maintenance history and all available historical and fleet data obtained using data mining and text mining" (Glaessgen, 2012).

For others, the interest lies in the availability of resources; that since the digital capabilities exist, they should be capitalized on to best support the physical asset:

"The DoD is actively undertaking the Industry 4.0 paradigm shift towards the use of large data sets to provide previously unobtainable, real-time insights into numerous systems, processes, or assets, the 2018 Digital Engineering Strategy (DES) serves to "guide the planning, development, and implementation of the digital engineering transformation across the DoD." (US Department of Defense , 2018)

Together, these put forth the drive towards digital twins. This generally describes "what" a digital twin is, but to get to a functional design and practical instantiation of a digital twin, the "why" (or objective of the twin) is imperative.

Researchers have put forth objectives for structural digital twins, or "why". This includes reducing maintenance costs, reducing risk in maintenance and operations, improving efficiency of operations of the vessel or of the fleet, and increasing availability, reliability, and resilience (VanDerHorn and Mahadevan, 2021). It also includes optimizing for elements of re-usability, interoperability, interchangeability, maintainability, extensibility, and autonomy across the entire lifecycle (Moyne et. al., 2020). Alternatively, objectives like decision-making support, cost reduction, remote control and monitoring, maintenance, condition monitoring, testing and simulation, and training personnel have been included (Assani et. al., 2022). Optimal route planning has also been posed as an objective for a digital twin (Lee et. al., 2022). In other discussions on digital twins, value creation is identified as being based on the actor-to-actor interactions for a particular solution, making it more complex to identify the objectives (or requirements) (West et. al. 2021). Complexity is key.

To form clear and coherent objectives, the complexity of the objective, the actor-to-actor interactions, and the broader context of decision making is critical. Thus, this paper proposes that the objective for a structural digital twin is, in fact, defined with a hierarchy of objectives as shown in Figure 1; the hierarchy is grouped into different levels:

- Vision: this is the highest level that expresses what the twin wants to be. It is the high-level purpose. For structural digital twins, this includes increasing the availability of the fleet, reducing costs for operations and maintenance, and optimizing performance.
- Strategic: this is the middle level that includes a targeted mechanism for meeting the vision. For structural digital twin, examples of this include supporting the operator via operator guidance and route planning, supporting the maintainer and planner for condition-based maintenance and fleet asset planning, and supporting designers to optimize the design.
- Operational: this is the lowest level, where the objectives are scoped to support the development of a practical instantiation of a structural digital twin. This is where the objective is scoped out for basic requirements and constraints including the complexity of the analysis needed, the planning time-horizon, the resources available, the flexibility of the vessel to stay/change course due to mission criticality, and the level of integration with the decision-making process, among others.

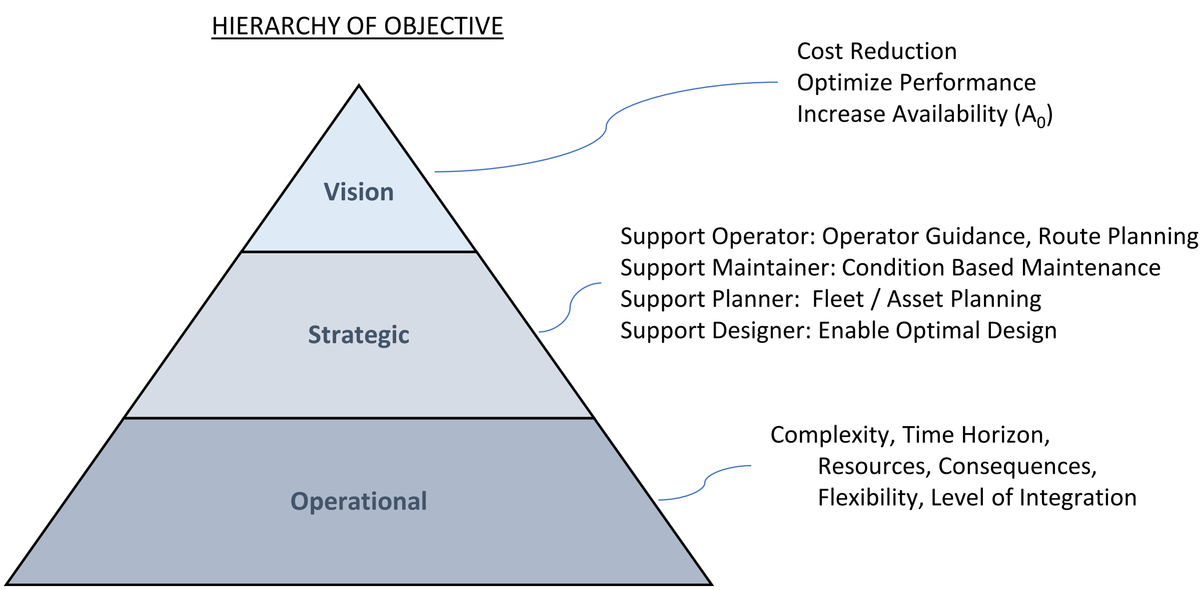


Figure 1. Hierarchy of Objectives as Applicable to Structural Digital Twins

The operational objective refines the scope of the strategic objective. For structural digital twin, this is essential for the tractable development of a functional product. Operational objectives include:

- Timeliness (Real Time, Near-Real Time, Time-Delayed): when is the decision support needed. If information and decision support is needed in real time, the complexity of the digital twin must be designed to support that objective. If it is near-real time, a different complexity of twin may be warranted given the trade-off between time and accuracy. If time-delayed decision support is warranted (as may be the case for condition-based maintenance recommendations that may be given after a deployment), the twin can take a different form.
- Use (Operational Guidance, Condition Based Maintenance, Route Planning, Fleet Planning): what is the decision that the information is intended to support.
- Time Horizon (Instantaneous, Near Future, Transit, Deployment): what is the required window of time that the data must account for.
- Consequences: Is the vessel manned, unmanned, autonomous, and operating as a part of a set?
- Operations: Is the vessel operating in routine conditions or critical conditions? Critical conditions may include operating in a storm, or critical missions or wartime for naval vessels. The twin may need to be flexible to support both.
- Level of Integration (Support Operator, Human-in-the-loop, full autonomy): How redundant, aware, capable of handling uncertainty, does the twin have to be?

An example between a strategic objective and an operational objective that both support vision objective of increasing operability is shown in Figure 2.

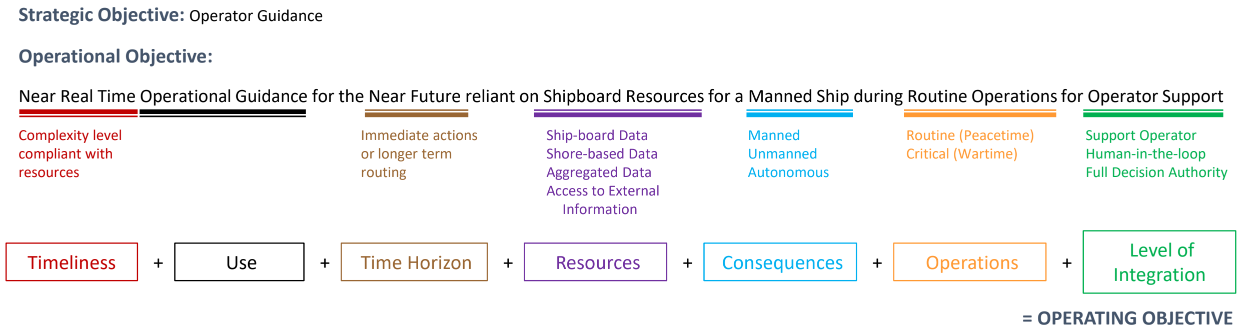


Figure 2. Operational Objective Decomposed

In the development of the above structural digital twin objectives definition, a review of class standards was performed. Like in research, there is high interest in the class societies to expand into the digital domain via digital twins (or “smart ships”, “intelligent ships”, and “condition monitoring”). A comparison of the class societies’ guidances and rules is provided throughout this document in the context of the chapter itself. Table 1 summarizes the class standards and the objectives and description provided. The classes have tended to break out guidances for Hull Monitoring (or similar title) and Smart Vessels (or similar title) when discussing concepts related to structural digital twin. The objectives and description are brief descriptions and/or excerpts from the referenced document. The Hull Monitoring documents tend to focus on ‘how’ the objective can be supported, with minimal discussion on the vision objective or the operation objective. The Smart vessel documents tend to focus on the vision objective, sometimes tying in flavors of the operational objective. One thing is apparent from the review of the standards: a full enumeration of the objective of the digital twin is not provided. This is not to say that this lack is a flaw, but that gap is (i) a statement on the complexity of the objectives (as indicated in the preceding paragraphs), (ii) a statement on the readiness of the field for standardization, (iii) both above, (or, as the authors are human, (iv) something the authors have not considered).

Table 1. Class Standards for Condition Monitoring and "Smart" Vessels

Class Standard	Objective/Description
ABS Hull Condition Monitoring (ABS, 2020)	Hull Condition Monitoring (HCM) is to monitor, visualize, and trend parameters relevant to environment, structural loads, and responses through sensor-based measurements. HCM typically involves onboard and/or onshore reporting and threshold-based alarms for operational guidance and post-voyage analysis.
ABS SMART (ABS, 2022)	"Provide the crew and support personnel with key information to aid in decision making. Common Smart Functions include structural and machinery health monitoring, asset efficiency monitoring, operational performance management, and crew assistance and augmentation to support vessel operations. The difference between Structural Health Monitoring (SHM), as a Smart Function, and the traditional HCM is primarily: SHM provides structural health diagnostics and prognostics through correlation of various parameters and integration with analysis and simulation. HCM handles parameter-based monitoring and covers the loads, responses, and identifiable damages from direct sensor measurements at certain sensor installed locations."
Rules for the Classification of Steel Ships, NR467 - JULY 2022, Part F, Additional Class Notations (BV, 2022)	Hull Monitoring System is a system which: Provides real-time data to the Master and officers of the ship on hull girder longitudinal stresses and vertical accelerations the ship experiences while navigating and during loading and unloading operations in harbor. Allows the real-time data to be condensed into a set of essential statistical results. The set is to be periodically updated, displayed, and stored on a removable medium.
BV NR675 Additional Service Feature SMART (BV, 2021)	A smart system is defined as a computer-based system that incorporate functions for the collection, the transmission, the analysis and the visualization of data. A function is a defined objective or characteristic action of a system or component. Smart functions may include monitoring, decision making support, remote monitoring, maintenance. Hull smart functionalities dedicated to operation: LI-S3, LI-S4, LI-HG-S3 or LI-HG-S4 Hull smart functionalities dedicated to hull maintenance: MON-HULL
DNVGL-RU-SHIP Pt.6 Ch.9 (DNV GL, 2017)	The system shall give warning when stress levels and the frequency and magnitude of ship accelerations approach levels that require corrective action. The owner shall decide how the hull monitoring system should be configured, i.e., which features to be included and how the measured and processed data shall be use as intended as an aid to the master's judgement and not as a substitute.
DNV GL Smart Vessel (DNV GL, 2020)	Use data and information to further optimize vessels' operations and reduce the environmental footprint. Operation and maintenance - hull and structure (OPH) enhancements include solutions that use data as an important element and provide options related to structural integrity management
Lloyds Register, ShipRight, Ship Event Analysis (Llyods Register, 2021)	Provide warning the ship's personnel that these stress levels or the frequency and magnitude of slamming motions are approaching a level where corrective action is advisable

Class Standard	Objective/Description
CCS, Rules for Intelligent Ships (CCS, 2020)	To provide assistant decision-making for hull and deck machinery maintenance and structural renewal during in-service period of the ship based on the establishment and maintenance of hull database system and three-dimensional hull structural models. The hull monitoring and assistant decision-making system is to include the following functions related to structures. - Development of hull inspection and maintenance scheme - Recording and assessment of hull structural conditions - Development of structural renewal plan Hull database system is to be able to integrate data of three-dimensional hull structural models, hull and deck machinery inspection and maintenance data, structural thickness measurement data and structural repair data.
Class NK (Class NK, 2020)	To monitor the behavior of hull girders during navigation, loading and unloading, and to provide real-time information on stress levels due to longitudinal bending moments and acceleration levels due to ship motion. Information is to be intended to aid the judgment of Shipmasters and crew members during navigational operations, it is not intended to be a substitute for the judgment and the responsibility of Shipmasters.

3.2. Where: Resources and Location

One of the major questions for structural digital twins, is “where does the twin exist”. Barring the natural, simple answer of “in the digital environment”, the response tends to be divided amongst shipboard or land based. The latter is the focus of the remainder of this section. The former is, in fact, not a simple answer as the digital environment is in development for some organizations or in flux for others.

Table 2 presents the breakdown on location, data and computational resources, and time-delay considerations as it pertains to the strategic objective of operational guidance support. The concepts can be generalized to the other objectives, but it is useful to track through the specific example. For the strategic objective of operational guidance support, the structural digital twin has the potential to be on the ship or on land with communication to and from ship. In both instances, the ship operator is the entity acting on the decision.

In the instance where the twin is a shipboard twin, it is most likely computationally limited and may not have access to datasets external to the ship (since connectivity may not always be present). However, the time delay between the twin getting data, making assessments, developing decision support findings, and delivering the information to the operator, is minimal. For the instance where the twin is land-based, the computational capacity is most likely more advanced, and there is access to external data sets. However, there may be a significant delay in providing decision support to the operator.

The largest concern that has been identified by researchers and class societies alike is the sense of agency (Woolley et. al., 2023). The digital twin provides decision support. For manned and human-in-the-loop systems, agency lies with the human, and they have authority (and almost the mandate) to take in additional information (such as information from other ship systems and qualitative information on ship performance) and make the final decision.

The concern over agency grows more when the digital twin is land based. It is often that the engineering support activity may, in other paradigms, have the authority to direct decisions. However, the shift to the digital twin paradigm and the existence of data, both quantitatively (other ship systems) and qualitatively (personnel and awareness on performance) means that the decision-making chain of command must be clearly established.

Table 2. Location and Resource Discussion for Operational Guidance Support

Location (general)	Shipboard	Land-Based
Location (specific)	Ship	Engineering Support Activity
Resources: Data	Data from Digital Twin	Data from Digital Twin*
	Summary Data from Digital Twin	Summary Data from Digital Twin
	Data from Other Ship Systems	-
	Qualitative Information	-
	-	External Data Sets (Environmental Forecasts, History, etc.)
Resources: Computational	Computational Capacity (limited)	Computational Capacity (advanced)
Resources: Infrastructure	Wired/Wireless on Ship	Digital Communication System
Time-delay	Minimal	Minimal to Substantial
* Depending on connectivity and digital infrastructure		

There is also the case where the twin can exist in both places, shipboard and land based. An example of this may be where there are two difference strategic objectives: one for providing operational guidance and one for developing condition-based maintenance plans. A single digital twin may not be feasible given different computational resource available and/or availability of data. As such, there may be nested digital twins, both stemming from the base twin, but adapted to meet the operational level objectives of the twin. The following describes the locations and differences in resources for the nested digital twin:

The structural digital twin established to support the two different operational level objectives may be a set of nested digital twins. This is shown visually in Figure 3 where a complex digital twin is developed and then a simple surrogate is developed out of it and transferred to the ship. Each twin supports the different operation-level objectives but stems from the same core. The nested set of twins will also have access to different data over time. The shipboard twin continuously has access to the operational data (strains, accelerations, speed, etc.) that it is acquiring. It may have the opportunity to offload information to the shore, but depending on the transfer capabilities, the size and type of data may vary. Bandwidth limits may mean that only down-sampled data or summary data are exported from ship to shore. The complex model may then be updated, and a revised surrogate developed and potentially pushed back as an update to the shipboard system. When the ship arrives in port, more information can be transferred to and from ship, and the nested digital twins can evolve.

Nested twins often vary in their access to data and the computational resources at their disposal. The digital infrastructure is then relied on to host the twins and support data and model transfers. For naval and marine applications, this is a critical component since connectivity of the vessel to shore and the capability of sending and receiving data may vary based over time.

The design of the SDT must consider the available infrastructure for shipboard communication and ship-to-shore connectivity. Wired and Wireless systems have different information security vulnerabilities, in addition to cost and feasibility of installation. Likewise, connectivity to the shore and bandwidth for transferring data must be considered when designing the SDT.

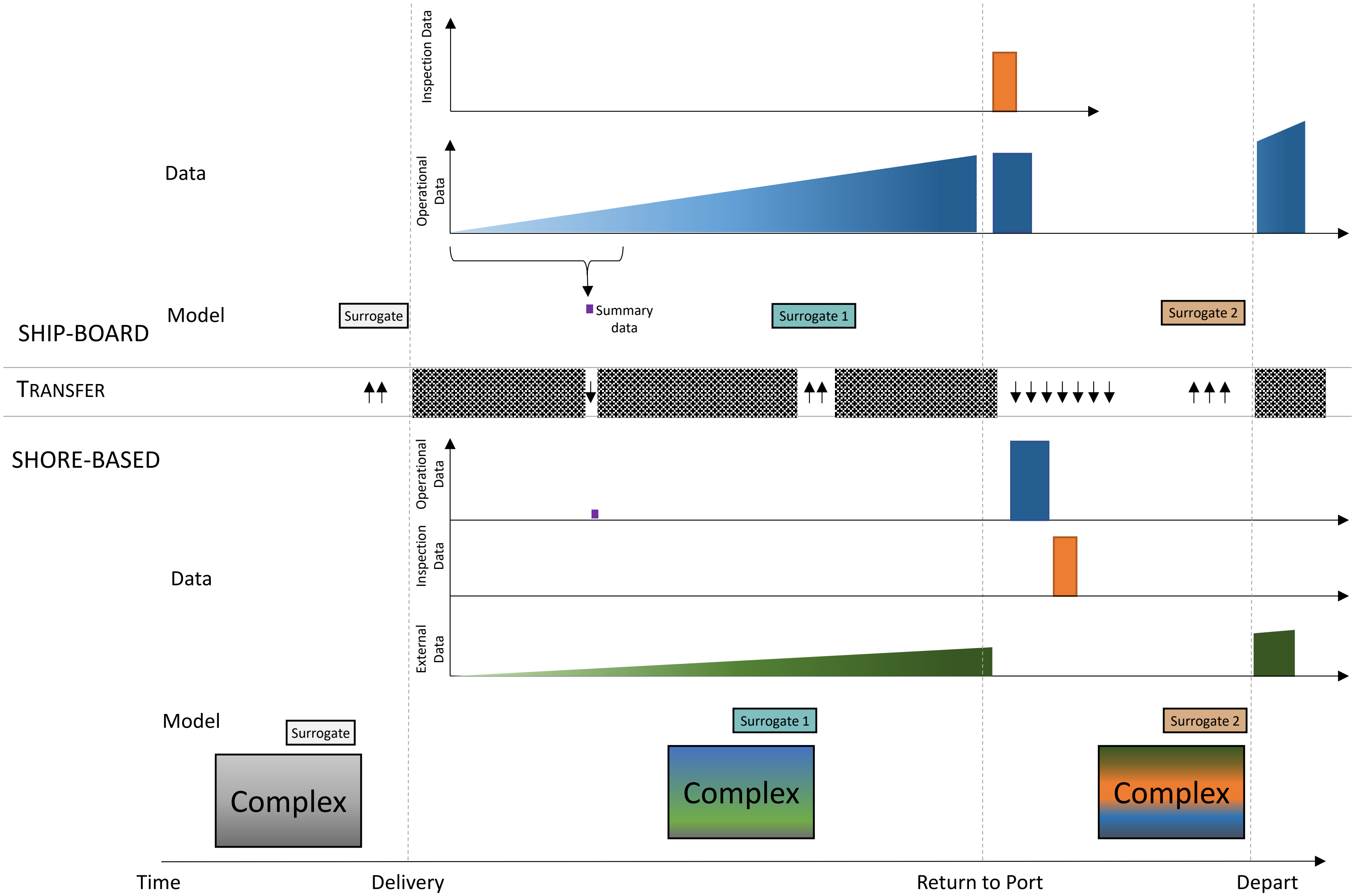


Figure 3. Nested Digital Twin Example

3.3. How: The Workings of Digital Twins

How digital twins work is the question met with the most skepticism in practice and in literature. This comes on two levels. The first is: "How can a digital twin improve availability?", typically asked with high skepticism. The question stems from the known complexity associated with maintenance, repairs, improvements, supply chain, availability of skilled labor, and accessibility. The answer to this level of "how" is often tied to the improved access to timely, useful, ship specific information on condition and prioritization. This necessity in turn leads to the next level of "how": "how can a digital twin provide useful information?". The latter is the emphasis of this section, and the functional examples that are presented in section 9.

"How can a digital twin provide useful information?": They are designed to. The design of the structural digital twin considers the objectives on all levels. Thus, the twin including the sensors, data, models, prognostics, and infrastructure is developed in a manner to enable processing, analysis, and communication of information. This considers the use, the resources available, the location of the twin, and the digital infrastructure available.

The above still does not answer "how" to the functional extent. The functional answer to "how does a digital twin provides useful information" includes the details on what data, what model, and what uncertainties are included, what assumptions are being made, and what risk is introduced. The answer for "how do structural digital twins' work" is found across this entire report in that it must cover how models, data, uncertainty, structural performance, and decision support are explicitly linked. Potential functional answers are in some of the examples in section 9, but certainly not all.

3.4. Concerns: On the Use of Structural Digital Twins

Trust and agency are amongst the most critical concerns identified for adopting the digital twin paradigm (Botín-Sanabria et. al., 2022; Danish Maritime Authority, 2018). This topic has been discussed briefly in section 3.2. Additionally, there is a concern on development and use of the topic. While the emphasis of this report is not the digital infrastructure or digital thread, it is worth noting the impact this has on the development of a practical instantiation of a digital twin.

The ship industry, commercial and naval, traditionally relies of specialized software tools for discrete problems. Furthermore, there are many different parties involved in the lifecycle of the ship as it progresses from planning to in-service and onward, as shown in Figure 4. The tools were not initially designed for interoperability but have been modified or revised to be able to communicate across tools. However, the challenge of an open marketplace means that there is a drive for a competitive advantage in ship building and design. This leads to challenges in managing the digital product models that would enable the digital twin (Fonseca and Gaspar, 2021). Moreover, if/when the challenges of the open marketplace are addressed, the communication between the tools may be limited, and the actual data and engineering processes may be obscured to protect proprietary data and tools. This greatly increases the risk in developing a coherent, reliable, and trustworthy digital twin.

It is important to note there are many more concerns associated with digital twins such as security of data, security of operations, accessibility, infrastructure, and integration into planning tools for maintenance, to name a few. To manage scope of this project, these topics are not detailed herein.

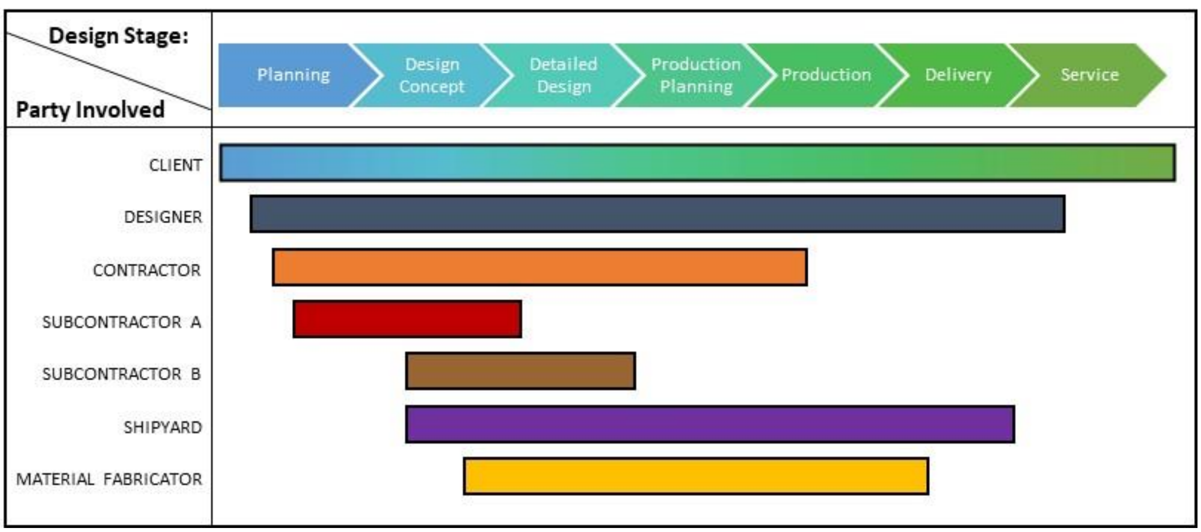


Figure 4. Compatibility Challenges may exist across Different Parties Involved in the Different Design Stage

4. Structural Digital Twin and Models and Data

A concept fundamental to Digital Twin is the merging of data with physics-based models. Researchers have scoped their DTs around a single type of data and a basic model, and have built models on an extensive, diverse database with multi-scale multi-physics models. This begs the question: are the “small” DTs and the “large” DTs both digital twins? The answer comes back to the concepts presented in section 1 – what does the DT need to do, to what accuracy and cost point, and with what available resources.

When “data” is discussed in research for DTs, it has been narrowly defined and broadly defined. The definitions again pose the question: “which is right”, and beg the answer: “it depends on your specific use case”. In general, data for SDTs can fall in categories:

- Direct structural measurements: strains, corrosion measurements, observed crack lengths, out-of-planeness of structural members.
- Indirect structural measurements: accelerometers for the intent of estimating slam pressures, strain gages for estimating global loads or slam pressures.
- Operational data: location as a function of time

The data can be viewed as continuous data sources and discrete data sources. Continuous data would be the time-history data acquired during operations (strain as function of time, accelerations as a function of time, etc.). Discrete data can be viewed as continuous data but sampled infrequently (on the order of years), such as crack lengths, corrosion wastage, and structural changes. Because of their different frequencies, the different data types may have different mechanisms for being incorporated into the SDT. The continuous data is integrated in an autonomous fashion; no human interaction is needed once deployed. The discrete data can be integrated autonomously but may also be a discrete-manmade update to the model.

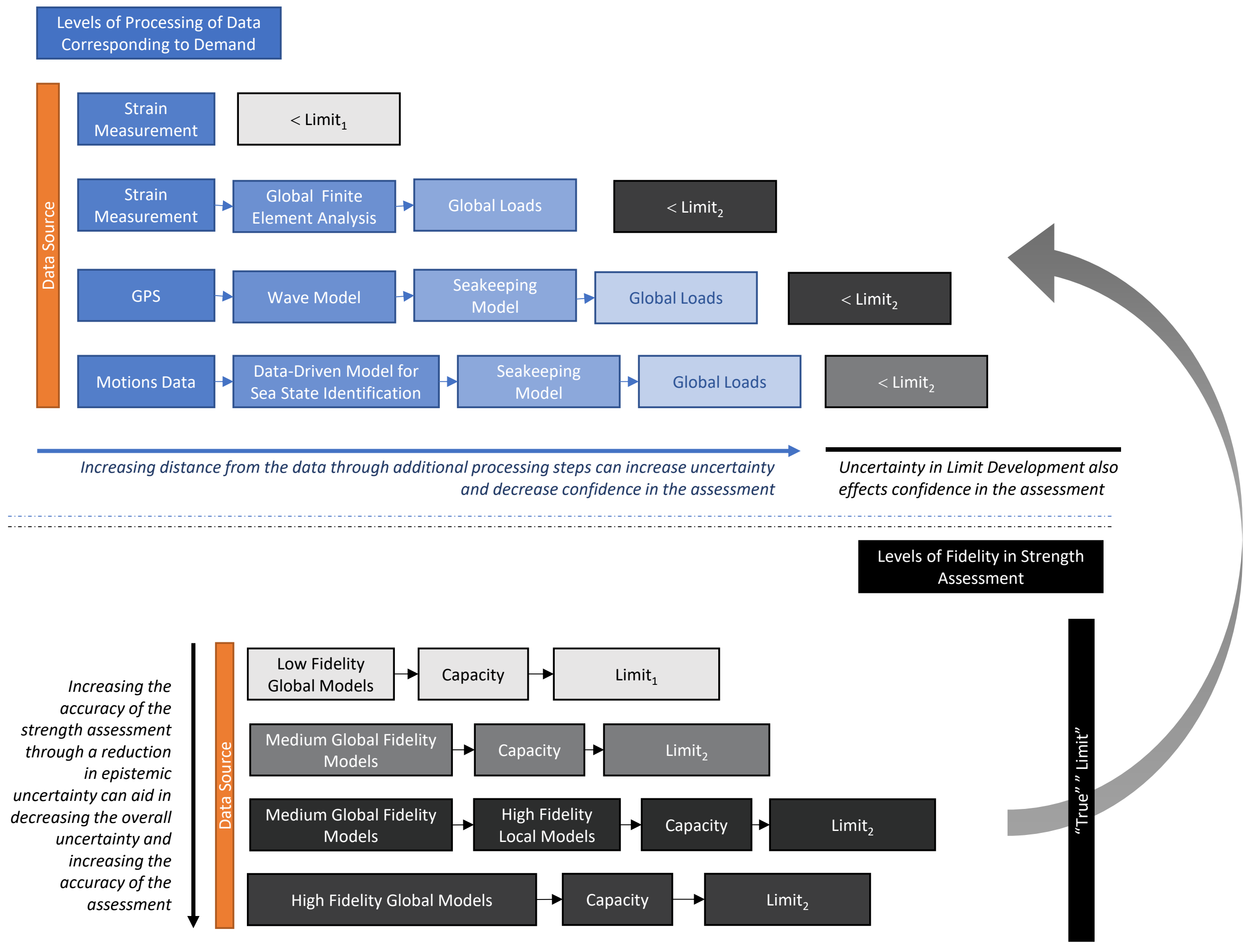
When “physics-based models” are discussed, there is as narrow of definition as a model that captures one type of behavior (structural response, crack growth, seakeeping and loads) to as broad of a definition to include all types of models integrated seamlessly across each other. To help formulate the design of a specific SDT, it may be helpful to look at the different groupings of physics-based models:

- **Structural Analysis Models**: models focused structural evaluation.
    - **Complex**: High-fidelity Non-Linear Inelastic Finite Element Model (FEM), Nested low- and high-fidelity FEM
    - **Intermediate**: Low Fidelity, Linear Elastic Finite Element Models
    - **Simple**: Fundamental theoretical behavior
    - **Surrogates**: Complex or Intermediate models that have been represented in a simpler configuration
- **Hydrodynamics Models**: models focused on hydrodynamics, seakeeping, and loads.
    - **Complex**: Computational Fluid Dynamics Models (CFD)
    - **Intermediate**: Lower Fidelity Hydrodynamic simulations
    - **Simple**: Fundamental theoretical behavior
    - **Surrogates**: Complex or Intermediate models that have been represented in a simpler configuration
- **Damage Models:** models thar represent propagation of damage.
    - **Corrosion Models**: uniform or pitting corrosion
    - **Crack Growth**: linear or nonlinear crack growth
- **Climatological Models:** models thar represent the waves and environmental conditions.
    - **Complex**: full 360 deg definition of the sea spectra
    - **Simple**: basic metrics of point spectrum including significant wave height, wave period, and direction

The ‘model’ in a SDT may in fact be the coupling of multiple models from the different groupings, integrated in a manner to use the available data to support the necessary evaluations.

The confluence of the two aspects “data” and “physics-based models” also leads to the presence of a third concept: data models. This includes Neural Networks, Machine Learning Models, among others, that are developed from physics-based simulations. These data models are then employed as the surrogates for the physics-based models with the intent of incorporating critical physics while working with only a subset of data with lower computational burdens.

When designing an SDT, data, models, and requirements (namely cost and accuracy) are essential to consider when establishing the workflow for data-model integration and performance assessment. Performance assessment refers to the evaluation of criteria relating the demand to the capacity. Figure 5 shows varying levels of integrations of data with models for demand and strength. Accuracy of the data, accuracy of the model(s), and uncertainty propagation throughout the schema all must be considered. For the demand side, increasing distance from the data through additional processing steps can increase uncertainty in the SDT and decrease confidence in the assessment. An example of this could be the use of GPS data that then gets mapped through climatological wave models to estimate the operating conditions; that then gets paired with seakeeping models to estimate the motions and loads. The cost of GPS data, however, is low, making this solution a low cost, higher uncertainty SDT. On the capacity side, the parallel can be drawn: lower cost, low fidelity models may have higher uncertainty.



Considerations for choosing data processing levels and level of fidelity of models include: cost of installation of measurement system, availability of space for measurement system, level and quality of information available on the structure, cost of building the analysis models, cost of running the analysis models, accuracy required for decision that is being supported

Figure 5. Levels of Data Processing and Levels of Model Fidelity

5. Structural Digital Twin and Uncertainty

The structural digital twin process, as described, is a complex process that incorporates a wide set of sources of uncertainties. The uncertainty can be divided into two categories: aleatory and epistemic. Aleatory variability is the natural randomness in a process; often called natural variability. This corresponds most readily to the natural variability in material properties or wave loads or other physical quantities. Epistemic uncertainty is the scientific uncertainty inherent to the model of the process. Epistemic Uncertainty derives its name from the Greek word episteme, which can be roughly translated as knowledge (Ang and Tang, 2017). So, another definition for epistemic uncertainty is the uncertainty derived from the lack of knowledge or information regarding the physical phenomena that dictate the behavior of a system, ultimately affecting the ability to quantify an outcome (i.e., response). For structural digital twin, aleatory and epistemic uncertainties are present.

To start, there are uncertainties associated with measurements. The variability in the measurement device, a strain gage, or accelerometer is an uncertainty introduced by the method used to quantify the physical phenomena, thus falling under a source of epistemic uncertainty. For measurements, there is also the measurement error that corresponds to the measurement method and the variation in the physical phenomena, which falls under both epistemic and aleatory uncertainty. Measurement uncertainty and error are a well discussed fields in test and evaluation that contributes to the overarching uncertainty in the structural digital twin.

There are uncertainties associated with the physics-based models, often referred to as model uncertainty. Model uncertainty is the lack of realism in the model, the inability to define the physical phenomena completely and accurately. This is often associated with the level of approximations for defining the structure and structural behavior. Methods for assessing and integrating model uncertainty for structural applications have been the topic of discussion in the field of structural engineering (Ditlevsen, 1982). There is often a trade space for higher fidelity models, which are intended to have less epistemic uncertainties, to reduced order models based on the availability of accurate data for the parameters of the models and the time and resources available for development and execution. The trend towards increased model uncertainty via surrogate models is often invoked as an essential step for propagating aleatory uncertainties through the model (Sudret, 2022). Note that model uncertainty is separate from the variability in the parameters in the model (parametric uncertainty), the variability in how the model is developed (modeler uncertainty), and the machine precision and errors (computational uncertainty).

Parametric uncertainty corresponds to the uncertainty in the inputs and model-inherent parameters. Examples of parametric uncertainty for structural models include the uncertainties in geometric parameters, Young's modulus, thermal properties, inelastic strength of materials, among others. This form of uncertainty pertains specifically to how the model incorporates these parameters, thus bridging the line between aleatory and epistemic.

Natural variability, or the inherent randomness in the physical environment, for Structural Digital Twin is a major component of uncertainty. The natural variation in structural material properties effects the uncertainty in capacity quantification. The environmental variability is another source; this includes both variability of loads in a seaway (natural variation), but also the evolution of the seaway (process stochasticity). The operations of a vessel within the seaway are a source of inherent randomness. That is, from the process perspective, the operational choices dictated by the mission and operational constrains and freedom impart randomness into the system. Sampling error, on the other hand, is an example of epistemic uncertainty, as it is the error associated with trying to measure the randomness in the stochastic processes and operational randomness.

Due to the breadth of uncertainties that exist within the purview of structural digital twin, and the duration of this project, the discussion herein is focused on aleatory uncertainty associated with natural variation, process stochasticity, and inherent randomness. The scope will also be limited to structural loads. It is recognized that the

SDT can and should be expanded to include methods for uncertainty quantification for the other sources of uncertainty.

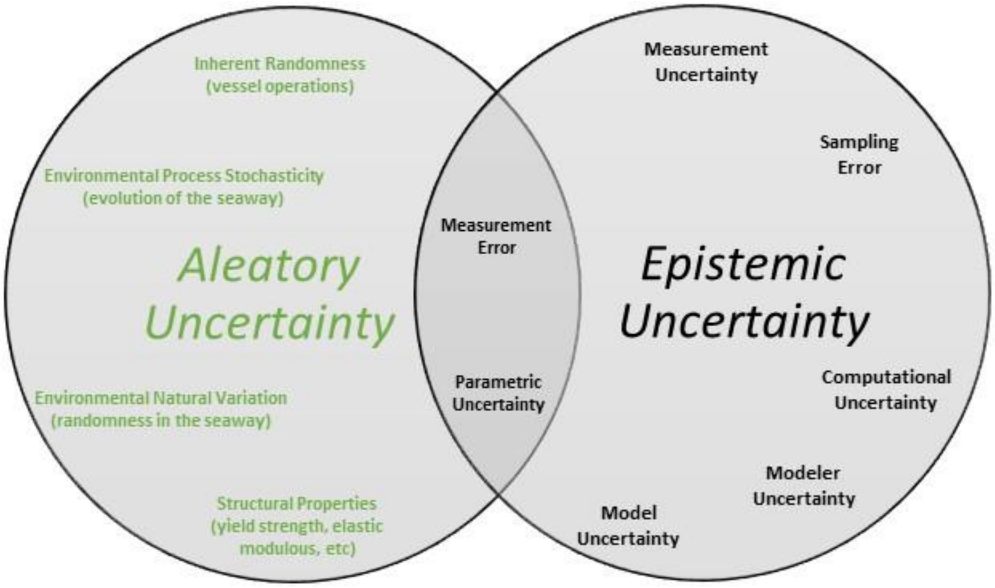


Figure 6. Categories of Uncertainties for Structural Digital Twin

5.1. Quantifying Structural Load Uncertainty

The uncertainties associated with structural loads that will be included in the scope of this project for structural digital twin are the uncertainties from inherent variability of loads within a seaway (i.e. natural variability), changes in loads due to evolving storm systems or the progress of the vessel to a new area (i.e. process stochasticity), and changes in load due to changes in speed or heading by the operator (i.e. inherent randomness). While changes in speed or heading are not in-fact random, but intentional, the resulting affect that those intentional decisions have on the structural loads leads to an otherwise unexplained and inherently different set of loads, which, to the structure, is random.

5.2. Incorporating Uncertainties: Conceptual

The need to incorporate uncertainties into design and assessment procedures is prevalent across the fields of structural engineering and marine engineering. Naturally, these concepts need to find their way into the field of structural digital twin for surface ships. In the design and assessment processes for civil and marine structures, the field has long been aware of the natural variability of structural loads (i.e., demand) and material properties and structural performance (i.e., capacity). Thus, both demand and capacity have been conceptualized as random variables, as shown in Figure 7a. The potential for the capacity to be less than demand then defines the probability of failure.

In load and resistance factored design (LRFD), and to some extent allowable strength design (ASD), the natural variability of the loads and strength are included. In LRFD, the demand and capacity are quantified deterministically and then factored to account to for natural variability and risk tolerances for the specific case. This is idealized in Figure 7b. The ASD approach does not explicitly formulate the design approach as such, but implicitly accounts for the natural variability and risk tolerances with a much broader brush.

As design practices evolve and as digital twins enter into design and operation stages, there is a high potential for a fusion of concepts for how to deal with natural variability. For example, data may be available to statistically quantify the structural loads within the digital twin. However, for performance assessment, the structural capacity is essential. The SDT may be capable of providing the most accurate representation of the structure but may not have the information available to statistically quantify the current strength. As such, it may lead to the fusion of concepts of a probabilistic quantification of demand, and a factored approach for capacity (see Figure 7c). The challenge therein is in setting the targeted probability of failure. Particularly for marine and naval applications, the decomposition of the factor and the capacity is challenging.

The SDT must, therefore, be able to quantify the natural variability in the load and have a clear understanding of failure mechanisms. The remainder of section 5 discusses the uncertainties in loads for SDT. Section 6 contains a discussion on failure mechanisms and strength assessments.

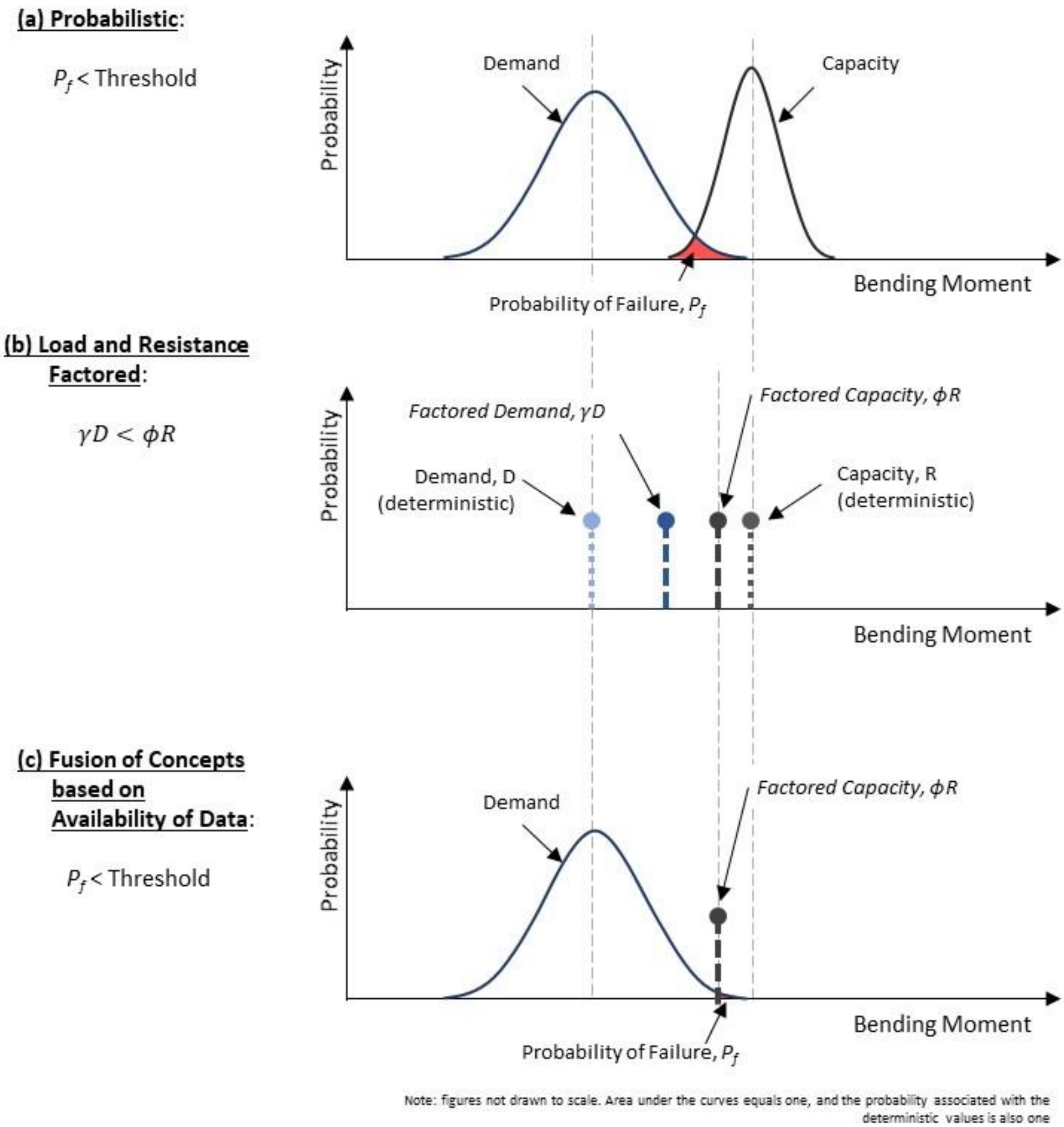


Figure 7. Concepts in Load and Resistance Factored Design and the Extension to Mixed Data Sources

For structural digital twins for surface ships, the objectives of the twin will dictate the extent to which uncertainty is included. For example, condition-based maintenance twins could be formulated in a way where the observed, deterministic, data is used to assess the state of the vessel. In that instance, there is no need to assess probabilistically since the question is “what did the ship actually experience?”. However, there could be the additional question that the twin is supporting: “is it likely that during the next voyage the vessel will incur damage, and should preventative maintenance be done?”. In this instance, an understanding for the natural variability would be coupled with process stochasticity and inherent randomness. On the other hand, operational guidance twins should incorporate natural variability.

For operational guidance SDTs, the data recorded on the vessel may include the response as a function of time, as shown in Figure 8a. The deterministic assessment of the data would indicate that the largest value, shown in yellow in Figure 8a and b, is less than the upper limit (L_u), and the largest negative value, shown in green, is less than the lower limit (L_L). However, when providing guidance, the assessment not only checks if the observed value exceeds the threshold, but also determines if the demand is likely to exceed the threshold if the current condition is continued. This could be supported by a statistical quantification of the response, shown in Figure 8d by the solid blue line. However, it is more common to define the response by the extrema (also referred to as peak and troughs), shown in grey and black in Figure 8c for hog and sag, respectively. The statistics of the extrema can then be quantified separately where hog extrema (peaks) and sag extrema (troughs) can be treated as random variables.

Additionally, SDTs should be developed around the concepts of process stochasticity, in that the probabilistic distribution of the response extrema has the potential to evolve with environmental changes (see next paragraph for additional changes). These types of changes can be integrated into the SDT through updating and/or predictive approaches. For example, the ship may be entering into a storm, or a storm may be developing where the ship is. In this case, the understanding of the past and current statistical fit can be used to update or predict what the future distribution may be. This concept is shown in Figure 9, and shows how not accounting for continuous changes in operations, such as a building storm, could lead to an underestimation of the probability of failure in the next time window.

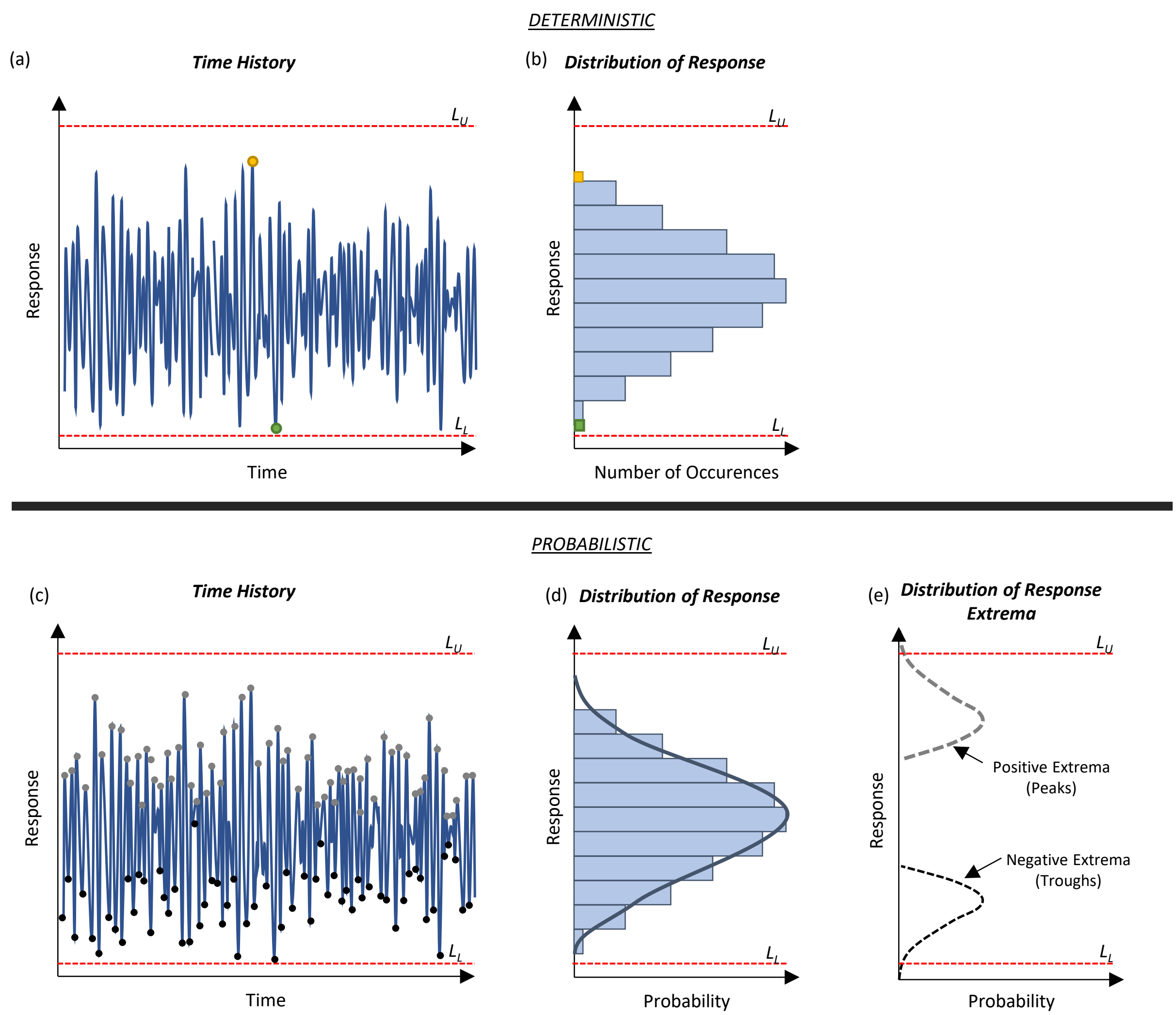


Figure 8. Deterministic and Probabilistic Representation of Observed Data

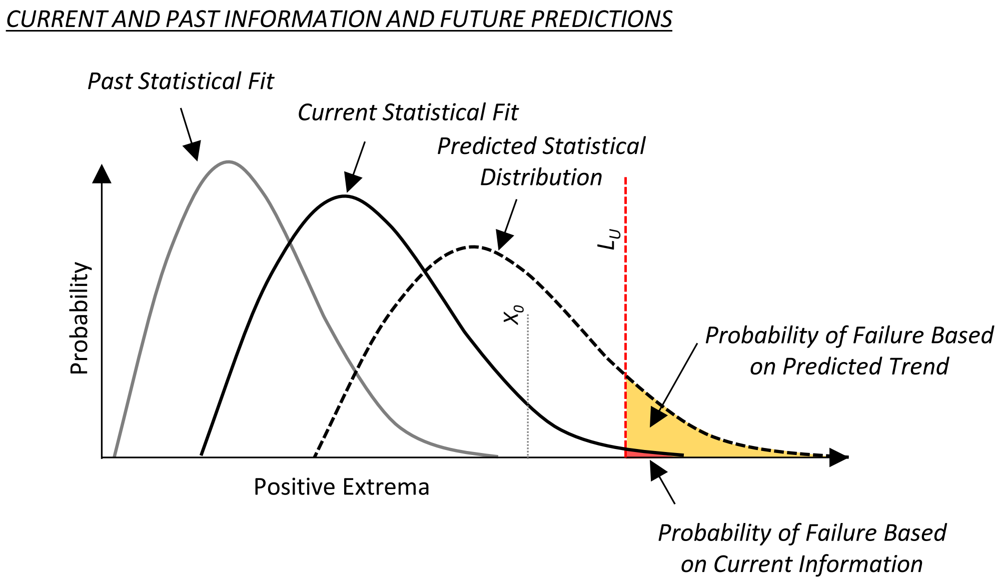


Figure 9. Information for Decision Support Using (top) Current Information Only and (bottom) Past and Current Information and Predictions

Predictive and prognostic methods to support SDT have been the topic of continued research. The complexity of the problem lends itself to solution approaches following Bayesian updating, regression modeling, machine learning and artificial intelligence. Lee et. al. (2022) developed deterministic predictions for wave trains incident on a ship and the resulting ship motions were conducted. Nielsen et. al. (2022) investigated a hybrid maneuvering model for predicting the speed of a ferry under model uncertainty and varying operating conditions. The work demonstrated the applicability of neural networks to capture complex, nonlinear behavior that is not an inherent component of first principle hydrodynamic models (Nielsen et. al., 2022).

One of the major challenges associated with statistical support for guidance, predictive and prognostic approaches is the inherent randomness in operations. That is, the effect of external factors on the progression of the vessel's response. This includes an immediate change in heading or speed. The statistical distribution of the response based on the past number of minutes is then no longer valid for quantifying the current state. Ideally, the exact operational conditions could be known and the SDT would divert to analytical or historical data while the observed dataset was being developed enough to support a statistical quantification. However, the limited technology associated with that real-time state awareness of wave environment to the required level of detail inhibits this path. As such, the SDT must be developed in another manner to account for this inherent randomness.

Class societies have developed and published rules, notes, and guidances regarding hull monitoring systems and smart systems. Inherently, these types of systems are narrowly scoped instantiations of structural digital twins. The varying classes address different characteristics of a system:

1) Approach: the authors have characterized the approaches in the class standards as pertaining to different means for accounting for natural variability, process stochasticity, and inherent randomness. The methods in the class standards include:
    (i) Diagnostic (or deterministic): reliant on data observed.
    (ii) Probabilistic: reliant on a statistical characterization of observed data wherein stationarity is assumed
    (iii) Prognostic (or predictive): reliant on the aggregation of current and past information with methods for forecasting
2) Calculation Period: the period that the system must use data from to support assessments.
3) Update to Display: defines the rate required to provide a visual update to the operator.
4) Limits: what structural considerations are being accounted for when interpreting the response data to provide context and make the information useful to the decision maker.

A summary of the review of rules and guidances is provided in Table 3.

In the review and comparison, there is an inconsistency in requirements for approaches. Most standards require a probabilistic assessment of the current data. Some require predictive approaches. However, missing in the class standards reviewed is a requirement to account for inherent randomness (i.e., the ability to incorporate abrupt speed, heading, or environment changes). This omission may lead to systems being designed to provide probabilistic assessments that are not valid and thus lead to false positives or false negatives while providing guidance.

In current research, Nielson (2022) suggests that 20-30 minute windows may be applicable for use in certain cases. However, the use cases should be considered when establishing what windows are used for probabilistic quantification. The 20-30 minute window provides a useful upper bound.

The calculation periods for assessing natural variability and process stochasticity are inconsistent across the standards. Anywhere from “a rolling basis” to 4 hours was provided as the time window to collect data for statistical and predictive assessments. The standards provided no referenced sources to indicate that the time window is based on rigorous review of the uncertainties expected in the application. BV was one of the few to provide a requirement based on a performance metric of the method: “the recording duration per cycle is to be adapted to produce results that are not to deviate by more than 10% from one wave encounter to the next in steady navigation conditions” (BV, 2022). In some instances, the standards indicated that the interval could be configurable, although it is unclear if “configurable” indicates that it is one-time configurable and set at the deployment of the system, or if it indicates that the system is self-aware and adapting the inputs for the configuration.

The complexity of the assessment approaches is alluded to in DNV’s guidance “for predictive assessments, past 4 hours for displacement ships and 30 minutes for high-speed vessels” (DNV GL, 2017). This could stem from the propensity for process stochasticity to vary based on ship type: displacement vessels tend to sit in the water consistently over longer periods of time. And larger vessels tend to follow a low-varying path. So, all other conditions being the same, the response is relatively constant (stationary). However, planing vessels, or short-mission vessels with variable paths, may only stay in constant (stationary) conditions for a short period of time.

The refresh rate of the display also varies extensively from real-time to 30 minutes. The variation in the requirement speaks to the variation in which different users may need information to affect change. However, it can also have an influential effect on the design of the system. The real-time or near real time relay of information dictates that the developer design the SDT or support analysis to be executed with the available computing resources. The architecture of the SDT may become more complex to account for the quick response needs and the need for higher fidelity assessments and statistical analysis.

As a side note, DNVGL - Rules for classification: Ships DNVGL-RU-SHIP Pt.6 Ch.9 (DNV, 2017), has a note regarding filtering of data: “Software shall include high-pass, low-pass and band-pass time domain digital filters. The cut-off frequency of the filters shall be configurable through the software and shall be stated in the configuration file; the hull monitoring system shall have the capability to optionally remove the strain due to temperature differences in the hull girder” (DNV, 2017). This is the only class society denotation found that alluded to the concept that strains associated with different sources need to be identified and accounted for appropriately.

The concept that strains come from different sources is a critical feature that is not clearly adjudicated or discussed in the class guidances. The variability of measurements to thermal changes, electromagnetic interference (EMI), or other sources on vessels must be accounted for when established the SDT performance assessment approach. The measured data must be processed to ensure that the measurements pertain to the limit for which they are being evaluated.

Table 3. Requirements for "decision support" based on class guidance.

Class Standard	Approach	Calculation Period	Update to Display
ABS Hull Condition Monitoring (ABS, 2020)	Not Identified	20-30 min; Rolling basis is accepted	Real time or near-real time
ABS SMART (ABS, 2022)	Diagnostic and Prognostic	NA	NA
Rules for the Classification of Steel Ships, NR467, Part F (BV, 2022)	Probabilistic	Not less than 10 minutes (the recording duration per cycle is to be adapted to produce results that are not to deviate by more than 10% from one wave encounter to the next in steady navigation conditions.)	Not provided
BV NR675 Additional Service Feature SMART (BV, 2021)	Prognostic	Invokes NR467 - JULY 2022, Part F	Invokes NR467 - JULY 2022, Part F
DNVGL - Rules for classification: Ships — DNVGL-RU-SHIP Pt.6 Ch.9 (DNV GL, 2017)	Probabilistic and Predictive	Time period for statistics shall be configurable. For predictive assessments, past 4 hours for displacement ships and 30 minutes for high-speed vessels.	5 minutes
DNV GL Smart Vessel (DNV GL, 2020)	Probabilistic and Predictive	Invokes DNVGL-RU-SHIP Pt.6 Ch.9 Sec.3,	Invokes DNVGL-RU-SHIP Pt.6 Ch.9 Sec.3,
Lloyds Register, ShipRight, Ship Event Analysis (Llyods Register, 2021)	Probabilistic (implicit)	Not provided	Not provided
CCS, Rules for Intelligent Ships (CCS, 2020)	Predictive	Time interval shall be stated in configuration file	No longer than 5 minutes
CCS, Hull monitoring and assistant decision--making system for operations in ice (CCS, 2018)	Predictive	*Forecast for the next 1-2 hours	Real-time
Class NK (Class NK, 2020)	Probabilistic (implicit)	4 hours	At least every 10 to 30 min
Notes: * only forecast window was identified			

6. Structural Digital Twin and Structural Performance

Structural performance assessment is the convergence of the concepts of structural loads, capacity, and failure mechanisms. This section focuses on structural performance assessment in the context of decision support for operational guidance, condition-based maintenance, and design. SDT is the prime candidate for the naval and marine community to incorporate performance-based decision support due to the reduced uncertainties that are associated with usage. Usage is no longer a set of possible operational conditions, but one immediate state of operations. And data is available to enhance the uncertainty quantification for demand. Thus, all that is needed for performance-based decision support is a clear quantification of failure.

Failure is commonly defined as “a state of inability to perform a normal function” (Merriam-Webster.com, 2022). Structural failure, therefore, can be defined as “a change in state such that the structure no longer provides a required function (load carrying or otherwise) or impacts some specified system performance to an unacceptable degree” (Hess, 2003). Similar definitions for failure have been used in the civil engineering industry for decades, spurring advancements in code organization and structural analysis methods.

Through recognizing the function of a structure as both resisting predictable internal and external loading and supporting the functional use of the structure, it can be concluded that conforming to prescriptive code requirements, which usually focus on structural capacity insofar as resisting design loading, does not guarantee structural system integrity. In the civil industry, buildings with special service functions (e.g., hospitals, laboratories, etc.) and extreme loading conditions (e.g., earthquakes, fire, tsunamis, etc.) have guided the design landscape into adopting a performance-based design approach to resist the unique structural demands that are above prescriptive design considerations. The critical functions of structures, like hospitals in the civil context and ships in the naval context, require that design practices meet the demand with intensive engineering design. The availability of demand information (i.e. seismic records, burn rates, wave profiles) and the advanced computational tools enables an understanding of performance, where layers of uncertainties have been removed because of the data and models, which makes this level of rigor not only possible but required for structural performance. Thus, these special loading and function cases can be designed accordingly using the modern computational tools and procedures normally outside the limited scope of prescriptive codes (Dusenberry, 2019). This parallels the use case for digital twins for surface ships well.

Surface ships (naval and marine) serve very particular critical functions (medical ships, supply ships, defense, rescue), are consistently under highly dynamic and extreme loads (which, in-part, can be quantified with gathered data), and can be assessed on a performance level. In order to do this, an in-depth analysis of what constitutes a systematic failure is needed. The following presents an initial discussion on failure through comparing the philosophy behind the civil and naval structures codes while proposing a new design framework to advance the methodologies by which failure of structures is determined for naval vessels.

The structural components and structural systems of naval and marine ships support a wide variety of functions which range from resistance of seaway loading through primary and secondary structures to support for on-board machinery or equipment. A failure can fit into one of two categories: ultimate failure or serviceability failure (Hess, 2003).

The first, ultimate failure, covers component failure (e.g., material yielding, buckling, fracture, etc.) which is often adequately defined in prescriptive codes and system failures (e.g., progressive collapse, non-linear buckling, disproportionate collapse, etc.) which require more complex analysis methods to study. The second, serviceability failure, is the limit state which defines all states at which the structure did not perform as intended, increased the risk of ultimate failure, or disrupted the normal function of the structure (Barrow, 2021). Defining these failure categories and how to create proper performance metrics for them is important for the design, resilience, and continuing operation of a naval vessel.

6.1. Ultimate Failure

Ultimate Failure is defined by the total or partial collapse of a structural component or system due to reaching an ultimate limit state (Hess, 2003). Both the civil and naval fields adopted probabilistic design philosophies in the form of Allowable Strength Design (ASD) (sometimes referred to as Working Stress Design) and Load and Resistance Factor Design (LRFD). These design methods determine failure to occur when the load/stress (demand), usually a probabilistic value, which is applied to the component is greater than the strength (capacity), usually a deterministic value, of the component. Both methodologies account for the variability and uncertainty in the strength of the structural members through a safety factor applied to the average expected capacity of the member. LRFD includes another factor to increase the average expected demand on the structural member to account for the variability in loading. As mentioned, these design philosophies are probabilistic in nature and require significant data sets to understand the variability in the strength, geometry, and loading for structural members. For the civil field, this variability in loading and materials is well documented and the reasoning behind the current factors used in the AISC/ANSI Specification (American Institute of Steel Construction, 2016) is detailed in an article by Theodore Galambos (Galambos and Ravindra, 1981).

However, the naval structures field does not have as well documented or studied distributions for loading and has some peculiarities in the application of LRFD design philosophies to code equations. In a study comparing multiple ship structures standards, the Ship Structures Committee found that those codes which had adopted an LRFD formulation were modifying the structural response factors so the result would match the pre-LRFD specifications (Kendrick et. al., 2006). A sample comparison of the compressive buckling equations from AISC/ANSI 360-16 and International Association of Classification Societies, IACS (International Association of Classification Societies, 2022), is shown in Table 4. A significant difference in the philosophy of the two approaches is that the AISC/ANSI specification accounts for inelastic buckling, which is why the two equations from AISC shown have lower coefficients, while the IACS formulation does not.

Because of the complicated nature of wave loading, whipping, slamming, and other highly dynamic loading patterns which ships experience, obtaining a generalized set of loading data to perform any probabilistic analysis on is a difficult task. Complex loading scenarios such as these also cannot be fully analyzed using first-principal mechanics analyses. This ambiguity on desired risk levels and probability of exceeding them places the engineer in a position to determine whether the design approach which is prescribed is conservative or not because it is not necessarily built-in to the method. It is therefore a logical progression that the naval engineering process must account for the specific scenarios which the ship will be found in and be designed to a clearly defined reliability requirement.

Table 4. Comparison of Compression/Compression Buckling Failure Definitions

	AISC/ANSI 360-16 (American Institute of Steel Construction, 2016)	**IACS** (International Association of Classification Societies, 2022)
Prescribed Reduction Factor on Capacity	$\phi_c = 0.9$	0.65, 0.75 (Load combination dependent)
Compressive Stress Capacity of Short, Non-Slender Elements	$F_{cr} = \left(0.658^{\frac{F_y}{\frac{\pi^2 E}{(KL/r)^2}}} \right) F_y$	Not Specified
Compressive Stress Capacity of Slender Elements	$F_{cr} = 8.66E \left(\frac{r}{KL} \right)^2 F_y$	$F_{cr} = \pi^2 E f_{end} \frac{I}{A \ell_{pill}{}^2}$

Prescriptive code methods for naval and marine structures that define stress limits with implicit factors of safety for all components have two drawbacks:

1. The structural behavior due to connected elements is not accounted for in the stress limit evaluation (Hess, 2003) so structural compatibility and secondary strength may not meet functional requirements.
2. Complicated mechanical mechanisms such as non-linear buckling, fatigue, and fracture, the most common failures in naval structures (Hess, 2003; Raju and Premanandh, 2018), are simplified down to a stress limit.

Because of the ambiguity in the formulation of capacity and demand safety factors, as well as the possible problems that arise from looking at only local component failures, a system-wide definition of failure and a more robust method of defining and calculating failure becomes necessary. This is one example of the "limit of validity" (Hughes and Paik, 2010) which Hughes brings up in defense of performance based (or rationally based) structural design.

Defining a system failure for a naval vessel is where the common ASD and LRFD approaches are no longer able to be used effectively (Czujko, 2018). A system failure is a critical component failure or series of component failures which results in a catastrophic loss of strength or stability for the entire structure. Identification of the critical components of such a failure and the mode of the collapse is the job of the designer or engineer. Some methods for prediction and designing against progressive collapse, one of the most prevalent forms of system failure where a component failure causes a series of other component failures, have been created, but significant knowledge gaps have been acknowledged (Czujko, 2018). Other system failures, such as accidental failures caused by collision, explosions, etc., are specialized to the naval design industry because of the possibility of water infiltration and major stability issues. Therefore, if no explicit analysis methods for these are set, the designer or engineer must set performance minimums to ensure that the risk of these failures are low.

6.2. Serviceability and Non-Structural Failures

Any failure that does not immediately result in a structural collapse will fall under the category of a serviceability failure or non-structural failure. These two types of failures occur due to structural behavior, but no ultimate limit state has been reached or no structural instability has been made. The civil and naval/marine structures industries use two different definitions for a serviceability failure:

1. AISC/ANSI 2016 (American Institute of Steel Construction, 2016): A serviceability limit state is a "limiting condition affecting the ability of a structure to preserve its

appearance, maintainability, durability, comfort of its occupants, or function of machinery, under typical usage."

2. NSWCCD-65-TR-2002/14 (Hess, 2003): "We may consider a serviceability failure to be an event which increases the risk of ultimate failure to unacceptable levels or degrades non-structural systems in an unacceptable manner."

Both definitions commonly associate deformation and damage to the structural system, especially that which places the structure in immediate danger of an ultimate limit state as a serviceability failure. Both imply that repairability and continuing operation are key goals of designing against serviceability failures. The civil structure definition includes more subjective goals such as "comfort of [the structure's] occupants". This, as well as the "function of machinery", are not explicitly delegated tasks for naval structural engineers when designing large structural systems. The other category of failure, non-structural failure, is one in which the structural behavior (i.e. vibration, thermal or electrical conductivity, appearance, etc.) directly impacts the functioning of a non-structural system. These types of failures usually do not have any inherent failure limit and require acceptable performance limits to be defined prior to design.

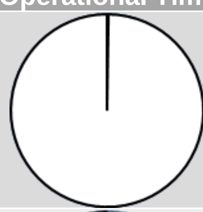
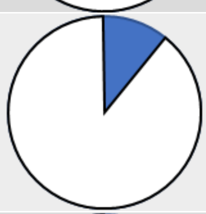
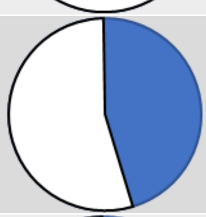
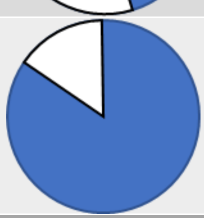
AISC directly states specific parameters to be controlled for serviceability of civil structures: deflections, drift, vibration, wind-induced motion, thermal expansion and contraction, and connection slip. Hess notes that naval serviceability design equations are usually based on deflections and material yield (Hess, 2003). The second is normally controlled in component and system failure design but deflection failures are a commonality between the two approaches. Naval structures could suffer susceptibility to buckling, unseating of cargo or equipment, or degrading of total seakeeping ability due to excessive deformation. Each possible failure state due to deformation of the structural system needs to be defined and limits set based on individual concerns.

Operational failures are also an important consideration for serviceability and non-structural failures. These occur when structural behavior directly inhibits or impedes equipment or personnel from performing normal in-service tasks. A common failure of this type is excessive vibration. This may occur when the structural system resonates at the frequency of the machinery attached to it causing significant vibrations, sounds, or both. These vibrations may also cause significant damage to structural components through fatigue and crack propagation (Vukelić & Vizentin, 2017). Another operational failure may be excessive acceleration (Hess, 2003). A structure will be designed against high accelerations, but if the structure allows for these to be common, then discomfort of human personnel and damage of machinery is possible. These are two examples of events which may be considered failures and, with proper design limits, could be designed against or detected in in-service ships.

Life-safety is one consideration which, though not explicit in naval design, is an important metric for serviceability. This refers to the probability of injury or death due to structural effects under different loading conditions. This is a well-documented metric often used in earthquake analysis of building structures (Applied Technology Council, 2018). This often takes the shape of defining safety levels or predicting the casualty risk for different loading scenarios. These levels are usually defined as, from least risk to most risk: Operational, Immediate Occupancy, Life Safety, and Collapse Prevention. Operational means that there is minimal to no damage under that category's loading and the structure can continue to function through the event. Immediate Occupancy means that minimal damage has occurred, none of the damage could be categorized as a serviceability failure, and function of the structure can continue after minor repair. The Life Safety level is the loading level when damage resulting in serviceability or component failure occurs to the structure, but minimal injury or deaths can be expected from the structural damage. The last safety level is when the loading causes structural damage which does not result in a system collapse but has no guarantee of continuing functionality or safety of occupants. Similar metrics can be adopted for ship structural systems for critical loading scenarios. An example of what that might look like is shown in Table 5. The "Seakeeping Viability" level may be the level at which the hull girder integrity is ensured but secondary structures are not, and the ship cannot continue its mission. "Life-Safety" could be the level at which there is significant structural damage to secondary structures, but none would endanger personnel or critical life support systems. "Mission Continuation" could be the category marking minor structural damage which will need repair post-mission. “Operational” would be the damage level associated with little to no damage and no repair needed. Characterizing different structural states under this type of system would offer significant guidance and knowledge regarding the in-service structural integrity.

A non-structural failure which may be significant to the design and continued operability of the ship is the cost of repairing the ship after a significant structural failure event. First, in design, repairability can be considered in the design of structural members and connections, especially in considering the areas of high risk of component failure. In service, the cost trade-off between repairing a damaged member vs. repairing a failed member or structural system needs to be considered for monitoring systems and alarms. There may also be a trade-off of the repair/non-operational time of preventative repair versus corrective repair (Applied Technology Council, 2018).

Table 5. Potential Vessel Damage Categories

Damage Category	Repair and Non-Operational Time	Initial Design Costs	Repair Costs	Life-Safety Rating
Operational		$$$$	$	High
Mission Continuation		$$$	$$	High
Life-Safety		$$	$$$	High
Seakeeping Viability		$	$$$$	Mid
Note: $ are used to provide rough approximation on costs for comparisons				

6.3. Performance-Based Design and Risk Assessment of Failures

A commonality between all the failure categories detailed above is that defining the limit states to determine when failure occurs requires both high-fidelity models and engineering judgement for the individual structures, especially for complicated systems such as naval ships. The knowledge gaps in the demand and capacity probabilistic properties are a significant hindrance to the creation of safe and thorough prescriptive codes, therefore a performance-based design philosophy should be adopted. Performance-based design would allow the engineers and designers to detail what failure looks like from a combination of past projects, codes, and engineering judgement (Dusenberry, 2019). With the only requirement being setting performance metrics, this enables the use of high-fidelity physical, mechanical, and numerical models to check the performance, instead of relying on prescriptive equations to check compliance. This style of design has been shown to result in better performing, efficient, reliable, and cost-effective marine and naval structures (Hughes & Paik, 2010). The application to SDT for operational guidance and condition-based management is logical.

As mentioned previously, probabilistic analysis is needed to determine the acceptable probability of failure in structural specifications. This requires a large data set of field loading conditions and structural responses. Not only would a data set like this assist in the creation of better safety factors to pair with the definition of failure, but it would also begin to build a baseline for development of risk levels associated with the component and system failures. These risk levels could become the performance metrics by which serviceability failures are defined. High-fidelity models could be studied and equilibrated to in-service ships. Combined, this information would be able to be used in a SDT which can warn of critical situations and offer guidance on how to decrease or avoid the risk.

6.4. Limit States for SDT

In a review of class society's rules for hull monitoring systems and smart vessels, the concepts of failure mechanisms appear in the definition of limits. Table 6 summarizes the limits identified. As the industry pursues concepts like structural digital twin and its practical implementation, there is a clear need for an enhanced understanding of structural failure and structural criteria. For performance assessment, ultimately it is coherency across the load definition, the failure definition, and the risk tolerance levels.

A brief aside on demand: the structural loads or demand (i.e., structural response) as developed in the SDT or hull monitoring system are derived from measurements coupled with physics-based models, statistical assessments, and, sometimes, data driven models. Thus, the demand, as suggested in the standards, has the potential to range from a deterministic value to an unspecified probabilistic value, or an unspecified predicted probabilistic value. The standards then provide the limits (shown in Table 6) without explicitly tying the load to the strength assessment or the risk tolerance for the use case. This lack of class-documented requirement is an area that requires immediate improvement. It is essential that the SDT be developed in a manner that coherently and appropriately accounts for loads, strength, and risk.

Coming back to strength: There is a broad variability in the limits that are required to be incorporated by the class societies. The catch-all approach is broadly taken; in that the explicit engineering decision tree for loads, strength, and risk is left to the discretion of the developer and then to be submitted for review and approval. This indicates a clear need to formalize the approach so as to ensure the safety of the vessel and the safety of vessels and personnel nearby.

The limits as defined by "[t]he requirements on the basis of which the hull structure is approved" (BV, 2022) enable a clear enumeration of strength and which failure mechanisms are accounted for. However, this does not afford consistency across the performance assessment method since the loads are being established by the data, as opposed to empirically calculated lifetime load values. Furthermore, the risk tolerances or safety margins are not being revised with respect to the change in the decision that is being support: immediate use as opposed to long-term design.

The further elaboration of limits is essential to the implementation of structural digital twins. The specific failure mechanism and component or system level assessment methodologies should be laid out. The coherent assessment approach for accounting for loads, strength, and risk needs to be established. The fundamental shift from “design” to “performance” is required for the rules societies to enable clear and coherent guidance for SDT for navigation performance support and performance (condition) based maintenance.

Table 6. Limits for "decision support" based on class guidance.

Class Standard	Limits
ABS Hull Condition Monitoring (ABS, 2020)	The warning levels are to be set with reference to the approved scantlings and their conditions of approval. Warning level settings are to be submitted for review.
ABS SMART (ABS, 2022)	Deformation, yielding and buckling, fatigue
Rules for the Classification of Steel Ships, NR467, Part F (BV, 2022)	The requirements on the basis of which the hull structure is approved
BV NR675 Additional Service Feature SMART (BV, 2021)	Invokes NR467 - JULY 2022, Part F,
DNVGL - Rules for classification: Ships — DNVGL-RU-SHIP Pt.6 Ch.9 (DNV GL, 2017)	User configurable
DNV GL Smart Vessel (DNV GL, 2020)	Invokes DNVGL-RU-SHIP Pt.6 Ch.9 Sec.3,
Lloyds Register, ShipRight, Ship Event Analysis (Llyods Register, 2021)	Not provided
CCS, Rules for Intelligent Ships (CCS, 2020)	Not explicitly defined; generically related to: Longitudinal strength of the hull structure. Stress in the critical structural areas; Temperature of the structural members affected by high or low temperature. Bow slamming pressure (for applicable ship types). Liquid sloshing in tanks (for applicable ship types). Structural stress of ice belt region of ice-reinforced ships.
CCS, Hull Monitoring and Assistant Decision-Making System for Operations in Ice (CCS, 2018)	Permissible stresses for shell plate, framing, bays, longitudinal strength through bending and shear.
Class NK (Class NK, 2020)	Index values for setting up alarms to judge danger to ships from stress and acceleration are to be decided by ship-owners in consultation with the Society

6.5. Recommendations for Strength Formulation and Integration with SDTs

This preliminary review of the ship failure definitions pertaining to structural design reveal that there is still much research and development that needs to be performed to properly create a performance-based definition of failure which accounts for the complexities in naval design and provides more freedom for the designers or engineers to create efficient, safe, and effective structures. Some key actions that should be taken are:

1. A literature review of all naval design specifications and handbooks to check the underlying physical assumptions and catalogue all safety factors. These can be compared to other maritime, shipping, civil, etc. structural specifications to analyze different design philosophies and judge which offer better safety for naval design. This can be followed by lab and model testing to analyze the probabilistic distribution of material properties, the buckling behavior of various components (with the goal of creating buckling curves that can be referenced for design), the fatigue, corrosion, and fracture behavior compared to requirements by specified standard details, etc.
2. A survey of common serviceability failures, especially those which have interrupted service or delayed mission activities. This can be followed by an analysis of the structural effects which caused these failures and development of structural guidelines to avoid them in the future.
3. Creation of a guidance document for system-wide structural failures. This will include the different types of system failures, how to identify the critical structural elements and their failure states, and criteria for additional structural integrity for the systems. This can be created through a review of past structural system failures of naval ships, a numerical analysis of multiple naval ships to determine critical structural elements and failure modes, and literature review on system-wide failures in other fields (e.g., progressive collapse in civil structures).
4. Development of design damage categories could offer guidance in the SDT recommendations for the operations and maintenance of the ship. This could aid

in prioritizing costs and time towards enabling performance, as opposed to enabling compliance with design-stage products that had high uncertainty in usage assumptions and may be over conservative. All previous research goals listed above feed into this goal as it covers all failure modes and compiles them into usable categories.

5. Determination of load combination effects on performance. Complex failure modes may exist that are otherwise unaccounted for in standard design processes where they were not explicitly included but implicitly covered because of safety factors and margins put in the criteria. It may also allow for better early detection of failure as the structural impact of combined environmental and internal loading is better known.

The development of a performance-based approach for structural failure of naval and marine ships is essential to the development of meaningful Structural Digital Twins because it offers a higher-fidelity and more rational approach to defining failure modes and classifying damage levels. This in-turn offers invaluable information for design, in-service guidance, preventive maintenance, and repair. More knowledge on component and system failures with better techniques for the prediction, detection, and classification of them offer more information which can be used to develop robust structural models and train predictive algorithms to better capture structural behavior.

7. Structural Digital Twin and Decision Support

SDTs that support operational guidance and condition-based maintenance are great candidates for a performance-based structural evaluation approach. When formulated for the specific problem, performance-based structural evaluation approach allows for the demand (deterministic, probabilistic, and/or predictive), capacity (component and/or system as quantified through low- or high- fidelity models), and risk tolerance to be included. The importance of different situations, such as peacetime or wartime, allows the STD to appropriately address situations. Likewise, a SDT can be designed to support operations in normal weather or rough weather, where risk tolerance levels may also differ. It allows for different considerations to be made if the ship is manned or unmanned, thus incorporating the fact that there are different consequences of failure.

For SDTs, performance-based decision support first involves the identification of the state that the vessel is in:

- Routine / Under Duress (Peacetime / Wartime)
- Manned / Unmanned
- “Normal Weather” / “Heavy Weather” (Day-to-Day / Hurricanes)

State identification helps establish the imperative for operations, consequences, and risk tolerances. Each branch may have different performance-based criteria. For instance, the safety margin for the “peacetime-manned-normal weather” branch may be different from the safety margin for the “wartime-unmanned-normal weather” branch.

Next, performance-based decision support requires a definition of capacity. This is typically done by first defining what failure is. Structural failure models can include (1) system-levels models structured around component failures and (2) explicit quantification of system performance. The first includes defining the individual components and their failure mechanisms, and then describing the system’s performance as a series, parallel, or series-parallel model with all the components included (Ang and Tang, 2007). The second is, as previously described, a complex nonlinear analysis of the full structural system to evaluate for failure, load shedding, and progressive collapse. While the latter more appropriately accounts for the physics in the failures, the variability inherent in the problem drives this approach towards a probabilistic characterization of the structural capacity. The probabilistic, nonlinear analysis of a large system becomes computationally expensive (if it has enough data on the stochastic parameters to be run at all).

The following logic tree framework is proposed for use when developing SDT solutions for operational guidance. First and foremost, the SDT must be developed to be able to place itself in the correct branch:

1. Routine – Manned - Normal Weather
2. Routine – Manned - Heavy Weather
3. Routine – Unmanned - Normal Weather
4. Routine – Unmanned - Heavy Weather
5. Under Duress – Manned - Normal Weather
6. Under Duress – Manned - Heavy Weather
7. Under Duress – Unmanned - Normal Weather
8. Under Duress – Unmanned - Heavy Weather

In this construct, "optionally manned" vessels can be supported. Where-in the vessel may be manned for certain operations and unmanned for others.

Until the point in time where a robust probabilistic solution that can account for steady state, gradual changes in state, and abrupt changes exists, the SDT must also can identify what state it is in:

- Steady state
- Gradual changes in state
- Abrupt changes

Then, a performance based structural assessment path can be developed for each path. This is conceptualized for the "Routine - Manned - Heavy Weather" branch (i.e.

branch 2) in Figure 10. In this figure, there are separate paths that are posed for each of the states.

- For the steady state, deterministic data provides valuable information on what the ship has recently been subjected to, and probabilistic data is useful for accounting for the natural variability in the response. Both could be used. In each case, the components included in the definition of the system would be evaluated, and then assimilated into the system level assessment. The deterministic and probabilistic evaluation would then have to be fused to provide a clear output from the SDT.
- For the gradual change branch, the same process could be applied, although now including the predictions.
- For the abrupt change branch, the deterministic value may be the only useful data to include in the evaluation. However, since it is the only data that is being used, the margin being mapped to the structural evaluations may be different from that used in the deterministic branch of the steady state (or gradual change) branch.

The logic tree shown in Figure 10 should be expanded for a complete SDT, with additional 7 branches.

The logic tree implicitly prescribes a system level structural assessment that is developed from a series, parallel, or series-parallel model containing the different components. Discussion on how that would be developed and validated is outside of the scope of this work.

This logic tree also implies that there is a systematic method to integrate deterministic assessments (i.e., based only on retrospective data), probabilistic assessments (that account for the current variation in the response) and predictive assessments (that track and account for trends over time). This is a challenging problem focused on risk and risk tolerances. Further discussion on this is needed for the development of a SDT, but falls outside of the scope of this paper,

It is worth noting that the performance-based approach allows for the SDT to account for monitoring data and the different branches of the situation. This is fundamentally different from design. In the design stage, rules and standards have to holistically account for all the branches at once, and all of the potential loading scenarios. The rules and standards often lack the clear enumeration of how load variability is included, how material variability is accounted for, how strength variability is included, or how consequences and risk tolerances play a part in the criteria. This makes it extremely challenging to pull out the core requirements for structural assessment when it comes to digital twins and related products (such as monitoring systems).

To parallel the proposed logic tree for SDTs for operational guidance, this paper proposes the same framework can be applied to condition based maintenance (CBM) SDTs with the appropriate adjustments for risk associated with the objective. The logic tree for CBM SDTs can be defined around the deterministic data. That is, only what happened to the ship is considered when identifying if any inspection actions need to be taken. This is shown in Figure 11.

CBM SDTs can also be expanded to include probabilistic and predictive quantities. However, the time window for predictions is larger (i.e., 1 transit, 1 year, 1 deployment, etc.) than the shorter time windows (i.e. 1 min to 4 hrs.) included for the operational guidance support. Thus, the methodologies for predictions will be different. They would need to account for the potential conditions that the ship may experience over the transit (or year or deployment), as well as the current damage state and damage growth models. This is shown in the logic tree for prognostic based CBM in Figure 12. This topic of prognostics and digital twins is being addressed by the broader DT community (Woolley et. al., 2023).

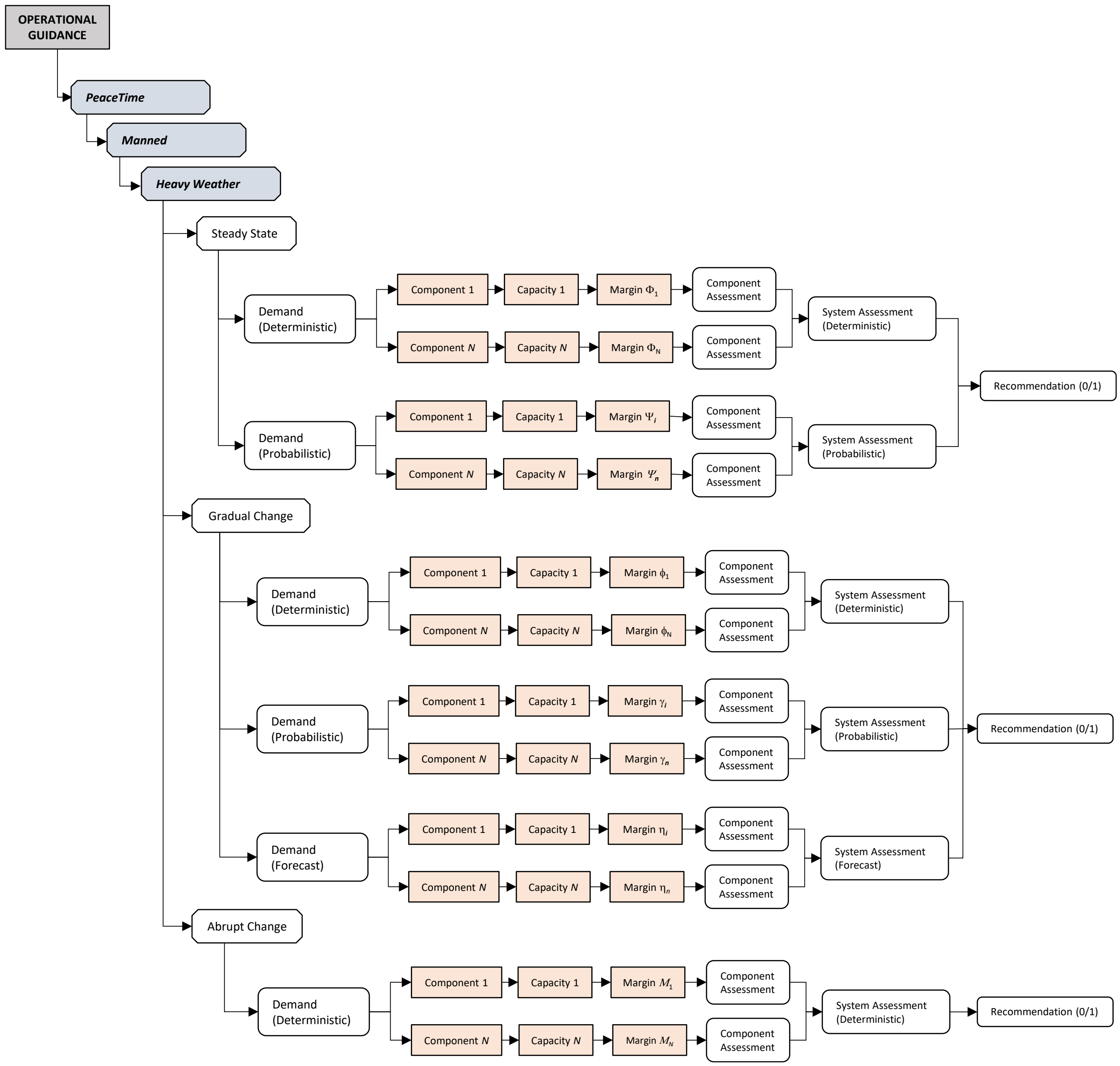


Figure 10. Logic Tree for SDT with a Performance Based Structural Assessment Approach

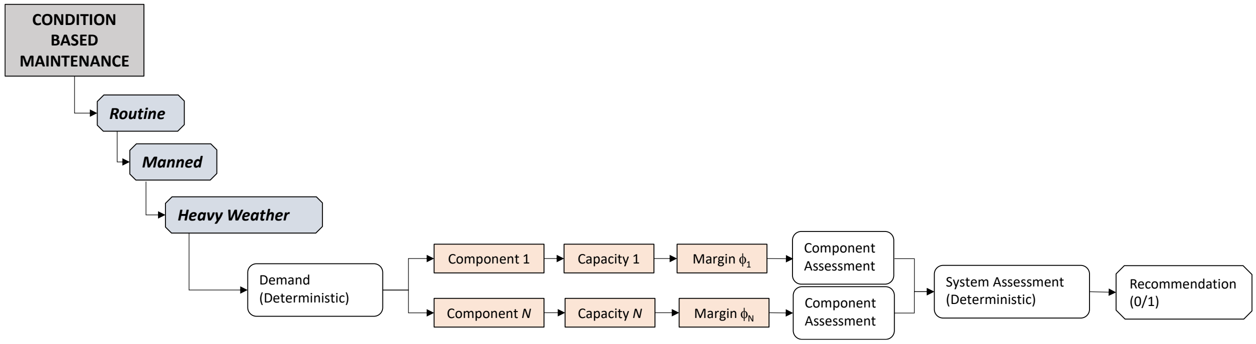


Figure 11. Example Decision Tree for Condition Based Maintenance based on Observed Data Only

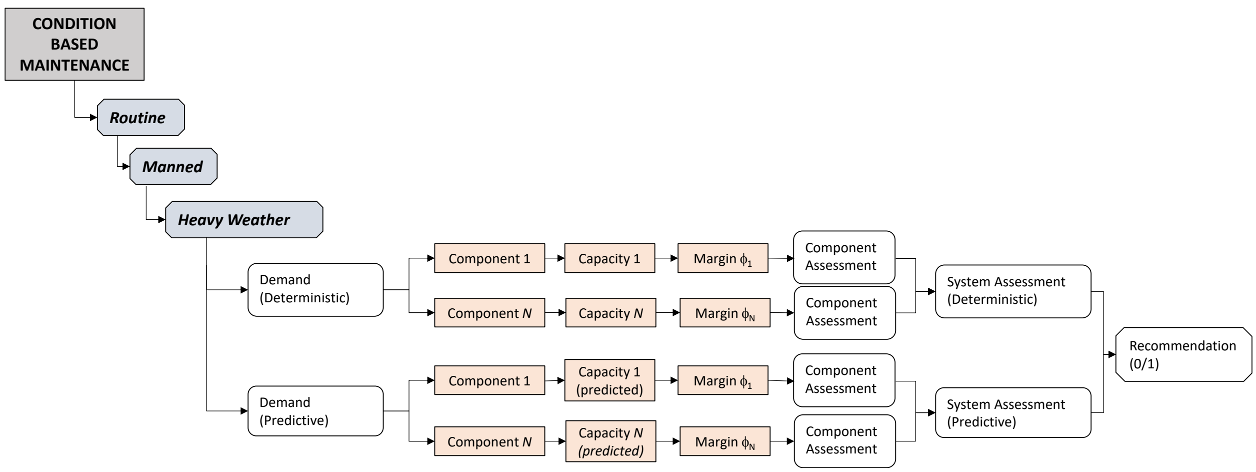


Figure 12. Example Decision Tree for Condition Based Maintenance that Includes Prognostics.

8. Validation of Structural Digital Twins

Validation of the SDT is essential prior to use. This statement may sound obvious, but it is critical, and therefore worth noting. Furthermore, the complexity of the design of SDTs dictate the need for validation and testing requirements.

The simplified definition for a SDT is data driven, physics-based models to support decisions. Unfurling the simplified definition provides a basic understanding for the validation needs for the:

1. Data: This includes the evaluation of the reliability, durability, accuracy, among others, of the (1) hardware components, (2) data acquisition system, and (3) data quality.
2. Models: This includes the evaluation of the model with respect to (1) the model’s ability to represent the physical as-built condition of the ship (or structural component), (2) the model’s ability to represent the demand (e.g., Load), and (3) the solver’s ability to appropriately assess response.
3. Decision Support Process: This includes the evaluation of the recommendations provided by the SDT (e.g., do the accuracy/uncertainty bounds meet requirements, were the correct requirements set for supporting the decision, etc.)

As the field of SDT (and the fields of Smart structures and Structural Health Monitoring) continues to develop, the validation and verification processes need to continue to develop as well. The class societies have started to document the expectations for validation and testing, as summarized in Table 7. Further work is needed to clarify the validation requirements and codify the testing methods necessary for validation.

Table 7. Class Standards and Validation & Testing Discussions

Class Standard	Validation / Testing
ABS Hull Condition Monitoring (ABS, 2020)	Operational verification procedure to be provided to ABS: Verification Procedure, covering initial set up and necessary calibration, is to be submitted for review. The procedure is to detail how to verify (1) that sensors are both operational and in adjustment or within calibration limits, as needed, and (2) how such confirmation will be needed for continued satisfactory operations to verify that the data collection, analysis, and display functions are still within function and calibration limits.
ABS SMART (ABS, 2022)	Indicates verification is needed for only tier 3 and 4 SHM systems. ABS is to verify and validate. The SHM function's capability and the health assessment results are to be demonstrated to the satisfaction of the ABS Surveyor. When employing approach of SHM Tier 3, the calibration approach utilizing data from structural sensors is to be reviewed by ABS. When employing approaches of SHM Tier 4, the risk framework, approach, and risk assessment are to be reviewed by ABS
Rules for the Classification of Steel Ships, NR467, Part F (BV, 2022)	Not required for Hull, but have requirements for trials and verification for other components
BV NR675 Additional Service Feature SMART (BV, 2021)	Not required for hull smart functions, but has requirements for Machinery and Navigation smart functions
DNVGL - Rules for classification: Ships — DNVGL-RU-SHIP Pt.6 Ch.9 (DNV GL, 2017)	Documentation requirements for Hull monitoring system - Test procedure for quay or sea trial - Report from quay and sea trial The initial readout of the sensor shall be checked against an agreed loading condition in calm water, with the attendance of a surveyor from the Society. If the difference is greater than 5% of the approved value or 10 N/mm2 occurs, whichever is the greater, the setup and subsequent checking shall be repeated. Calibration shall be verified by a surveyor from the Society. The operation of the hull monitoring system shall be verified upon installation by a surveyor from the Society: witness that the relevant procedures for testing the system are carried out. ensure that the recorded data is according to the requirement. verify that the maintenance and calibration log is complying with the relevant procedures.
DNV GL Smart Vessel (DNV GL, 2020)	Invokes DNVGL-RU-SHIP Pt.6 Ch.9 Sec.3, or system qualification (SQ) process
Lloyds Register, ShipRight, Ship Event Analysis (Llyods Register, 2021)	Not provided / not found
CCS, Rules for Intelligent Ships (CCS, 2020)	For Hull: The System testing procedure is to be submitted to CCS for information. After completion of installation of the system and equipment, survey and test are to be carried out in accordance with testing procedures to verify system function and effectiveness.
CCS, Hull monitoring and assistant decision--making system for operations in ice (CCS, 2018)	Installation, calibration, and actual verification of some tests and analysis if deemed necessary by the surveyor.
Class NK (Class NK, 2020)	Verification of installation requirement. Validation of strain level changes associating with changes in draught to be completed within 3 months of initial setup in the presence of a surveyor.

It is useful to note that CCS, Rules for Intelligent Ships have further elaboration of test and validation processes for machinery systems. This includes concepts such as

- The baseline data of the equipment and systems are to be measured in the initial healthy condition (after the run-in period) or obtained by other means. The reference condition during measurement is to be documented.
- The baseline data are in general to be measured during shipboard trials and the following requirements are to be complied with:
    1) Baseline data are to be measured by designated personnel.
    2) The measured baseline data are to cover the expected operating conditions of the equipment and systems.
    3) The effectiveness of the measured baseline data used for fault diagnosis and health assessment is to be assessed.
    4) For new equipment or equipment after major conversion, the baseline data is to be measured after a period of running.

9. Functional Approaches for Structural Digital Twin

Thus far, this document has focused on the philosophy and ontology around SDTs. This section will present the practical approaches and examples.

The approaches explored in the project address the question: “If I have a measurement at a discrete location, can the response in uninstrumented locations be approximated?”. To further bound the problem to a more feasible scope, the question is further guided by the use cases: hull girder fatigue and major structural damage. For fatigue, vertical bending moment cycles and the induced stress is the governing feature for design (DNV-GL, 2018; ABS, 2017). The stress concentration and detail related assessments all stem from the quantification of the vertical bending moment range histogram, at least in early stages of design. For operational guidance and major structural damage, typically the concern is performance of the hull girder. The smaller, component level failures have less consequence. Given that these two concepts are the main drivers for this discussion on SDT, it then gives way to scoping of the problem to focus on vertical bending moments. The philosophical question then becomes: If moment data is available for the instrumented location, can moment information in other locations be inferred? This is conceptualized in Figure 13.

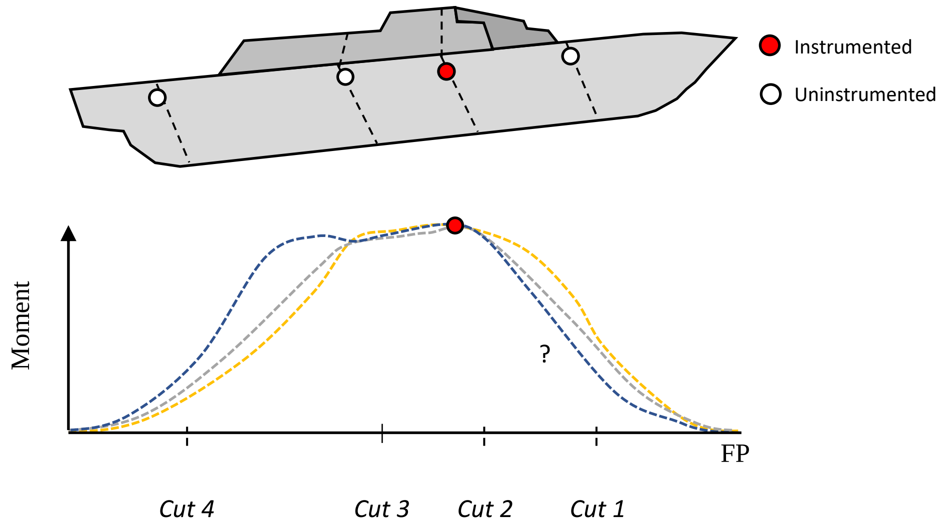


Figure 13. Philosophical Question: If data (shown as a red dot) is available for the instrumented location, can information in other locations be inferred (shown as the dashed lines)?

Therefore, the bounded scope of this project pertains to vertical bending moments (VBM) for monohulls. This is driven by the need to narrow the scope to formulate a functional solution and by the prevalence of VBM for overload and fatigue. The former supports operational guidance needs and both support condition-based inspection needs. Furthermore, the bounded scope for this project pertains to the use of strain gage data as the data sources and finite element analysis (FEA). This was due to the availability of data and the technical merit in enhancing methods for coupling data with FEA.

Since there is no direct method for measuring vertical bending moment, vertical bending moment is approximated by measuring strain in the hull girder, in areas isolated from secondary and tertiary strains, and using finite element analysis (FEA) derived calibrations to relate primary loads to strains. Assuming the strain response at a given location is the combination of vertical bending and lateral bending contributions

$$\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix} = \begin{bmatrix} CF_{1,VB} & CF_{1,LB} \\ CF_{2,VB} & CF_{2,LB} \end{bmatrix} \begin{bmatrix} VB \\ LB \end{bmatrix} \tag{1}$$

where ε_i is the strain at location *i*, *VB* and *LB* are the vertical and lateral bending moments at the frame, respectively, and $CF_{i,j}$ is the calibration factor relating the response *j* to the measurement at location *i*. To summarize at a conceptual level: this analysis requires that use of data (strains) with other physics-based models (FEA for establishing the cal factors), thus putting this into the category of a higher-order data-model fusion approach.

Section 9.1 presents a method that uses the actual strain response and the calibration approach described above. The method presented in Section 9.2 relies on the actual strain response data to be coupled with alternative FEA derived modal factors, and as such is a low-order data model coupling approach. It should be noted that "high order" and "low order" are loosely used to denote the layers of analysis that get applied to the data to support the SDT methods.

9.1. Approach A: VBM Estimation via Data, FEA, & Surrogates

9.1.1. Method: Envelope Approach

A SDT method for approximating the vertical bending moment (shortened to Moment herein) along the length of the ship based on measurement data at one location was developed and evaluated. The underlying concept for this approach borrowed from design practices, wherein the evaluation of the hull girder is performed with respect to the bending moment envelope. Typically derived from empirical equations, analytical simulations, or model testing, the vertical bending moments are quantified along the length of the ship. The envelope represents the largest possible value for the location, as shown in black in the top plot of Figure 14. For cut 1, the moment may be the largest at time a, with a moment distribution along the length of the ship shown in light blue in Figure 14. The maxima for cuts 2, 3, and 4 do not occur at the same time. Their maxima correspond to different instances in time, with different longitudinal distributions of moments.

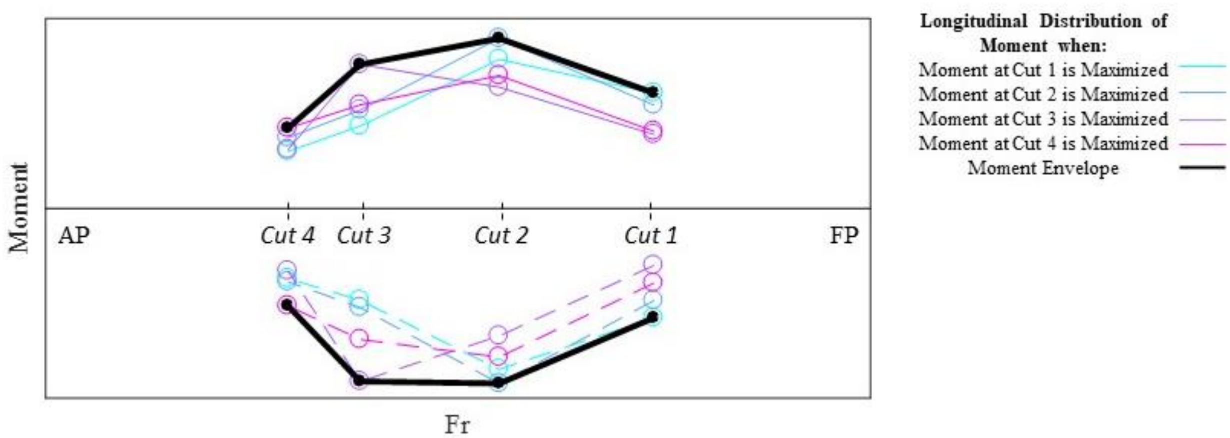
It is imperative to note that this approach is developed to support two dimensional assessments of the hull girder for bending capacity. As shown in the bottom plot in Figure 14, the derivative of the moment envelop is not the shear envelope. The shear envelope (not shown) would be the largest values of shear expected along the length of the vessel.

At all cuts, the derivative of the moment envelope has a shear value less than that of at least the cut-specific maxima.

Nevertheless, the use of moment envelope enables the evaluation of the structure to certain failure mechanisms. And, for this example application, those failure mechanisms were deemed sufficient for performance assessment. The methodology can be extended to apply to the different loads, and potentially load interactions, but for sake of time and priority, it was limited to just vertical bending moment.

The moment envelope method is developed as follows:

1. Develop an analytical model to support the estimate of seaway induced loads acting on a vessel in each operational condition.
2. Evaluate the analytical model for a given operational condition (Sea State, Speed, Relative Heading) for the vertical bending moments at the desired, uninstrumented locations, and the instrumented location.
3. Develop the moment envelope for the single operational condition (see Figure 15).
4. Repeat the simulation for all arrays of operational conditions applicable to the specific vessel and develop the associated moment envelopes.
5. Develop an aggregate moment envelope (see Figure 16).



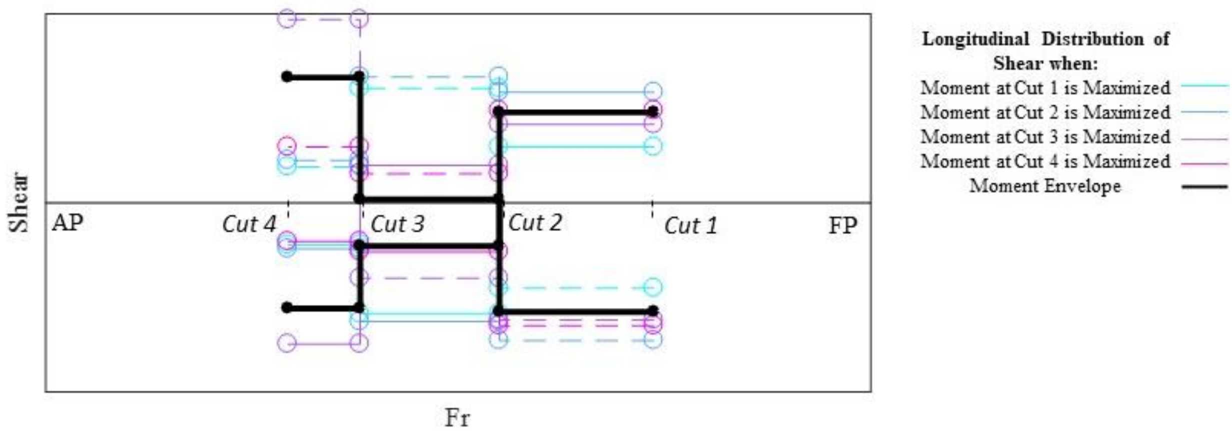


Figure 14. Moment and Shear Longitudinal Distribution

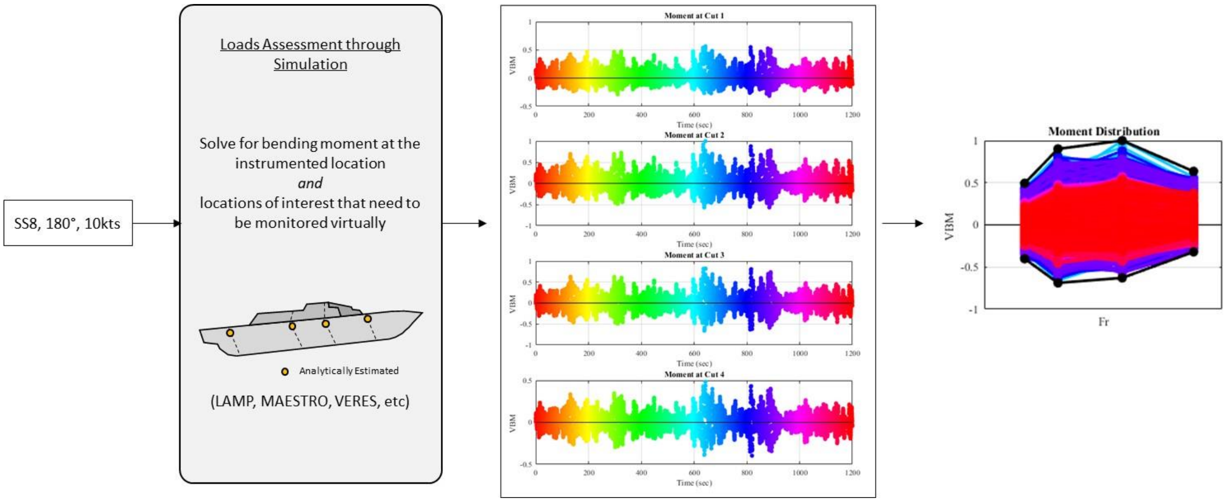


Figure 15. Process for Developing a Moment Envelope for a Single Operational Condition via Seakeeping Simulation

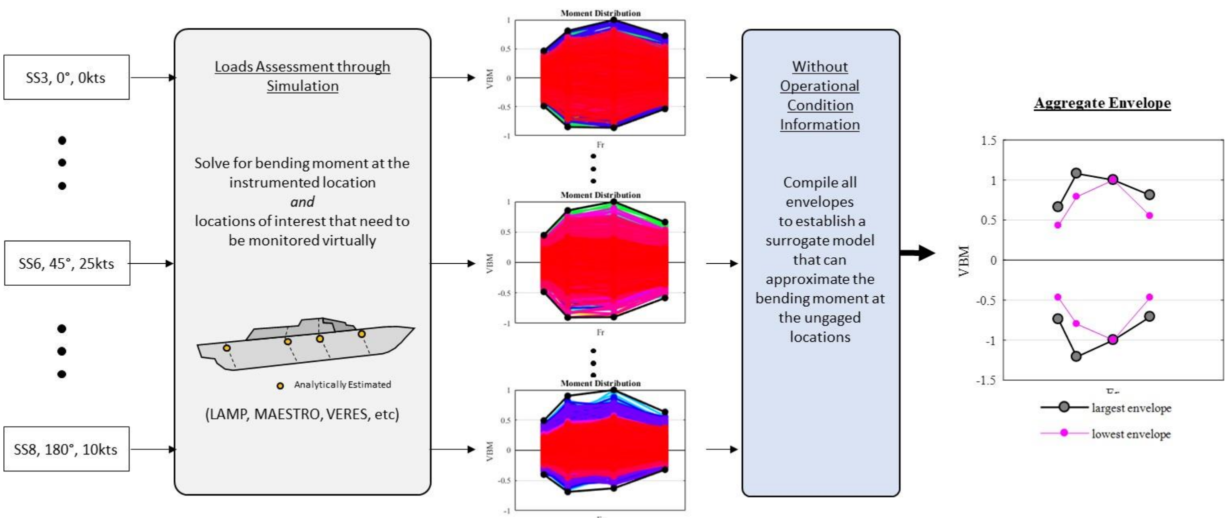


Figure 16. Process for Developing the Aggregate Envelope via Seakeeping Simulations for all Relevant Operational Conditions

The aggregate moment envelope then defines the most conservative* relationship between the moment at the measured location and the moment in other locations. Thus, the uninstrumented locations bending moment can be estimated based on the ratios from the moment envelope.

*Conservative was called out in the preceding description of the approach. This is because it is conservative when confined to the context of the analytical realm. However, the analytically derived relationship is subject to the assumptions, limitations, and accuracy of the analytical model itself, which, unless validated, contributes to unquantifiable epistemic uncertainty (see section 5). For example, Large Amplitude Motions Program (LAMP) (Lin et. al., 2008) can be developed with respect to different hydrodynamic approaches: LAMP-1 (body-linearized 3-D method) through LAMP-4 (large-amplitude 3-D body-nonlinear method). It can also be used to develop whipping-induced responses if the whipping routine is evaluated.

Another concept that falls out of the development of the largest moment envelope is the lowest moment envelop. The delta (i.e., the vertical distance) between the two curves represents the level of (quantifiable) conservativism that can be introduced by this method. Note that the lowest envelope is equal to the largest envelop only at the instrumented location.

The method was developed and applied to a notional naval vessel. Operational conditions from the expected sea states, forward speeds, and relative headings were considered. Table 8 and Table 9 present the aggregate envelopes developed for Hog and Sag. Assuming the data is available for cut 2, the largest envelope can be used to approximate the vertical bending moment response at cuts 1, 3, and 4. For both Hog and Sag, the level of conservativism in this approach is roughly around 50%. This level of conservatism could lead to over-restrictive guidance that could hinder the operations of the vessel and negatively impact the mission/transit unduly.

As such, it was hypothesized that increased knowledge, such as known speed or known speed and sea state, could improve the performance of the methods if the moment envelope was developed with respect to the conditioned information. The aggregate envelopes were thus developed from the data set, conditioned on speed. The results are shown in Table 10. No improvement was noted. Next, the aggregate envelopes were developed from the data set, first conditioned on sea state and then on speed. The results are shown in Table 11. Moderate improvement was noted when this additional information is available; the average conservativism was reduced to approximately 30%.

The envelope method is a simple to implement method for estimating bending moments in unmeasured locations. However, its natural conservativism may be too constraining to support the transition of this approach to support shipboard guidance.

Table 8. Aggregate Envelope: Hog

Normalized such that Cut 2 =1

	Cut 1	Cut 2	Cut 3	Cut 4
Largest Envelope	0.81	-	1.08	0.66
Lowest Envelope	0.55	-	0.79	0.43
Potential Conservativism	48%	-	36%	51%

Table 9. Aggregate Envelope: Sag

Normalized such that Cut 2 =1

	Cut 1	Cut 2	Cut 3	Cut 4
Largest Envelope	0.71	-	1.21	0.74
Lowest Envelope	0.47	-	0.80	0.47
Potential Conservativism	52%	-	50%	56%

Table 10. Aggregate Envelope Conditioned on Speed Only

Speed (kn)	Field	Hog: Normalized such that Cut 2=1				Sag: Normalized such that Cut 2=1			
		Cut 1	Cut 2	Cut 3	Cut 4	Cut 1	Cut 2	Cut 3	Cut 4
5	Largest Envelope	0.79	-	1.05	0.62	0.80	-	1.31	0.74
	Lowest Envelope	0.54	-	0.79	0.45	0.51	-	0.86	0.48
	Potential Conservativism	45%	-	32%	37%	57%	-	51%	55%
10	Largest Envelope	0.81	-	1.04	0.56	0.75	-	1.24	0.70
	Lowest Envelope	0.56	-	0.77	0.41	0.47	-	0.84	0.47
	Potential Conservativism	45%	-	35%	38%	61%	-	48%	48%
15	Largest Envelope	0.80	-	1.08	0.66	0.73	-	1.23	0.71
	Lowest Envelope	0.57	-	0.77	0.40	0.46	-	0.82	0.45
	Potential Conservativism	41%	-	40%	62%	56%	-	51%	58%
20	Largest Envelope	0.77	-	1.01	0.57	0.77	-	1.17	0.70
	Lowest Envelope	0.54	-	0.82	0.46	0.51	-	0.75	0.46
	Potential Conservativism	41%	-	24%	23%	51%	-	55%	52%
25	Largest Envelope	0.74	-	1.04	0.58	0.71	-	1.13	0.61
	Lowest Envelope	0.52	-	0.80	0.43	0.53	-	0.80	0.44
	Potential Conservativism	42%	-	31%	34%	33%	-	41%	41%

Table 11. Aggregate Envelope Conditioned on Speed and Sea State

Sea State	Speed (kn)	Field	Hog: Normalized to Mid Ship =1				Sag: Normalized to Mid Ship =1			
			Cut 1	Cut 2	Cut 3	Cut 4	Cut 1	Cut 2	Cut 3	Cut 4
6	5	Largest Envelope	0.79	-	0.97	0.55	0.68	-	1.20	0.69
		Lowest Envelope	0.61	-	0.83	0.47	0.55	-	0.86	0.51
		Potential Conservativism	28%	-	17%	18%	24%	-	39%	35%
6	10	Largest Envelope	0.81	-	0.99	0.56	0.67	-	1.06	0.63
		Lowest Envelope	0.63	-	0.84	0.47	0.52	-	0.93	0.53
		Potential Conservativism	29%	-	19%	20%	28%	-	14%	18%
6	15	Largest Envelope	0.75	-	0.96	0.53	0.69	-	1.11	0.71
		Lowest Envelope	0.58	-	0.82	0.44	0.54	-	0.84	0.52
		Potential Conservativism	30%	-	17%	20%	27%	-	31%	37%
6	20	Largest Envelope	0.76	-	0.97	0.53	0.67	-	1.13	0.70
		Lowest Envelope	0.57	-	0.83	0.46	0.54	-	0.88	0.47
		Potential Conservativism	33%	-	16%	14%	24%	-	28%	48%

9.1.2. Method: Multivariate Lagged Regressive Model

Discrete sensing along a ship lends insight into the bending moments the ship is experiencing only at the discrete sensor locations. However, the ship's responses at all locations are highly correlated with each other. That is, the response in one location is highly correlated to the response at other locations. This fact leads to the existence of a model which relates the ship response in one location to expected response in a separate location. The ship response is directly correlated through the ship's dynamic, section, and elastic properties which are not always available and may be affected by the operational regime. One class of solutions to this problem is data driven regressive modeling. These solutions seek to derive a model for one state, based on a combination of other states. One of these methods is a lagged regression. Lagged regression (LR) refers to the estimation of one state variable based on a combination of past readings of a different state variable, defined below.

$$x_t = C + B_0 y_t + B_1 y_{t-1} + B_2 y_{t-2} + \dots + B_n y_{t-n} \tag{2}$$

9.1.2.1. Initial Assessment

In the context of ship primary loads, the unknown state may be the vertical bending moment at a non-instrumented location, and the independent variable may be the measured vertical bending moment at an instrumented location. The LR model is built by finding the coefficients which minimize the residuals between the estimated state and the actual measurement. The initial assessment for this method was completed by examining the residual error incurred at different lag lengths using the vertical bending moment measured from a single experimental run condition of a vessel underway.

$$\boldsymbol{VBM_{1_t} = C + \sum_{i=0}^{n} B_i VBM_{1_{t-i}}} \tag{3}$$

The results from the initial assessment are shown in Table 12. Since the residual errors are minimized for the development of the model, the low errors in standard deviation are to be expected. Typically, this data is used to support performance assessment, in which the maximum (max) and minimum (min) are the most critical. Therefore, the errors in the max and min are reported.

The results suggest that by taking more fitting terms, or lags, the model will more closely match the maximum and the minimum values. However, as more terms are taken into the model, the error does not decrease as quickly and in the case of max value error, the residual even increases. Moreover, as more terms are included, there is a higher risk of over-fitting and having an unstable model.

Table 12. The residual error, as measured by the difference in maximum, minimum, and standard deviation.

Case	Lag (s)	Dependent Variable	Independent Variable	Error in |Max Value|	Error in |Min Value|	Error in Standard Deviation
1	0.05	VBM_1	VBM_2	-21%	-50%	-2%
2	0.5	VBM_1	VBM_2	-9%	-23%	-1%
3	1	VBM_1	VBM_2	-5%	-20%	-1%
4	2	VBM_1	VBM_2	-4%	-20%	-1%
5	4	VBM_1	VBM_2	-5%	-16%	-1%
6	6	VBM_1	VBM_2	-5%	-15%	-1%
7	8	VBM_1	VBM_2	-7%	-13%	-1%

Another possible method to reducing the residual error incurred is to use a multi-variate lagged regression (MVLR) which adds additional sensor data into the model. One example may be to include another VBM reading from a different location.

$$\boldsymbol{VBM_{1_t} = K + \sum_{i=0}^{n} P_i VBM_{2_{t-i}} + \sum_{i=0}^{n} Q_i VBM_{3_{t-i}}} \quad (4)$$

Table 13 shows the results from including another sensor reading into the MVLR model. In the case of .05s lag, the new model exhibited lower error than the single independent variable counterpart. However, for the other cases, little advantage was found.

Table 13. The residual error from the MVLR model as measured by the difference in maximum, minimum, and standard deviation.

Lag (s)	Dependent Variable	Independent Variable	Error in |Max Value|	Error in |Min Value|	Error in Standard Deviation
0.05	VBM_1	VBM_2	-21%	-50%	2%
0.5	VBM_1	VBM_2	-9%	-23%	1%
1	VBM_1	VBM_2	-5%	-20%	1%
2	VBM_1	VBM_2	-4%	-20%	1%
4	VBM_1	VBM_2	-5%	-16%	1%
6	VBM_1	VBM_2	-5%	-15%	1%
8	VBM_1	VBM_2	-7%	-13%	1%
0.05	VBM_1	VBM_2, VBM_3	-12%	-34%	2%
0.5	VBM_1	VBM_2, VBM_3	-7%	-23%	1%
1	VBM_1	VBM_2, VBM_3	-4%	-20%	1%
2	VBM_1	VBM_2, VBM_3	-3%	-20%	1%
4	VBM_1	VBM_2, VBM_3	-3%	-18%	1%
6	VBM_1	VBM_2, VBM_3	0%	-12%	4%
8	VBM_1	VBM_2, VBM_3	-5%	-14%	1%

9.1.2.2. Effects of Operating Condition

The residual errors shown above are measures of how well the models describe the actual relationship between sensors for a single run. However, the relationship between the sensors may change depending on operating condition. Figure 17 shows the cross correlation between two sensors at various speeds. The different operating conditions lead to cross-correlation function which show vastly different behavior. This implies that the relationship between the two sensors depends on the operating condition and the MVLR method may be susceptible to error when the model is applied across operating conditions.

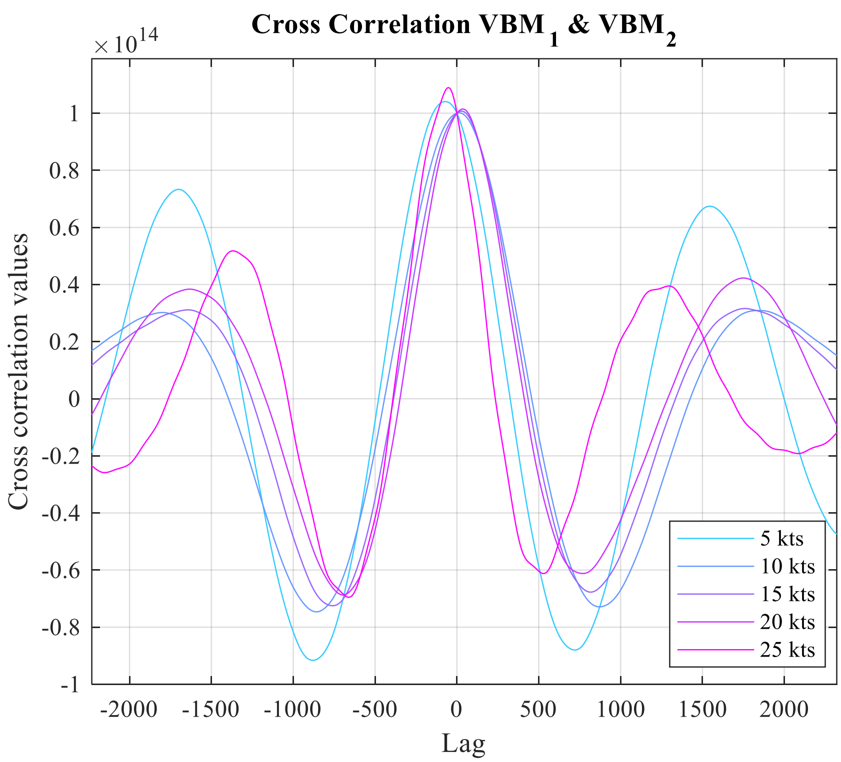


Figure 17. The cross correlation between two VBM measurements varies with respect to the vessel's speed.

Further validation efforts for the MVLR method were done using cross validation where the model is trained on one set of data and then applied to another. It should be noted that this method requires training data to be available at the non-instrumented location. Once the model is developed, however, it may be applied to non-training data which does not include measurements at the non-instrumented locations and the model error may be assessed. A new normalized error metric was included in the assessment to scrutinize the performance of the MVLR model more closely. Table 14 shows the mean squared error for each training set, testing set pair with respect to vessel speed. The error values represent the similarity between two signals (Perlin & Bustamante, 2016). In this case the two signals are the predicted signal from the MVLR model, and the actual signal as measured. The values for the mean squared error vary between zero and one; zero being a perfect similarity and one being complete dis-similarity. The first take away is a significant difference in error depending on which speed was used to train and which speed was used to test. This implies that a model created using data from one speed condition will have varying error depending on the speed of the test data. Secondly, the lowest error does not always occur when the training speed and the testing speed are the same. This may be due to the presence of other moderating variables such as heading.

Table 14. The mean squared error shows variation depending on what speed was used to train the model as well as the speed at which the model was tested.

		Testing Speed (kn)				
		5	**10**	**15**	**20**	**25**
Training Speed (kn)	**5**	0.176	0.164	0.160	0.176	0.202
	10	0.172	0.167	0.159	0.172	0.206
	15	0.174	0.169	0.161	0.172	0.205
	20	0.177	0.173	0.168	0.179	0.205
	25	0.230	0.216	0.225	0.232	0.187

9.1.2.3. Diverse Model Training

The above results show that the MVLR error is influenced by changes in operating condition. Therefore, operating condition must be considered to expand the applicability of the MVLR model. One option is to train the model using data from a diverse set of operating conditions. A lagged regression (LR) model was constructed by optimizing the coefficients across a full day of the trials. This set includes 35 unique conditions which span a wide variety of speeds and headings. The resulting model was validated using the remaining trials from other days. The model resulted in an average mean squared error of .21 and a median mean squared error of .17. The overall error distribution is shown in Figure 18. The errors are heavily skewed with the most errors being below .20 across all operating conditions.

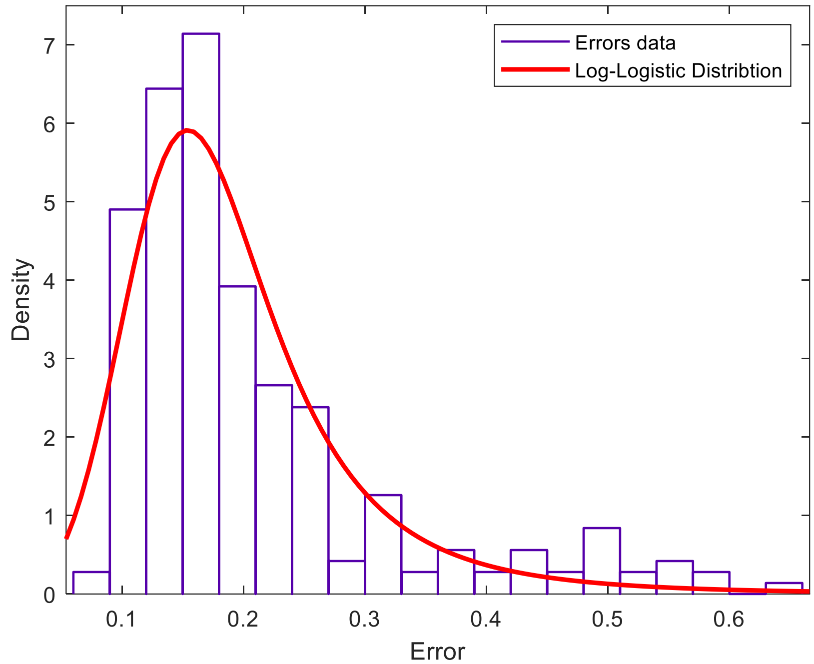


Figure 18. Error distribution for the single dependent variable LR model trained on a diverse set of operating conditions.

Table 15 shows the error values for the best and the worst test cases respectively. Figure 19 and Figure 20 show the worst and the best test cases tested using the diverse trained LR model, respectively. In the best case, the low mean squared error shows a high degree of similarity between the predicted and the actual signal. The difference in sign for the errors in maximum and minimum values lends insight into the possible source of error. A positive error shows the predicted value is further from zero than the actual value. Therefore, for run Condition 1, a large portion of the error comes from a bias shift. In other words, some error could be accounted for by shifting the mean of the predicted signal. On the other hand, trial Condition 2 does not show this issue. The maximum and minimum value errors are both negative which implies that the prediction is dilated. However, this dilation does not account for the much larger mean squared error. Indeed, Figure 19 shows that a large portion of the error is being contributed from an apparent phase shift which is introduced by the predicted signal. This behavior is often a product of non-modeled phenomena.

Table 15 Error values for the diverse trained LR model for the worst and best test cases

	Error			
	Mean squared	Max Value	Min Value	Standard Deviation
Condition 1	0.087	11.9%	-7.5%	4.80%
Condition 2	0.683	-20.4%	-19.1%	2.80%

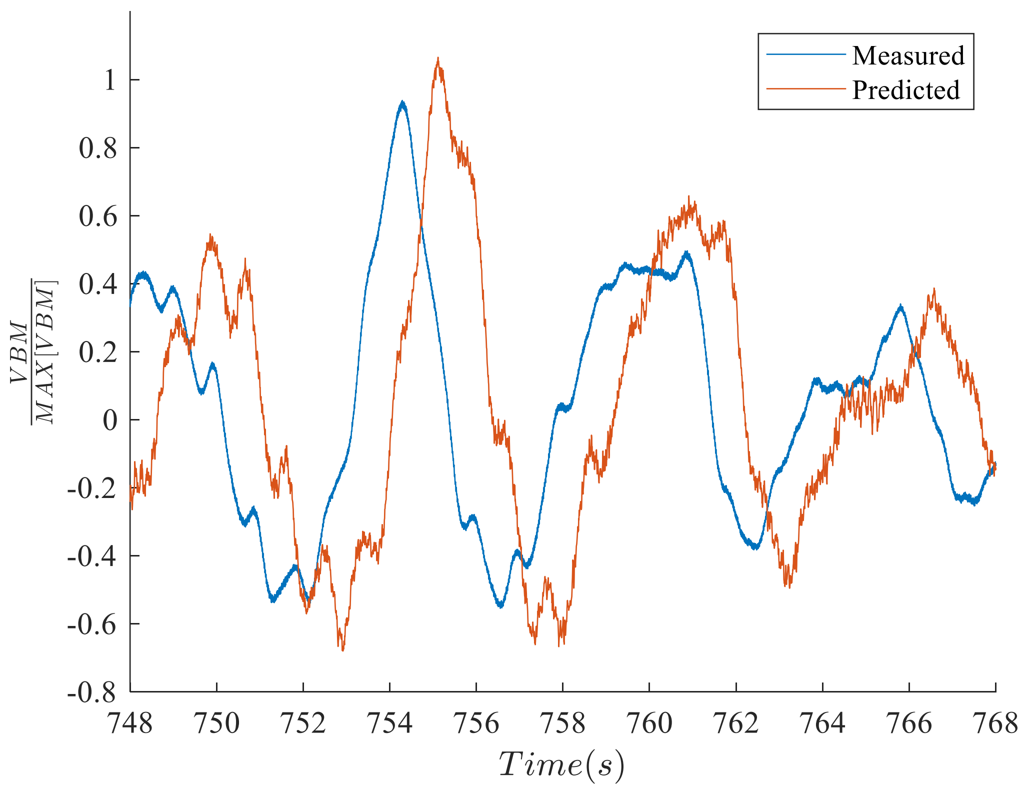


Figure 19.Measured and LR prediction response for Condition 1

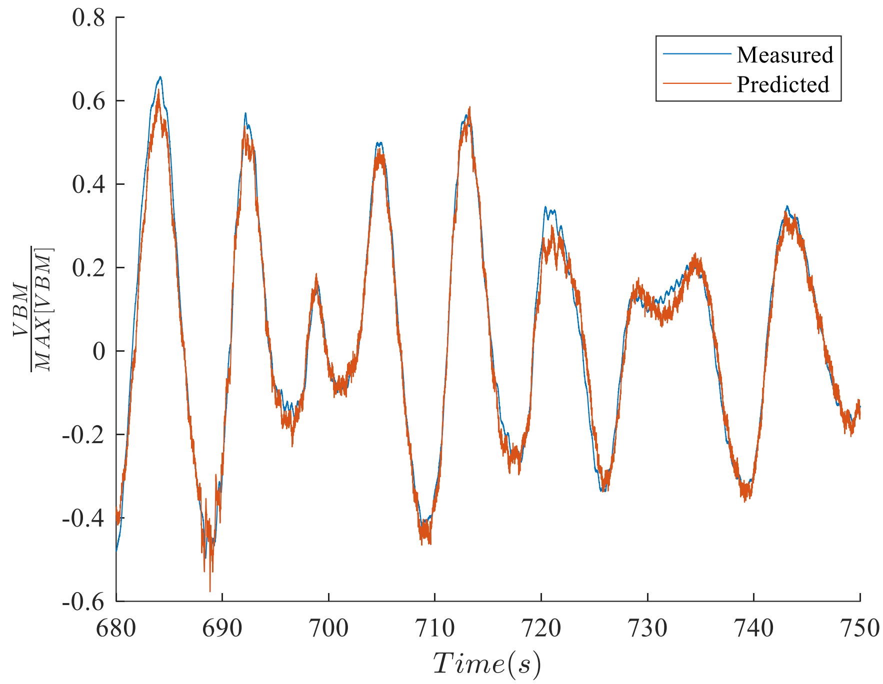


Figure 20. Measured and LR prediction response for Condition 2

Figure 21 through Figure 23 show the error distribution associated with the LR model at each operating condition. For the heading and the sea state cases, the central tendency of the error remains consistent across conditions. However, this is not the case for speed. As the speed varies, the central tendency for the error distributions also changes significantly even when the model was trained using a diverse set of conditions. Also, in all three cases, the skewness and the variance of the distributions are inconsistent across operating conditions. Therefore, it may be concluded that even when a diverse set of training data is used, the error between the predicted signal and the actual signal remains highly associated with the operating condition.

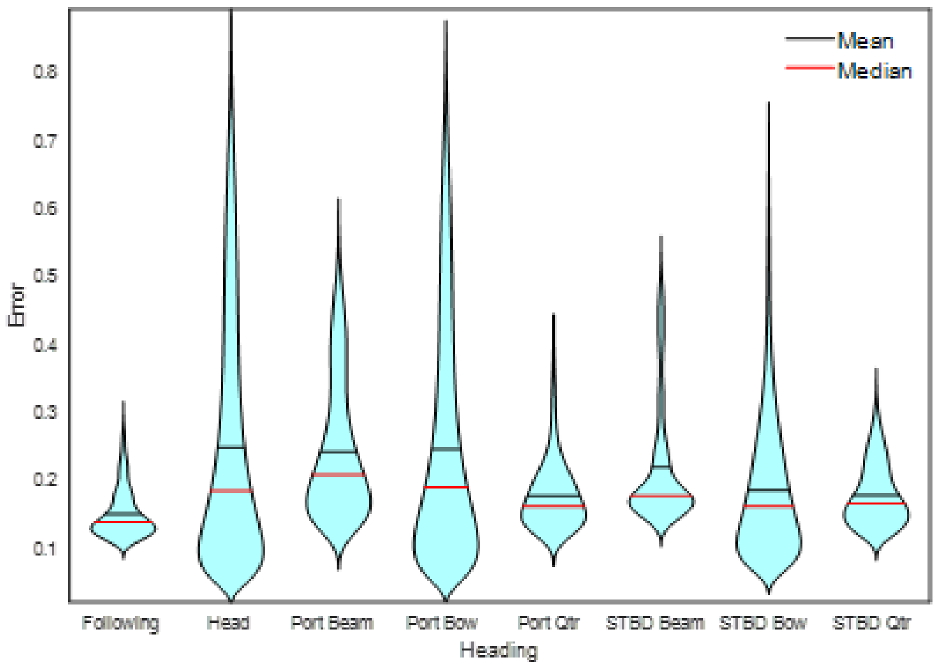


Figure 21. The error distributions at each heading tested show similar central tendencies and some variation in skewness and variance.

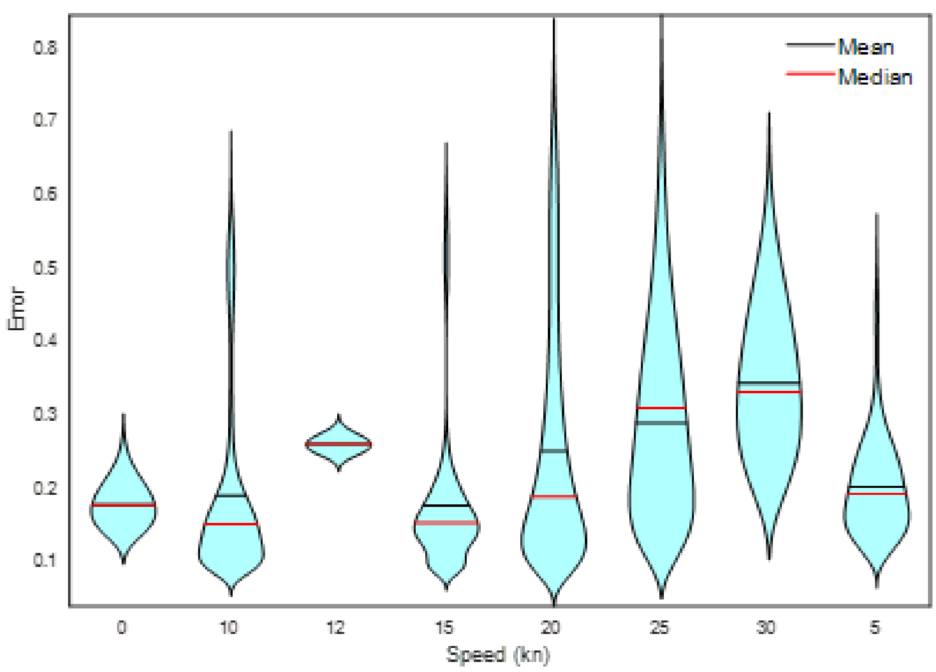


Figure 22. The error distribution at each speed tested shows significant differences in central tendency as well as skewness and variance.

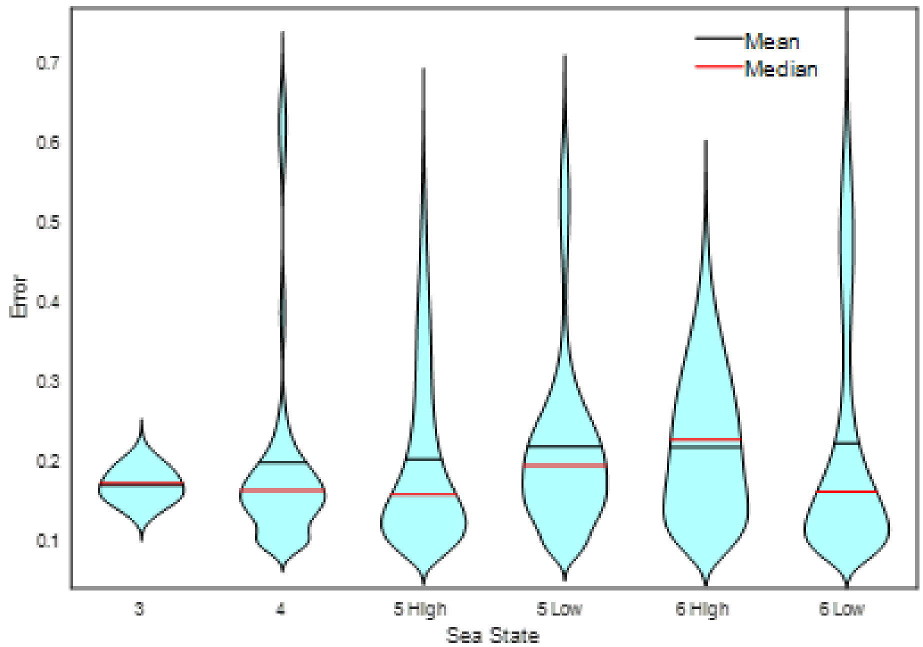


Figure 23. The error distributions at each sea state tested show similar central tendencies and some variation in skewness and variance.

A MVLR model was derived using a diverse set of training data to counter the effects of operating conditions. However, the operating conditions were not included in the model itself. Therefore, it is possible that the dilation and the phase error are due to operating condition effects which are not modeled.

In section 9.1.2.1, the residual error was shown to decrease when a second sensor was used in the modeling effort. As such, a multi-variate lagged regression (MVLR) model was trained using two sensors instead of one at the same training conditions as before. The overall error distribution is shown in Figure 24, and Table 16 shows the cross-validation error values when two sensors are used in the MVLR model. For both cases, the mean squared error decreased. For the other three errors, namely the maximum value, minimum value, and standard deviation error, the model performance shows split results. Condition 1 exhibited lower errors across the board and Condition 2 exhibited higher errors across the board. This divergence in results further implies that the response under the conditions of Condition 2 is not modeled well using the response at other conditions, and operating conditions must therefore be accounted for separately.

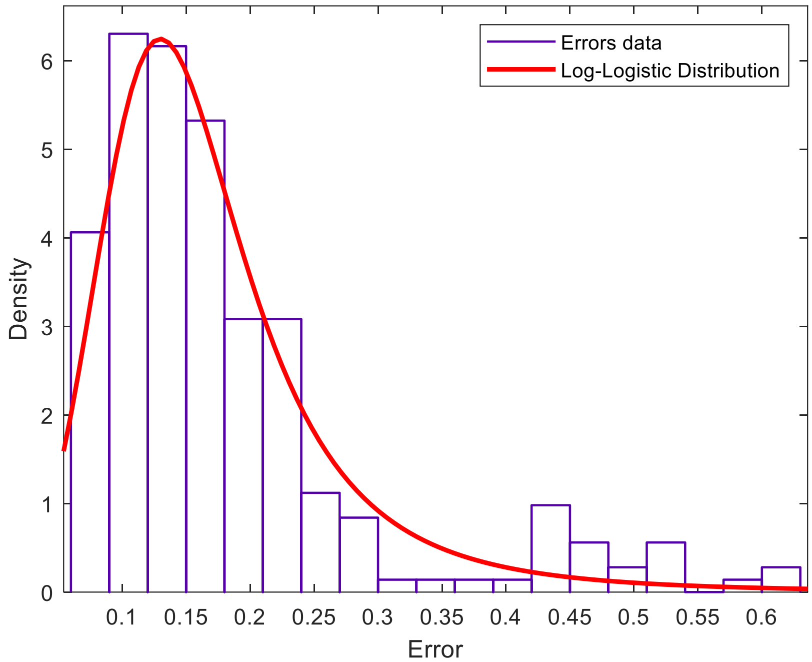


Figure 24. Error distribution for the two-regression variable MVLR model trained on a diverse set of operating conditions.

Table 16: Error values for the two-regression variable MVLR model trained on a diverse set of operating conditions.

Condition	Error			
	Mean squared	Max Value	Min Value	Standard Deviation
Condition 1	0.074	7.0%	-2.5%	3.5%
Condition 2	0.636	-26.3%	-26.1%	9.1%

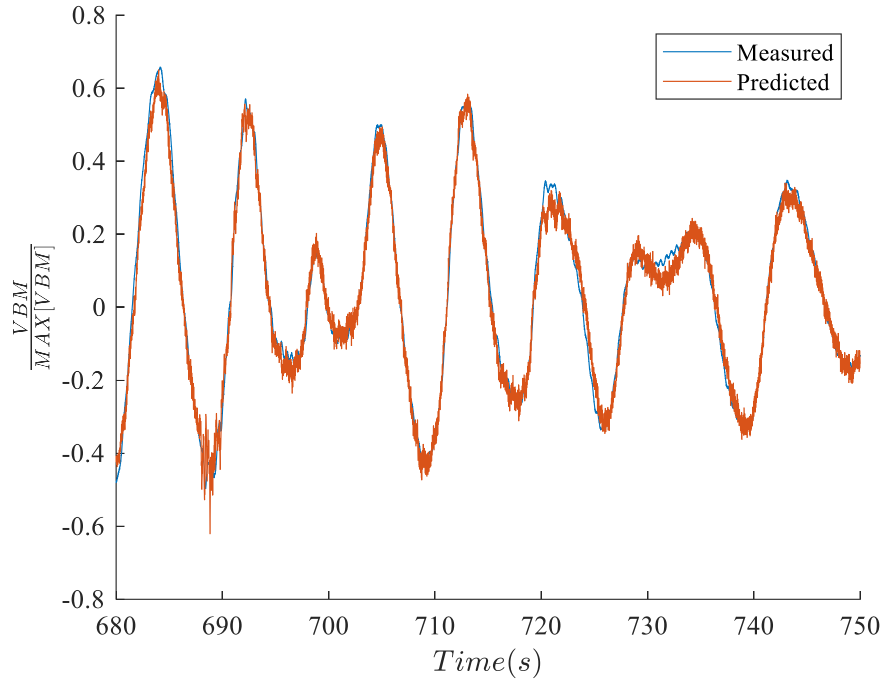


Figure 25. Measured and two-regression variable MVLR model trained on a diverse set prediction response for Condition 1

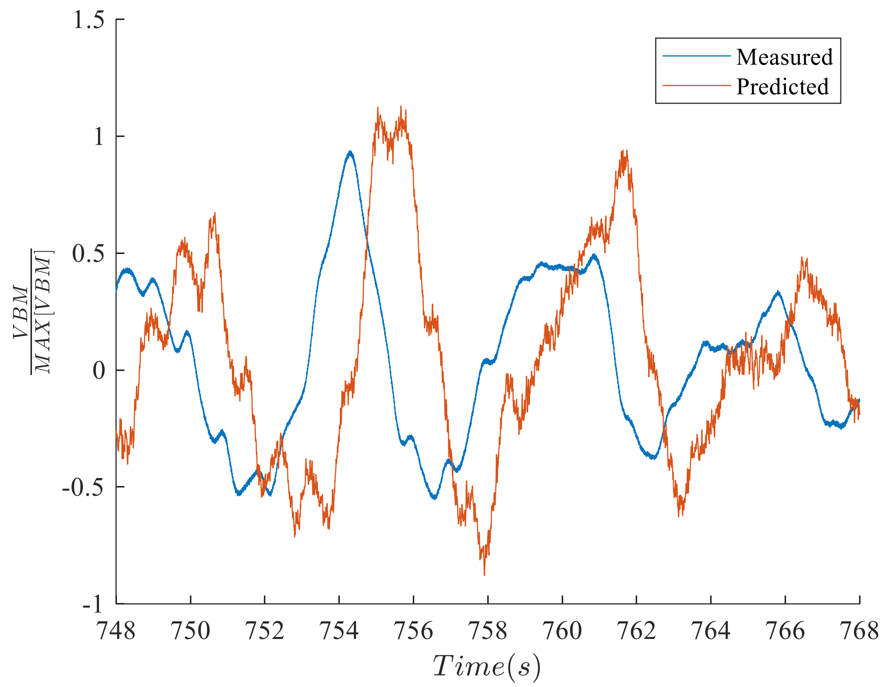


Figure 26. Measured and two-regression variable MVLR model trained on a diverse set prediction response for Condition 2.

9.2. Approach B: VBM Estimation via Data, FEA, & Structural Modes

Approach A (presented in Section 8.1) starts with a formulation for estimating structural loads (also referred to as internal loads) from strain gages. Approach B keys in on this part of the SDT: how are loads estimated from measurements.

9.2.1. Theory: Formulating the Strain-to-Load Conversion Problem

Many methods have been developed to infer the internal loads developed within structures from imperfect strain, displacement, velocity, and/or acceleration instrumentation measurements. This is, generally, a very ill-posed, underdetermined problem, as only a few discrete measurements are used to infer the response of the entire structure. This is especially true for very large structures such as surface ships, where the global loads experienced by the hull girder are typically calculated via measurements from a series of perhaps ten to twenty primary strain gauges.

For the evaluation of surface ship global loads, the assumption is typically made that the relationship between strain and the internal loads induced in the hull girder are linear, with these linear relationships quantified in a conversion matrix. As will soon be discussed, a variety of methods exist to generate these matrices and compute internal hull girder loads from measured strains. The differences between these methods stem, primarily, from the selection of basis vectors used in defining the conversion matrix, and the method of solution applied to the resulting system of equations.

The basis vector selection gives the method its physical significance; typically, these vectors are linearly independent modes of the structural response, as computed from finite element (FE) model of the hull girder. The method of solution is independent of information preserved within the selected basis. However, the relative number of strain gauge measurements, basis vectors, and hull girder loads to be computed dictates the dimensions of the system of equations to be evaluated. These relative dimensions dictate whether the system is determinate or indeterminate (overdetermined, or underdetermined), each requiring different methods of solution.

To compute a well-conditioned, relatively insensitive solution for the internal loads, provided some discrete set of imperfect strain measurements, the system of equations must be, at a minimum, determinate or, preferably, overdetermined. Due to the errors inherent in strain sensor measurements, it is preferrable to have an overdetermined system, as multiple conditions/constraints can be applied to the same variables/DOFs of the model, while not strictly enforcing any of them (as this would increase the sensitivity of the system); i.e, soft constraints.

As such, a novel methodology is presented herein that employs the assumptions of linear elasticity, as well as the scalar amplification and linear independence properties of eigenmodes, in order to generate a well-conditioned, overdetermined system of equations for computing internal load.

9.2.2. Formulating the Strain-to-Load Conversion Problem

Bigot et. al. (2013) details the definition of the strain-to-load conversion matrix and mathematical formulation of the overall problem. It is summarized here for the sake of completeness, though the notation used herein differs from the original summary. Bold letters denote vectors and matrices, whereas scalars are denoted by normal text. Superscripts denote the dimension(s) of the given variable, and subscripts denote variable counters.

As with many problems of practical importance, this problem is formulated as a linear relationship between a set of inputs and a set of outputs. In this case, the input is a set of p strain measurements, $\boldsymbol{\varepsilon}^{p}$, extracted from gauges/sensors instrumented on an in-service platform, and the output is a set of q corresponding cross-sectional hull girder internal loads, $\boldsymbol{f}^{q}$. They are related through a conversion matrix, $\boldsymbol{C}^{qxp}$, where

$$\boldsymbol{C}^{qxp}\boldsymbol{\varepsilon}^{p} = \boldsymbol{f}^{q} \tag{5}$$

It should be noted that, in general, the quantity of strain measurements is not equal to the desired quantity of cross-sectional hull girder loads, i.e., $p \neq q$.

By defining the measured strain data as the superposition of a set of orthogonal modal basis vectors, $\boldsymbol{\varepsilon}^p$ can be defined as

$$\boldsymbol{\varepsilon}^p = \boldsymbol{\Psi}^{pxm}\boldsymbol{\alpha}^m \equiv \sum_{i=1}^{m} \boldsymbol{\psi}_i^p \alpha_i \tag{6}$$

where $\boldsymbol{\Psi}^{pxm}$ is a matrix composed, column-wise, of m modal strain basis vectors $\boldsymbol{\psi}_i^p$, and $\boldsymbol{\alpha}^m$ is the corresponding vector of m scalar modal amplitudes α_i. Note that i is the counter for the modal basis vectors, where $i = 1, \ldots, m$.

If displacements are small and strains can be approximated as "infinitesimal", the relationship between strain and internal load can be assumed linear. This assumption allows for the computation of the cross-sectional loads via the superposition of modal load vectors, corresponding to the same basis as the set of orthogonal modal strain vectors. That is

$$\boldsymbol{f}^q = \mathbf{R}^{qxm}\boldsymbol{\alpha}^m \equiv \sum_{i=1}^{m} \boldsymbol{r}_i^q \alpha_i \tag{7}$$

where $\mathbf{R}^{qxm}$ is a matrix composed, column-wise, of m modal load basis vectors $\boldsymbol{r}_i^q$, and $\boldsymbol{\alpha}^m$ is, again, the corresponding vector of m scalar modal amplitudes α_i. Note that the vectors of modal amplitudes $\boldsymbol{\alpha}^m$ in equations 6 and 7 are equivalent. Under this assumption, they can be manipulated and combined to yield the equation below

$$\boldsymbol{f}^q = \mathbf{R}^{qxm}(\boldsymbol{\Psi}^{pxm})^{-1}\boldsymbol{\varepsilon}^p \tag{8}$$

This methodology relies on the inversion of $\boldsymbol{\Psi}^{pxm}$. However, if the number of strain measurements does not equal the number of vectors composing the modal basis, $\boldsymbol{\Psi}^{pxm}$ is not invertible, and the system of equations becomes either underdetermined ($p < m$) or overdetermined ($p > m$).

9.2.3. General Notes Regarding Load-to-Strain Conversion Methods

The standard approaches to the strain-to-load conversion problem have their shortcomings. The goal of these methodologies is for $\boldsymbol{C}^{qxp}$ to contain, implicitly, all possible physical relationships between strain and internal load. Since all physical relationships are, to some extent, nonlinear, each element of $\boldsymbol{C}^{qxp}$ should be a function of $\boldsymbol{\varepsilon}^{p}$ and $\boldsymbol{f}^{q}$ (and, most likely, many other additional factors), not a constant. Additionally, since the data is so sparse and the system of equations is so underdetermined, knowledge of these nonlinear relationships only for the DOFs included in the qxp system in equation 5 does not provide sufficient information to infer the response at other DOFs. This is because, unlike linear systems, the ratio of responses between DOFs isn't constant.

This is one of the major issues with all underdetermined inverse problems; as the phenomena being simulated become more complex and nonlinear, more data and physical relationships are required to properly model them. As such, the ability to employ the assumptions of linear phenomena is ideal, whenever possible, as it provides additional relationships to exploit between non-local DOFs. Luckily, in the case of surface ship hull girder response, the assumption of linear, small displacement, infinitesimal strain response is typically reasonable, as most sea states experienced by the hull girder do not induce large structural response; at least globally. As such, the elements of $\boldsymbol{C}^{qxp}$ can be reasonably approximated as constants.

The next issue involves the determination and computation of the basis vectors composing $\boldsymbol{C}^{qxp}$. These bases are what give these methodologies their physical significance. Typically, these vectors correspond to set of linearly independent distortion modes (i.e, a generalized modal basis) computed via a numerical model of the ship's hull girder. However, a wide variety of methods exist in the selection and computation of these modal responses. Ultimately, the methodology selected depends on whether the generalized modal basis employed can produce a span capable of reconstructing the response for the given loading.

One of the simplest approaches is to generate the modal generalized basis from modes of response computed by an FE model of the ship hull girder, subjected to a series of pre-computed shear forces along its length, in order to induce specified shear and moment distributions. Basis vectors have also been generated via linear, frequency-domain seakeeping simulations, where loads corresponding to waves of varying heading and frequency are applied to the FE model of the hull girder; see Bigot et. al. (2015), Jiang (2021) and Wang et. al. (2021) Note, however, that for both cases, the computed responses are not guaranteed to be linearly independent, or even to have low cross-correlation. Though, in the latter case, methods have been semi-automated to extract the set of least-correlated modes from the pool of extracted modal responses. Alternatively, the natural modes of vibration of the hull girder, computed via a structural dynamic FE eigen analysis, can also be employed to generate the basis; see Bigot et. al (2013), Baudin et. al. (2013), and Jiang (2021).

While the methodology discussed herein applies more to the solution of the strain-to-load conversion, it is important to note that the basis selected must be physically accurate to be efficient. Bigot et. al. (2013) compared the use of a dry free vibration modal basis with a basis generated from linear, frequency-domain seakeeping simulation loads, for the case of an Ultra Large Container Ship (ULCS). It was determined that the modes generated from the responses to linear seakeeping simulation-generated loads were more efficient than the dry modal basis. However, as the authors note, this requires further investigation. While the dry free vibration modes may not provide the best modal basis for these methods, that does not preclude the use of free vibration modes. It is recommended that research into the use of a wetted free vibration modal basis be performed. Simple boundary element methods can be used to compute eigenmodes with fluid added mass This type of method could potentially provide more physical accuracy than a dry modal basis, more generality and efficiency than using a linear seakeeping simulation to compute distortion modes, and it would ensure linear independence of the modal generalized basis.

9.2.4. *Improved Strain-to-Load Conversion Methodology*

This section presents an improved strain-to-load conversion methodology. The inputs are the displacement modes of the hull girder, as computed via an FE structural dynamic eigen-analysis. As such, linear independence of the responses is ensured. The measured data and relationships provided by the FE model are then manipulated in such a way as to generate an overdetermined system of equations for computing internal load.

Regarding notation, bold, capitalized variables refer to matrices, bold, lower-case variables refer to vectors, and non-bold characters refer to scalars. Superscripts define the dimensions of the given variable, and subscripts define its counter. Also note that, when describing dimensions (superscripts), all dimensions in parentheses correspond to a single dimension; e.g, a variable $\boldsymbol{a}^{(bxc)}$ is a vector of length (bxc), and $\boldsymbol{A}^{(bxc)x(dxe)}$ is a matrix with (bxc) rows and (dxe) columns.

This process will first be demonstrated for the case of computing the hull girder cross-sectional load at q section cuts. The first step is to explicitly define the elements of the conversion factor matrix. For this method, this will be an $(mxqxp)xp$ matrix, where, like in the previous sections, m is the total number of modes forming the modal generalized basis, p is the total number of strain gauge measurements composing $\boldsymbol{\varepsilon}^p$, and q is the total number of section cuts at which to compute cross-sectional internal loads; the counters employed for the totals m, p, and q are i, j, and k, respectively. Each element of the conversion matrix corresponds to the i^{th} modal ratio of the k^{th} modal cross-sectional load r_k^i to the corresponding j^{th} strain gauge measurement location modal response ψ_j^i, as extracted from the FE model of the hull girder; i.e,

$$\boldsymbol{C}^{(mxqxp)xp} = \begin{bmatrix} \boldsymbol{c}_1^{(mxq)} & \mathbf{0}_1^{(mxq)} & \cdots & \mathbf{0}_1^{(mxq)} \\ \mathbf{0}_2^{(mxq)} & \boldsymbol{c}_2^{(mxq)} & \cdots & \mathbf{0}_2^{(mxq)} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_p^{(mxq)} & \mathbf{0}_p^{(mxq)} & \cdots & \boldsymbol{c}_p^{(mxq)} \end{bmatrix} \tag{9}$$

where each vector $\boldsymbol{c}_j^{(mxq)}$ consists of the modal load-to-strain ratios corresponding to the j^{th} strain gauge measurement response ψ_j^i; i.e,

$$\boldsymbol{c}_j^{(mxq)} = \left\{ \frac{r_1^1}{\psi_j^1} \quad \frac{r_1^2}{\psi_j^2} \quad \cdots \quad \frac{r_1^m}{\psi_j^m} \quad \frac{r_2^1}{\psi_j^1} \quad \cdots \quad \frac{r_2^m}{\psi_j^m} \quad \cdots \quad \frac{r_k^1}{\psi_j^1} \quad \cdots \quad \frac{r_k^m}{\psi_j^m} \right\}^T \tag{10}$$

Multiplication of the conversion matrix $\boldsymbol{C}^{(mxqxp)xp}$ by the vector of strain measurements $\bar{\boldsymbol{\varepsilon}}^p$ yields the corresponding vector of modal cross-sectional loads at each cross-section, and each strain gauge measurement, $\boldsymbol{\gamma}^{(mxqxp)}$; i.e,

$$\boldsymbol{C}^{(mxqxp)xp}\bar{\boldsymbol{\varepsilon}}^p = \boldsymbol{\gamma}^{(mxqxp)} \tag{11}$$

where, in this formulation, the overbar in $\bar{\boldsymbol{\varepsilon}}^p$ represents measurement data.

Note that the strain gauge measurement vector simply acts as a scalar amplifier of the modal loads used to define the coefficients in $\boldsymbol{C}^{(mxqxp)xp}$; i.e.,

$$\gamma_{i,j,k} = \frac{r_k^i}{\psi_j^i} \cdot \bar{\varepsilon}_j = \left(\frac{\bar{\varepsilon}_j}{\psi_j^i}\right) r_k^i \tag{12}$$

where $\gamma_{i,j,k}$ is one of the scalar elements of the vector $\boldsymbol{\gamma}^{(mxqxp)}$.

Also note that each vector $\boldsymbol{c}_j^{(mxq)}$ in $\boldsymbol{C}^{(mxqxp)xp}$ relates the j^{th} strain measurement to all modal cross-sectional loads. Thus, multiplication of the strain measurement vector by the entire conversion factor matrix yields the modal cross-sectional loads for all strain measurements, separately. As such, the right-hand side (RHS) vector $\boldsymbol{\gamma}^{(mxqxp)}$ can be defined as a vector consisting of p sub-vectors, each corresponding to one of the p strain gauge measurements; i.e,

$$\boldsymbol{\gamma}^{(mxqxp)} = \left\{ \boldsymbol{\gamma}_1^{(mxq)} \quad \boldsymbol{\gamma}_2^{(mxq)} \quad \cdots \quad \boldsymbol{\gamma}_p^{(mxq)} \right\}^T \tag{13}$$

where each sub-vector $\boldsymbol{\gamma}_j^{(mxq)}$ is defined as

$$\boldsymbol{\gamma}_j^{(mxq)} = \bar{\varepsilon}_j \left\{ \frac{r_1^1}{\psi_j^1} \quad \frac{r_1^2}{\psi_j^2} \quad \cdots \quad \frac{r_1^m}{\psi_j^m} \quad \frac{r_2^1}{\psi_j^1} \quad \cdots \quad \frac{r_2^m}{\psi_j^m} \quad \cdots \quad \frac{r_k^1}{\psi_j^1} \quad \cdots \quad \frac{r_k^m}{\psi_j^m} \right\}^T \tag{14}$$

Once the cross-sectional modal loads have been computed, the strain measurement vector can be defined via mode superposition as

$$\bar{\boldsymbol{\varepsilon}}^p = \boldsymbol{\Psi}^{pxm} \boldsymbol{\alpha}^m \equiv \sum_{i=1}^{m} \alpha_i \boldsymbol{\psi}_i^p \tag{15}$$

This becomes

$$\boldsymbol{C}^{(mxqxp)xp} \boldsymbol{\Psi}^{pxm} \boldsymbol{\alpha}^m = \boldsymbol{\gamma}^{(mxqxp)} \tag{16}$$

Multiplication of $\boldsymbol{C}^{(mxqxp)xp}$ and $\boldsymbol{\Psi}^{pxm}$ yields a new conversion matrix that relates the modal cross-sectional loads in $\boldsymbol{\gamma}^{(mxqxp)}$ to the modal amplitudes in $\boldsymbol{\alpha}^m$, not directly to the strain measurements $\bar{\boldsymbol{\varepsilon}}^p$; this new matrix will be defined as $\mathbf{X}^{(mxqxp)xm}$; i.e,

$$\mathbf{X}^{(mxqxp)xm} \boldsymbol{\alpha}^m = \boldsymbol{\gamma}^{(mxqxp)} \tag{17}$$

Solution of the overdetermined least-squares system yields the vector of modal amplitudes $\boldsymbol{\alpha}^m$, from which the resultant cross-sectional loads corresponding to the strain measurement data in $\bar{\boldsymbol{\varepsilon}}^p$ can be computed via mode superposition. However, this first requires the conversion matrix to be reassembled and redefined as

$$\mathbf{X}^{(mxq)xm} = \left[\boldsymbol{x}_1^{(mxq)} \quad \boldsymbol{x}_2^{(mxq)} \quad \cdots \quad \boldsymbol{x}_m^{(mxq)} \right] \tag{18}$$

where

$$\mathbf{X}^{(mxq)xm} = \boldsymbol{W}^{(mxq)x(mxq)} \boldsymbol{C}^{(mxq)xp} \boldsymbol{\Psi}^{pxm} \tag{19}$$

and

$$\boldsymbol{C}^{(mxq)xp} = \left[\boldsymbol{c}_1^{(mxq)} \quad \boldsymbol{c}_2^{(mxq)} \quad \cdots \quad \boldsymbol{c}_p^{(mxq)} \right] \tag{20}$$

Where $\boldsymbol{W}^{(mxq)x(mxq)}$ is a block diagonal weighting matrix used to scale the contributions of each vector $\boldsymbol{c}_j^{(mxq)}$ (corresponding to a given strain gauge) to the total solution and, now, since $\boldsymbol{\alpha}^m$ provides the best fit solution corresponding to all strain gauge measurements in $\bar{\boldsymbol{\varepsilon}}^p$, multiplication of $\mathbf{X}^{(mxq)xm}$ with $\boldsymbol{\alpha}^m$ will yield the vector of modal cross-sectional loads that best approximates all vectors $\boldsymbol{\gamma}_j^{(mxq)}$ i.e, the solution $\boldsymbol{\gamma}^{(mxq)}$ best approximates all p sub-vectors $\boldsymbol{\gamma}_j^{(mxq)}$ of $\boldsymbol{\gamma}^{(mxqxp)}$, $j = 1, \dots, p$. Thus,

$$\boldsymbol{\gamma}^{(mxq)} = \mathbf{X}^{(mxq)xm}\boldsymbol{\alpha}^m \equiv \sum_{i=1}^{m} \alpha_i \boldsymbol{x}_i^q \tag{21}$$

where $\boldsymbol{\gamma}^{(mxq)}$ is the modal cross-sectional load vector with each of its (mxq) terms equal to some weighted average of their corresponding p terms (per strain gauge) in $\boldsymbol{\gamma}^{(mxqxp)}$ (corresponding to the weights in $\boldsymbol{W}^{(mxq)x(mxq)}$). Note, however, that $\boldsymbol{\gamma}^{(mxq)}$ is still a vector containing modal components of each cross-sectional load; i.e,

$$\boldsymbol{\gamma}^{(mxq)} = \{\gamma_{1,1} \quad \gamma_{2,1} \quad \cdots \quad \gamma_{m,1} \quad \gamma_{1,2} \quad \cdots \quad \gamma_{m,2} \quad \cdots \quad \gamma_{1,q} \quad \cdots \quad \gamma_{m,q}\}^T \tag{22}$$

Where $\gamma_{i,k}$ is the weighted average modal cross-sectional load corresponding to the i^{th} mode and cross-section cut k. Thus, subsequent superposition of all modal components for each cross-section cut yields the best fit RHS force vector $\boldsymbol{f}^k$; i.e,

$$\boldsymbol{f}^k = \{f_1 \quad f_2 \quad \cdots \quad f_q\}^T \tag{23}$$

$$f_k = \sum_{i=1}^{m} \gamma_{i,k} \tag{24}$$

Additionally, the mode superposition definition can be used to compute the approximation of the measurement strain vector $\bar{\boldsymbol{\varepsilon}}^p$; the approximation will be defined as $\boldsymbol{\varepsilon}^p$ (no overbar); i.e,

$$\boldsymbol{\varepsilon}^p = \boldsymbol{\Psi}^{pxm}\boldsymbol{\alpha}^m \equiv \sum_{i=1}^{m} \alpha_i \boldsymbol{\psi}_i^p \tag{25}$$

The vector of errors between the strain measurement data and the reconstructed strains can then be computed,

$$\Delta\boldsymbol{\varepsilon}^p = \bar{\boldsymbol{\varepsilon}}^p - \boldsymbol{\varepsilon}^p \tag{26}$$

as well as the Euclidean norm of the error,

$$\mathrm{norm}(\Delta\boldsymbol{\varepsilon}^p, 2) = \|\bar{\boldsymbol{\varepsilon}}^p - \boldsymbol{\varepsilon}^p\|_2 \tag{27}$$

The key elements of this method are summarized below:

1. As opposed to the standard method of computing the conversion matrix as a product of modal bases, the components of the conversion matrix $\boldsymbol{C}^{(mxqxp)xp}$ are defined explicitly as ratios of cross-sectional load to elemental strains, extracted from the FE model of the structure.
2. Since the responses forming the generalized modal basis are global in nature, the strains and cross-sectional loads for every strain gauge-representative location j, and cross-section cut k, respectively, will likely be non-zero. Thus, the load-to-strain ratios are unique, non-zero scalar coefficients for every mode, strain, and cross-sectional load.
3. Since all DOFs of linear modes scale by a single amplitude, the products of the conversion matrix coefficients in $\boldsymbol{C}^{(mxqxp)xp}$ with the corresponding gauge strain measurements in $\bar{\boldsymbol{\varepsilon}}^p$ should, theoretically, yield $\boldsymbol{\gamma}^{(mxqxp)}$ with all sub-vectors $\boldsymbol{\gamma}_j^{(mxq)}$, $j = 1, \ldots, p$, being equal; i.e,

$$\boldsymbol{c}_j^{(mxq)}\bar{\boldsymbol{\varepsilon}}^p = \boldsymbol{\gamma}_j^{(mxq)} \tag{28}$$

$$\boldsymbol{\gamma}_j^{(mxq)} = \boldsymbol{\gamma}_l^{(mxq)}, \qquad j \neq l \tag{29}$$

   However, since the strain measurement data will contain some amount of error, the condition in equation 29 will likely not be the case, thus, creating the need for a best-fit solution to $\boldsymbol{\alpha}^m$.

4. By defining the conversion matrices $\boldsymbol{C}^{(mxqxp)xp}$ and $\mathbf{X}^{(mxqxp)xm}$ in the manner described above, the system is always overdetermined. As previously discussed, this is much more preferrable to an underdetermined system of equations (i.e, having more equations than unknowns as opposed to more unknowns than equations).

Equations 28 and 29 provide quantitative error metrics to determine whether the vectors of the generalized modal basis defining $\boldsymbol{\Psi}^{pxm}$ create a span sufficient to approximate the loads corresponding to the input strain data, or if additional basis vectors are required; again, this assumes that the structural response is linear.

9.2.5. Recommendations for the Formulation of Loads from Discrete Measurements

The following are the recommended follow-on work associated with approach B:

1. Compare the proposed methodology with current practices. Assess the applicability of the proposed method and the value added. The added complexity of solving for modal properties increases the amount of information needed to develop the model, as well as adds the layer of modal identification. The added time needed to complete the more complex and increased number of steps needs to be mapped against the value added. This will provide the information necessary to support the design-trade space studies for digital twin.
2. Approach B utilizes the fundamental mode shapes to develop the strain-to-load conversion matrices. The theories of structural dynamics suggest that an observed response in the time domain can be decomposed into modal components. The applicability of which is worth pursuing as a method to estimating the response along the length of the ship.

Thus, a viable path to explore going forward (to improve the estimation of loads by capitalizing on sparse, spatially distributed measurements) could include the use of the developed modes to identify the modal participations factors as a function of time. This enables the estimation of loads in uninstrumented areas to a higher level of accuracy.

An anticipated challenge is the prevalence of the forcing load profile in the response profile. The development of the future approach should consider this, as the structural vibration modes may be limited at systematically representing the structural response to the loading and vibrational behavior of the ship structure.

9.3. Approach C: Accounting for Uncertainties in Loads

9.3.1. Incorporating Uncertainties

The statistical assessment of the natural variability in primary loads extrema is an essential step in the development of an SDT. Fundamentally, the analysis of the primary loads assumes that they are stationary and ergodic random processes (amongst other statistical characteristics; see Rice (1944) for full details), their extrema (i.e., peaks) have been shown to be characterized with the Weibull distributions. In the select case where the load is a stationary, narrow-banded, Gaussian process, the extrema can be shown to be Rayleigh distributed, which is the equivalent of a Weibull distribution with a shape parameter of 2. For secondary loads, it has been shown that the Weibull distribution is also a good closed-form distribution to use to quantify the response (Lewis, 1989).

The Weibull distribution is a robust, closed-form distribution that is actually a family of distributions that can have an infinite number of shapes depending on the parameter β. The Weibull distribution contains both the Rayleigh distribution (β = 2) and the exponential distribution (β = 1). The shape parameter, β, is also referred to as the slope, which comes from the fact that the Weibull distribution can be transformed into a linearized form by taking the natural log of both sides twice. While the Weibull distribution is a useful distribution to quantify extrema, it is also a single-sided, unbounded distribution with non-zero probabilities for values of x approaching infinity. Thus, it does not account for any physical limitations on maxima.

The most generic form of the Weibull distribution is the three-term parameterization, expressed as:

$$\mathrm{F}_X(\mathrm{x}) = \begin{cases} 1 - e^{-\left(\frac{x - X_0}{\eta - X_0}\right)^{\beta}} & x \geq X_0 \\ 0 & x < X_0 \end{cases} \quad (30)$$

where,

$F_X(x)$ = cumulative density function (CDF) (i.e., the probability of non-exceedance of x)

x_o = threshold value of x (i.e., the value below which there is no measured data)

β = the Weibull shape parameter

η = the Weibull scale parameter

There are multiple methods for estimating the Weibull parameters (x_o, β, η) for a given dataset, including linear regression and moment-methods. Linear regression capitalizes on the linearized form of the Weibull distribution where independent and dependent variables can be defined and fit with a straight line. In doing so, parameters can be estimated with least square regression (a stable numerical solver), but the lower peaks are disproportionately weighted due to the transformation into the log-log space. Moment methods rely on solving the nonlinear equations for the central moments of the observed dataset and the assumed Weibull distribution. This approach weighs all events equally but, in some cases, is sensitive to numerical instabilities. A full description of the solution techniques is presented in past reports (Abernathy et. al., 1983; Lewis, 1989) including the two- and three- term linear regression methods (LR2 and LR3, respectively) and the two- and three- term moment methods (MM2 and MM3, respectively). The alternative fits are developed for the peaks, and then the method that has the best goodness of fit in the upper region is chosen to quantify the statistical distribution.

Considering guidance is concerned with the strength of the vessel being greater than the loads, the extreme values is a useful statistical quantity for SDT: The largest values for a sample of a finite size are also random variables. The distribution of the largest values, herein referred to as the extreme, are related to the distribution of their initial variate (or sample set) (Ang and Tang, 2007). The initial variate distribution defines the probability of the event for the tested time duration. However, the lifetime exposure is typically larger than that which was tested. As such, the number of instances, *N*, of the initial variate, *x*, that occurs over the lifetime can be used to define probability of the extreme. Assuming that the events are statistically independent and identically distributed, the cumulative distribution of the extreme value is:

$$F_Y(y) = F_X(y)^N \tag{31}$$

where *y* is the extreme lifetime value, and $F_Y(y)$ is the probability of non-exceedance of the extreme value. Given the CDF of the initial variate $F_X(x)$ is defined by the Weibull CDF, this can be rewritten as

$$F_Y(y) = [1 - e^{-\left(\frac{y - x_o}{\eta - x_o}\right)^{\beta}}]^N \tag{32}$$

The value of the extreme value associated with a specific probability of non-exceedance (PNE) can be expressed as

$$y = (\eta - x_o)\left[-\ln\left(1 - PNE^{1/N}\right)\right]^{1/\beta} + x_o \tag{33}$$

where PNE is the probability of non-exceedance.

This fundamental approach to statistically quantifying extreme events can be adapted to support near-real time assessments for SDT. The idea being, if there is data being measured by the system and shown in Figure 27 (top), the extrema can be identified (bottom). The extrema can then be pulled for a given time window, t_{window}, and statistically fit, as shown in Figure 28.

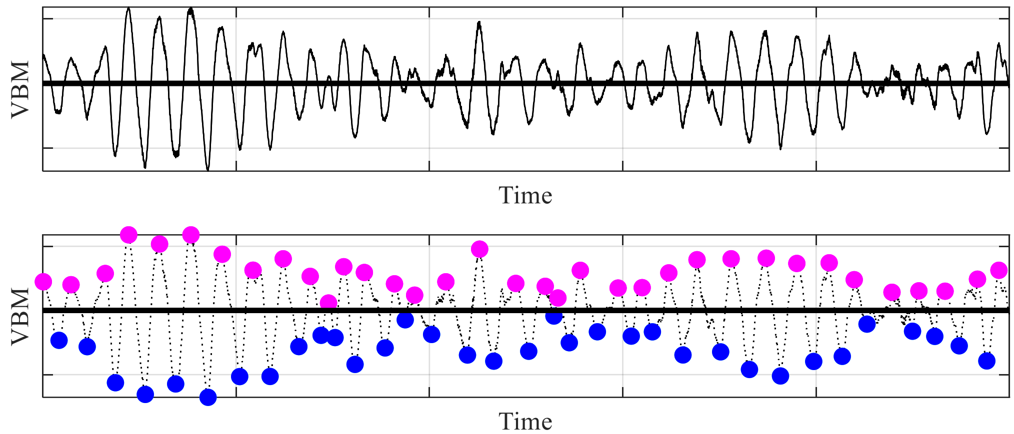


Figure 27. Top: Sample vertical bending moment time history; bottom: hog (magenta) and sag (blue) peaks identified in the top time history.

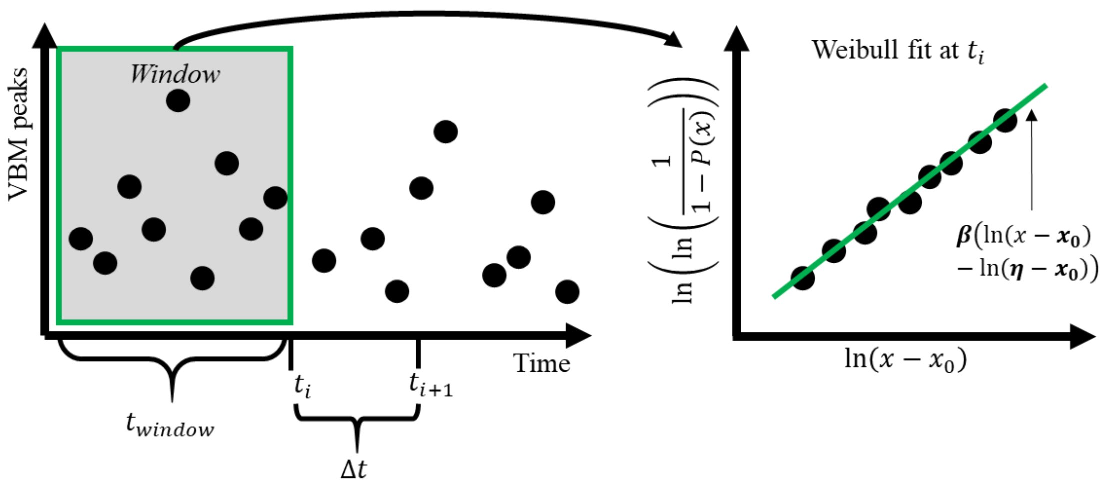


Figure 28. Illustration of sliding Weibull fit analysis

Once the Weibull fit parameters are computed for a sample window, it is possible to use them to produce a forecasted extreme load $x_{forecast}$. The forecasted value depends on the PNE and the forecast window $t_{forecast}$, which is the duration into the future to extrapolate. Equations (34) and (35) were used to extrapolate the future extreme load at the i^{th} time step, using fitted distribution parameters β_i, η_i, and $x_{0,i}$, and the number of events in the sample window $n_{events,i}$.

$$x_{forecast,i} = \left(-\ln\left(1 - PNE^{\frac{1}{N_i}}\right)\right)^{\frac{1}{\beta_i}} \left(\eta_i - x_{0,i}\right) + x_{0,i} \tag{34}$$

$$N_i = t_{forecast} * \frac{n_{events,i}}{t_{window}} \tag{35}$$

This method was implemented on a data set (see next section) and evaluated. To evaluate the performance of the predictive approach, the forecasted value was assessed with respect to the actual observed value within the forecast window. For the forecast window, the peak time history data $x_{data}(t)$ was used to extract the actual future maximum load for each Weibull fit (see equation (36)). With this value, one could also obtain a relative forecast deviation δ (see equation (37)).

$$x_{actual,i} = \max(x_{data}(t))\ where\ t_i \leq t \leq t_i + t_{forecast} \tag{36}$$

$$\delta_i = \frac{x_{forecast,i}}{x_{actual,i}} - 1 \tag{37}$$

By plotting these derived quantities and then updating them for different selected values of PNE and forecast window, one could assess the effect of these parameters on the accuracy of forecasted data.

9.3.2. Application

The dataset used in this application is a dataset of continuously recorded strain gage measurements from various, discrete locations on a notional naval vessel. The strain gage data was coupled with the physics-based model calibrations to infer vertical bending, as described in Section 9.

The dataset was chosen to include severe structural responses that induced slamming and whipping on the vessel. The data pertains to the vertical bending response at midship. The vertical bending response data represented only the ordinary wave response with whipping and did not include the still-water bending moment. A peak-finding algorithm was applied to the vertical bending signal to isolate the hog and sag extrema for each wave encounter. In this way, the observed, continuous, data (i.e., random process) was distilled into random variables (hog and sag extrema). This example is representative of content in the ship's response. The physical response, as described, includes multiple physical responses. These different events may be better classified separately and extrapolated for the extremes. However, this would require an understanding of dependencies and/or joint occurrences of the events and the ability to integrate them into statistic of the extremes. This is recommended for further investigation, but falls outside of this study.

The example dataset comprised 17 continuous days' worth of measurements where the vessel is at sea. The initial study is formulated around the most critical day in the dataset (section 9.3.3). This deep dive isolates and identifies key features of the approach in order to highlight critical design parameters for Approach C (the forecasting approach). The expanded study pushes the evaluation to the full data set (section 9.3.4. Alternative approaches are then presented and explored in sections 9.3.5 and 9.3.7.

9.3.3. Initial Study

The following subsections summarize the evaluation of Approach C with respect to the application. Multiple sensitivity studies were performed to investigate the performance of the forecasting approach with respect to its parameters, including sample size, probability of non-exceedance, forecast window, and non-stationary conditions.

9.3.3.1. Sample Size

The forecasting approach (Approach C) was applied to the data set and evaluated for 4-time windows t_{window} = 300, 600, 1200, and 2400 seconds. The approach was applied to the entire data set. The review of the complete dataset identified consistent concepts as discussed below.

Figure 29 shows the time histories of the three Weibull parameters, η, β, and x_0 as computed using all four sample window durations (300, 600, 1200, and 2400 seconds).

From it, one can conclude that (1) as sample duration increases, the parameter estimates become smoother, and (2) the graphs for longer sample windows lag those for smaller sample windows. The first observation is reinforced by the comparison of standard deviations, presented in Table 17 (data has been normalized by the values in the first column). This table shows that standard deviation of each parameter decreases as the sample window size increases.

Neither observation is unexpected: The graphs for longer sample windows are smoother because the parameter estimates are based on larger sample sizes, whose statistical characteristics change more slowly between time steps. And by including peaks from further in the past, the Weibull fit for a larger sample window will naturally tend to reflect the data from further in the past, hence explaining the time lag.

Table 18 shows the effect of altering the window duration on the forecast deviation δ statistics for a fixed PNE and forecast window. One observes that the mean, maximum, and minimum δ tend to shift in the positive direction with increasing window size. However, when focused on a smaller portion of the day's data, this observation no longer holds true: There are hours when the mean δ decreases with window size. This tends to occur when there is a significant positive gradient in the η parameter, as occurs roughly halfway through the day.

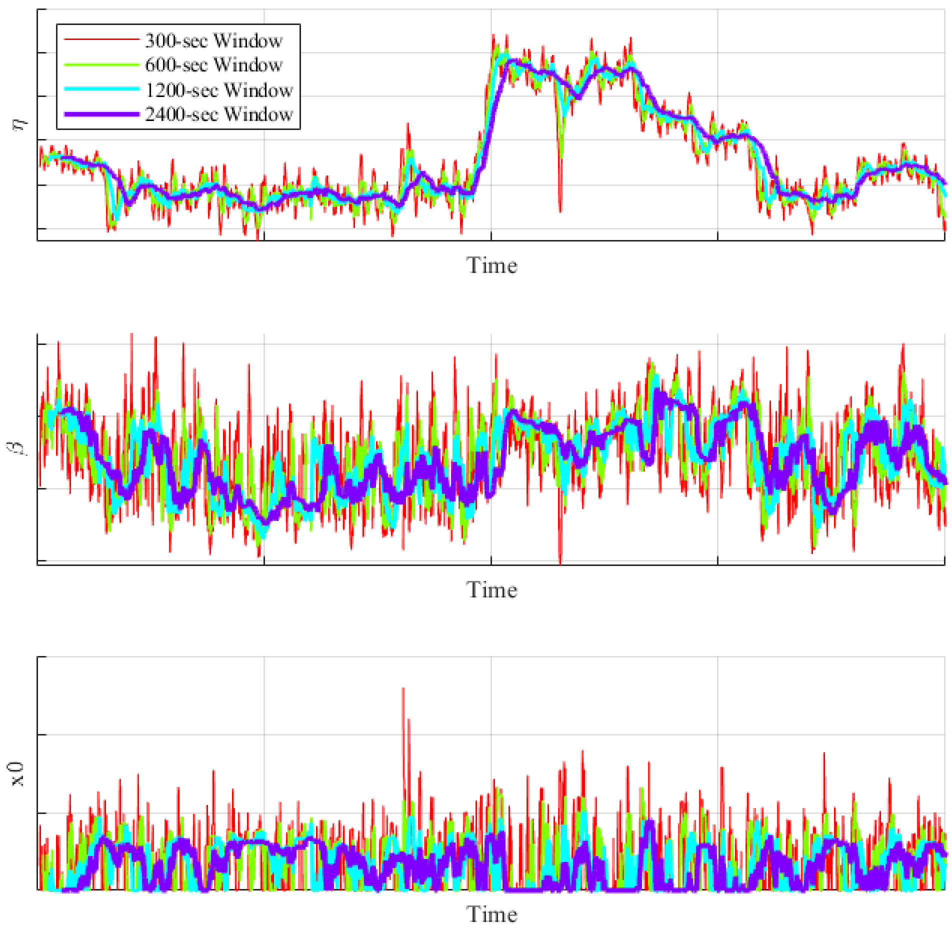


Figure 29. Weibull Parameter Estimate Time Histories.

Table 17. Normalized Standard Deviations of Weibull Parameter Estimates.

	t_{window} [seconds]			
	300	600	1200	2400
$\boldsymbol{\eta}$	1.000	0.978	0.965	0.957
$\boldsymbol{\beta}$	1.000	0.875	0.812	0.757
$\boldsymbol{x_0}$	1.000	0.811	0.664	0.572

Table 18. Deviation statistics vs. sample window size for PNE=0.9 and forecast window=600 seconds

Deviation Statistics	t_{window} [seconds]			
	300	600	1200	2400
$\boldsymbol{\mu}$	0.157	0.164	0.174	0.185
$\boldsymbol{\sigma}$	0.237	0.228	0.217	0.211
$\boldsymbol{median}$	0.131	0.137	0.137	0.145
$\boldsymbol{max}$	6.142	7.074	6.882	6.612
$\boldsymbol{min}$	-0.455	-0.409	-0.309	-0.238

9.3.3.2. Probability of Non-Exceedance (PNE)

The forecasting approach (Approach C) was applied to the data set and evaluated considering a range of PNE. The effect of PNE on forecast deviation can be examined analytically by taking the partial derivative of (37), as shown in (38). This equation includes the partial derivative of (34) with respect to PNE, which is provided below in equation (39). As shown in the equation, this partial derivative is proportional to $(\eta - x_0)$. Its dependence on β and N is less obvious. Figure 30 plots this relation as a function of PNE for different values of β. The plot shows that the partial derivative tends towards infinity as the PNE approaches 1. The growth rate decreases as β increases.

$$\frac{\partial \delta}{\partial PNE} = \frac{1}{x_{actual}} \frac{\partial x_{forecast}}{\partial PNE} \tag{38}$$

$$\frac{\partial x_{forecast}}{\partial PNE} = \frac{(\eta - x_0)}{\beta} \left(-\ln\left(1 - PNE^{\frac{1}{N}}\right)\right)^{\frac{1}{\beta}-1} \frac{PNE^{\left(\frac{1}{N}-1\right)}}{N\left(1 - PNE^{\frac{1}{N}}\right)} \tag{39}$$

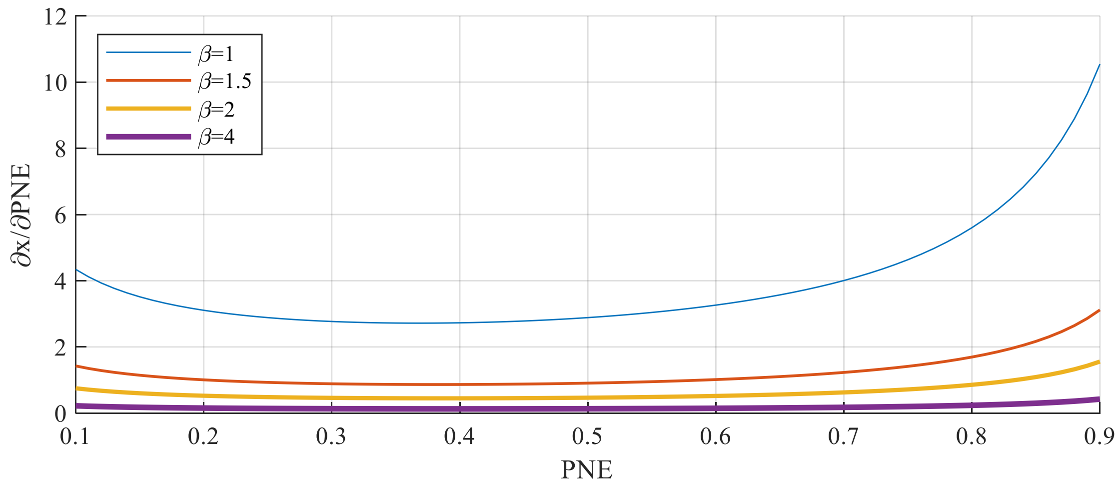


Figure 30. $\partial x/\partial PNE$ vs. PNE for various β for $\eta - x_0 = 0$ and $N = 100$

Based on these equations and Figure 30, one can deduce the following general statements:

- As PNE increases, the forecast deviation δ will increase.
- The increase will be greater for larger $\eta - x_0$
- The increase will be greater for smaller β
- The increase will be greater for smaller x_{actual}
- The forecast deviation will grow without bound as PNE approaches 1.

As further confirmation of some of these statements, Figure 31 shows the change in forecast deviation against time when the PNE is increased from 0.368 to 0.9. The plot shows that the increase is always positive, as expected. Moreover, the change is not uniform, indicating it must depend on other time-varying quantities, namely the distribution fit parameters.

Put simply, the PNE gives control over the level of conservatism in the forecasted extreme value. By setting the PNE close to 1.0, one can generally ensure the forecast deviation will be positive, i.e., one will almost certainly over predict the actual future extreme value.

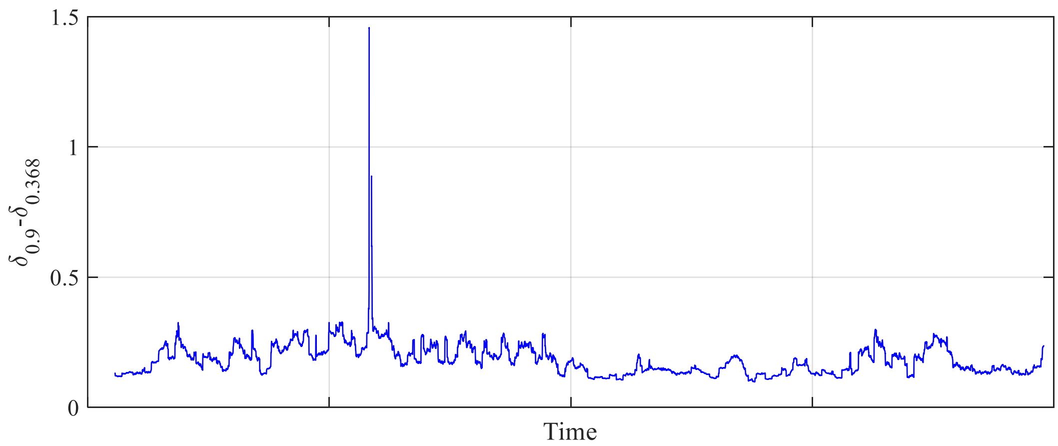


Figure 31. Change in forecast deviation when increasing PNE from 0.368 to 0.9 (for 2400-second sample window and 600 second forecast window)

9.3.3.3. Forecast Window

The forecast window $t_{forecast}$ affects the forecast deviation from two aspects. First, it defines the expected number of future events N through equation (35). To be precise, N is proportional to $t_{forecast}$. According to equation (34), the value of the extrapolated extreme event $x_{forecast}$ increases monotonically with N. Thus, an increase in $t_{forecast}$ will cause an increase in $x_{forecast}$. The second place where $t_{forecast}$ enters the forecast deviation is through equation (36), which defines the actual future extreme event x_{actual}. As $t_{forecast}$ increases, x_{actual} can either increase or stay the same.

Since both x_{actual} and $x_{forecast}$ might increase at different rates with larger $t_{forecast}$, one cannot make any more general statement about δ other than it could change value in either direction. One may instead look at the δ-statistics for different values of $t_{forecast}$, as are reported in Table 19. This table indicates that the mean, median, and standard deviation of δ decrease for larger $t_{forecast}$, i.e., the deviation becomes both smoother and smaller. The increased smoothness can also be discerned from Figure 32: The curve corresponding to the 3600 second forecast window is much more compressed and closer to 0 than the other curves. Unfortunately, it has one section of negative values around midday, corresponding to the section of increasing η in Figure 29. The other deviations for the other forecast windows still tend to overpredict in this interval, indicating temporarily better performance.

In brief, the sample dataset suggests that increasing the forecast window time up to 3600 seconds leads to less volatile forecast deviations. Indeed, the relatively limited range in deviations for the 3600 second forecast window on its own indicates that the forecasting methodology produced generally favorable results on the sample data set. However, during times of rapid change in distribution parameters, the forecast will tend to either under- or overpredict the actual future extreme event by a significant margin. In these situations, an intermediate forecast window such as 600 seconds provides slightly more reliable predictions.

Table 19. Deviation statistics for different forecast windows

Deviation Statistics	$t_{forecast}$ [seconds]			
	60	300	600	3600
μ	0.374	0.221	0.185	0.126
σ	0.396	0.233	0.211	0.153
Median	0.278	0.185	0.145	0.129
max	5.116	6.192	6.612	0.553
min	-0.357	-0.271	-0.238	-0.333
Note: PNE=0.9 and t_{window}=2400s				

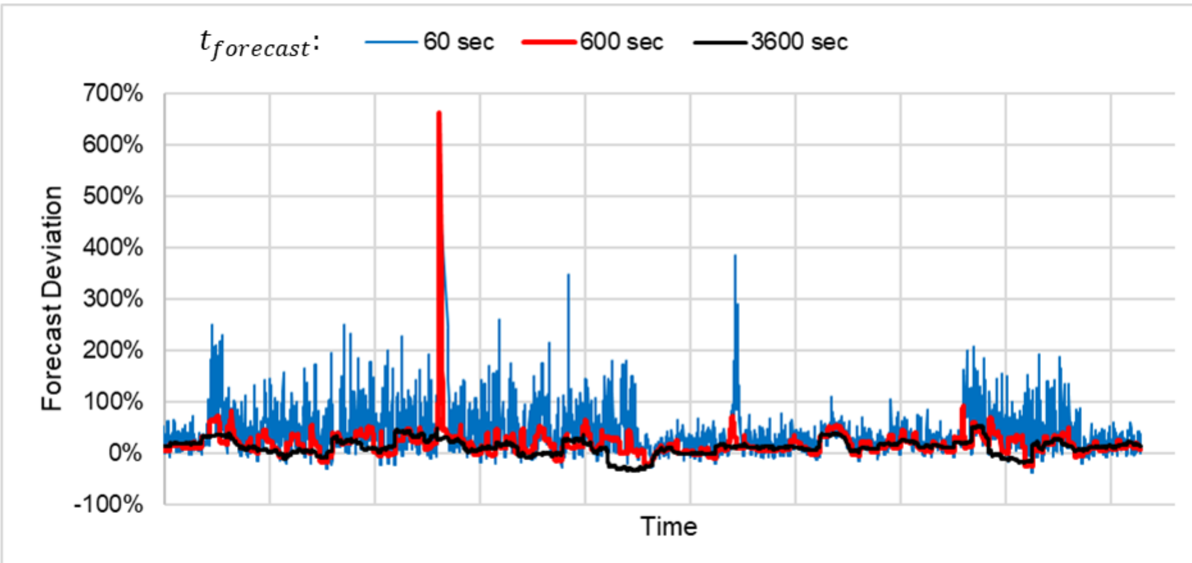


Figure 32. Forecast deviation vs. time for different forecast windows (PNE=0.9, t_{window}=2400s)

9.3.3.4. Non-Stationary Conditions

The previous section hinted at a challenge intrinsic in the adopted forecasting approach, namely the limitations of assuming that the peak distribution is stationary. It was already evident from the previously discussed distribution parameter time histories that the peak statistics can hardly be considered stationary from hour to hour. To illustrate the effect of non-stationary conditions on forecasting accuracy, consider the plots in Figure 33. Each plot shows the performance of a forecast based on different sample and forecast window combinations. Each forecast applies to the same interval in the sample data, namely the interval enclosing the rapid rise in η around midday, as can also be seen in Figure 29.

The top-most plot shows the forecast for the largest sample window and the longest examined forecast window of 1 hour. There is a wide interval in the middle where the actual future extreme event is larger than the forecasted value. This interval of underestimation begins well before η even begins to increase. This happens because the forecast window is so large that it includes extreme events from the time when η increases at an earlier point in time. Contrast this plot with the one below it, which uses a forecast time of only 300 seconds. The underestimate zone there is significantly shorter and occurs later in time. Thus, one might conclude that during non-stationary conditions, the forecast performance may be improved by decreasing the forecast window. Finally, the bottom plot shows the forecast when the sample window is also decreased to the lowest value of 300 seconds. The forecast becomes much more erratic, and an underestimate zone still exists, though it is the shortest of all the plots. The plot therefore suggests that little is gained by using a smaller sample size to estimate Weibull parameters in this case.

This initial exploration of forecasting during a non-stationary condition indicated that performance could be improved by reducing the size of the forecast window. A possible solution to the stationarity problem would be to find an optimum forecast window, or to adaptively change the forecast window based on how much the peak statistics appear to be changing. Beyond this, improvements might be obtained by extrapolating the fitted distribution parameters based on recent behavior and using the extrapolated values to predict future extreme events.

When explored under the context of operating conditions, neither the significant wave height nor the wave period changed much during the day (top plot of Figure 34). By contrast, one could immediately identify a correlation between the rise in η around midday and a change in heading from head seas to stern quartering (middle plot of Figure 34). This change in heading additionally corresponded with a change in mean speed (bottom plot of Figure 34). Thus, one may conclude that the increase in load peak magnitudes was principally due to the change in heading and speed.

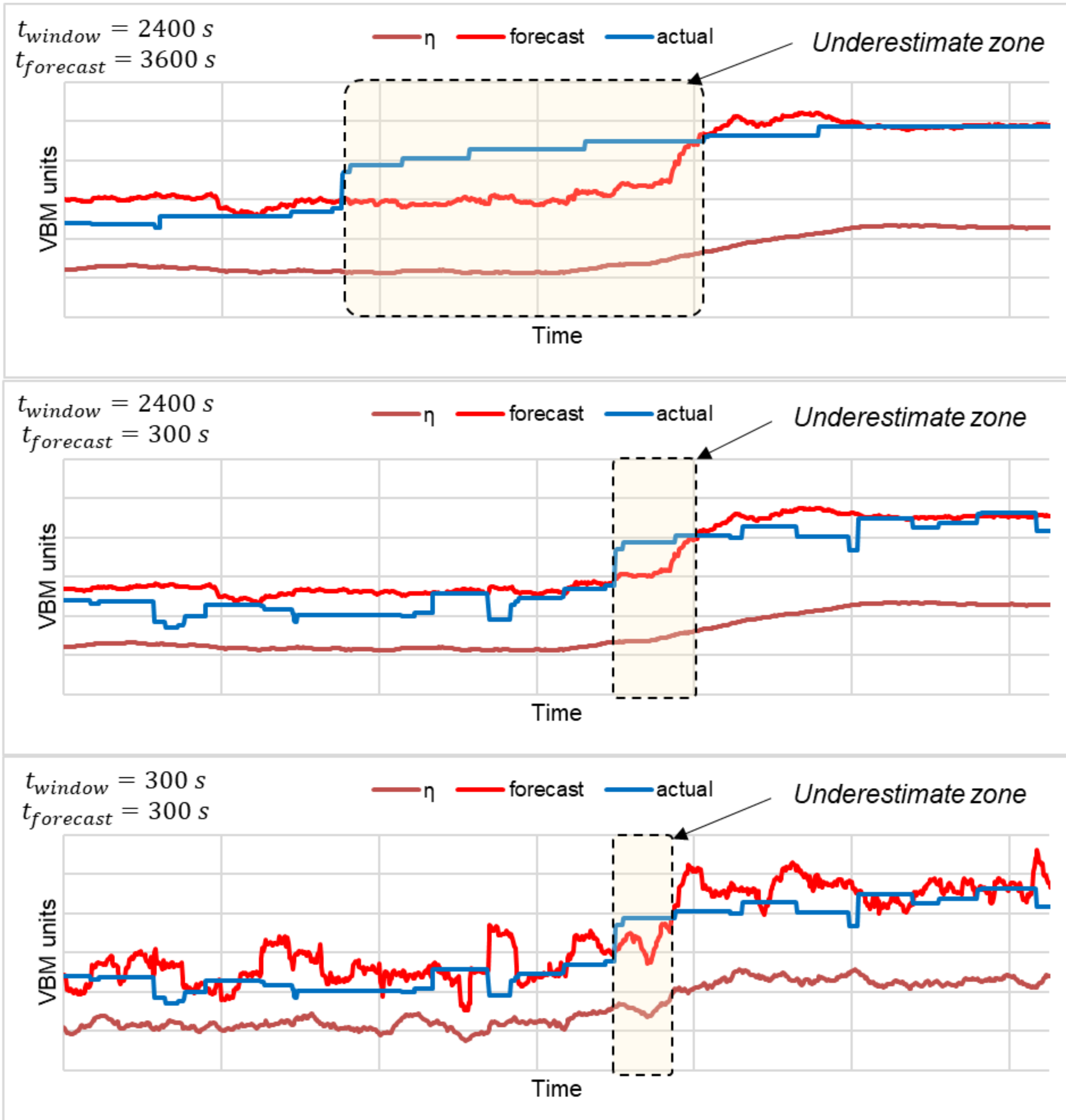


Figure 33. Performance of three forecasts during a non-stationary interval (PNE=0.9)

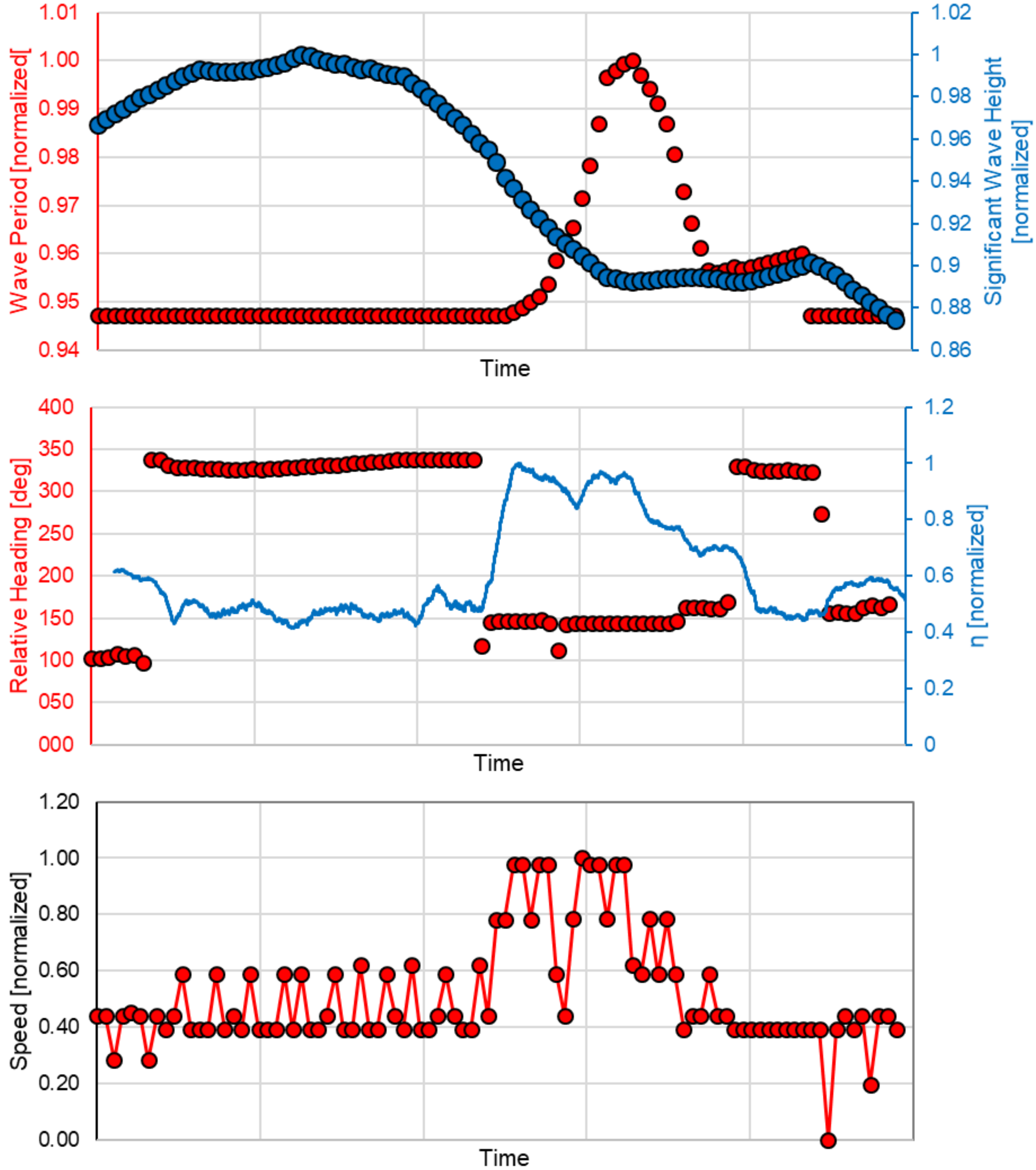


Figure 34. Wave Period, Significant Wave Heigh, Relative Heading, Characteristic Value of Response, and Speed. (Note: the time axis for all figures is the same)

9.3.4. Expanded Study

The initial exploration of the dataset was expanded to incorporate the full 17 days' worth of data. The same analysis was performed for all 17 days as was applied to the 1 day in the Initial Study (section 9.3.3). The related plots and figures did not exhibit different tendencies that those presented for the critical day of data and therefore are omitted for brevity. The review of the 17 days' worth of data does, however, show that the observations made based on the initial exploratory study's dataset applied generally:

1. As the sample window increased, the Weibull fit time histories became smoother, but time delayed.
2. As the sample window increased for the same forecast horizon, the forecast deviation time histories became smoother.
3. As the forecast horizon increased for the same Weibull fit sample window, the forecast deviation time histories became smoother and shifted downward and exhibited instances of negative plateauing that correlated with periods of increasing Weibull scale parameter, indicating that peak values were increasing in time, i.e., non-stationary. For all days but one, the median and mean of the forecast deviation were positive, i.e., the forecast over predicted the actual future extremum.

9.3.5. Alternative Forecasting Approach 1: Revised Accuracy

The expanded study underlined the weakness of the extrapolation-based forecasting method in non-stationary conditions. Therefore, the next phase of research focused on discovering and evaluating methods to improve the forecast accuracy. This section describes some of the methods tested for this purpose.

9.3.5.1. Scale Parameter Forecasting

The first alternative at improving forecasting accuracy was to extrapolate future extreme loads using a Weibull scale parameter that would itself be a forecast based on past scale parameter values. This approach was based on the observation that instances of underprediction were correlated with periods when the scale parameter was increasing. The scale parameter is an indirect measurement of the magnitude of peak values: an increasing scale parameter is indicative of worsening conditions. Consequently, the extrapolation tended to underpredict the future extreme because it used a scale parameter based on the less severe past conditions. It was hypothesized that a more accurate extrapolation might be obtained using an estimate of the future, larger scale parameter.

To test the hypothesis, the expanded study was repeated, but with a small change: Instead of using the scale parameter at the current time step to extrapolate a future extreme value, the maximum scale parameter within the forecast horizon was used for that purpose. In a real-time environment, such a calculation would be impossible, since the maximum scale parameter in the forecast horizon would not yet have been observed; it could, however, be estimated in some fashion. Using the actual maximum scale parameter was an idealization meant to test whether extreme load forecasting accuracy might be improved given perfect knowledge of the future values of just the scale parameter.

$$\eta_{perfect,i} = \max\big(\eta(t)\big)\, for\ t_i \leq t \leq t_i + t_{forecast} \tag{40}$$

$$x_{\eta,perfect,i} = \left(-\ln\left(1 - PNE^{\frac{1}{N_i}}\right)\right)^{\frac{1}{\beta_i}}\left(\eta_{perfect,i} - x_{0,i}\right) + x_{0,i} \tag{41}$$

As expected, the forecasting study using perfectly predicted maximum scale parameters resulted in slightly more conservative forecasted extreme loads. The change in the minimum forecast deviation for each day’s data was used as a primary metric for assessing the improvement in forecasting accuracy. Without exception, the minimum forecast deviation was shifted in the positive direction relative to the original expanded study that used the causal value of scale parameter. The degree of change in minimum forecast deviation seemed uncorrelated to the sample window size used for Weibull fitting. However, after breaking down the metric according to forecast window, it was possible to see that the minimum value tended to increase for forecast windows of 600 and 3600 seconds (see Figure 35); samples for forecast windows of 60 seconds were mostly unaffected. This trend is to be expected, since the longer the forecast window, the more time there is for the scale parameter to increase to a more severe value. After excluding all samples for a forecast window of 60 seconds, the average minimum forecast deviation of the remainder was roughly -0.26, compared to an average minimum value of -0.39 for the original expanded study.

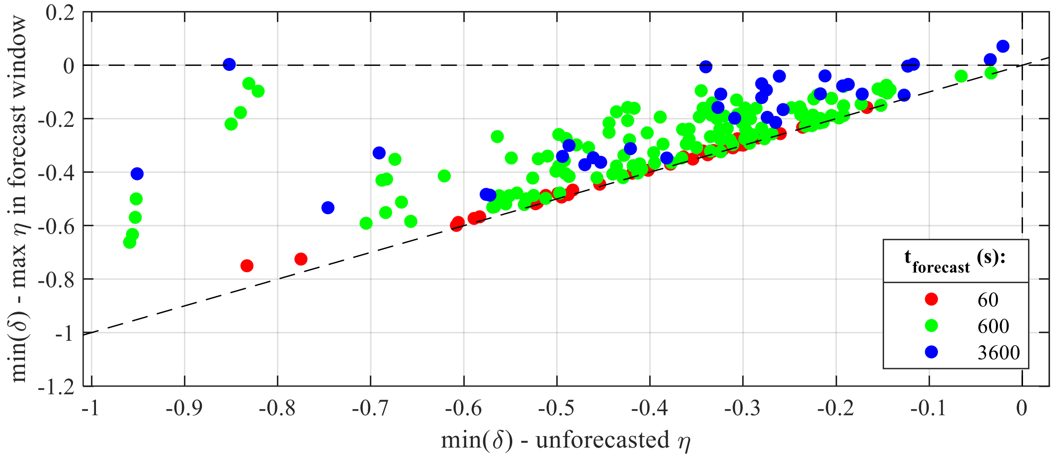


Figure 35: Minimum daily forecast deviation using maximum η in forecast horizon vs. minimum daily forecast deviation using present time η

Figure 36 through Figure 39 offer a more detailed view of the effect of η-forecasting. These figures show the extreme load forecasting results for the expanded dataset using a forecast horizon of 60 minutes and a Weibull fitting window of 40 minutes. They compare the results without η-forecasting against results using $\eta_{perfect}$ and a naïve η-forecasting method. The scale parameter for this method is denoted η_{naive} and is computed as:

$$\eta_{naive,i} = \begin{cases} \eta_i, \; if \; \eta_i < \eta_{i-1} \\ \eta_i + (\eta_i - \eta_{i-1}) * \frac{t_{forecast}}{t_i - t_{i-1}}, \; otherwise \end{cases} \tag{42}$$

$$x_{\eta,naive,i} = \left(-\ln\left(1 - PNE^{\frac{1}{N_i}}\right)\right)^{\frac{1}{\beta_i}} (\eta_{naive,i} - x_{0,i}) + x_{0,i} \tag{43}$$

In effect, η_{naive} assumes the trend in scale parameter values that held between the current and previous time steps will hold for the entire forecast window, unless that trend is downward. In that case, the current scale parameter is used. Such a forecasting method could be implemented in real time, as it makes use only of current and past observations. To reduce the variability in η_{naive}, the original dataset, which provided Weibull estimates every 10 seconds, was downsampled to 1/5 minutes for this study.

Figure 36 shows a time history excerpt of forecasted and actual extreme loads. It shows that of the three forecasting methods, using the current scale parameter value leads to the greatest time lag in forecasted extreme load compared to the actual curve. The model using perfect knowledge of the future scale parameter leads to the smallest time delay, but the delay persists. The naïve scale parameter forecasting methods leads to less of a time delay than using the current value, but also results in more erratic predictions, leading in some cases to significant overestimation.

Figure 37 plots the forecast deviation against the actual future extreme load for all points in time. The scatter shows that underestimation occurs throughout the range of loadings, i.e., it is not concentrated in specific bands of extreme loads. Moreover, it reinforces the observation that underestimation occurs for all forecasting methods, even the one using a perfect model of future scale parameters.

Figure 38 provides a visual to compare the results using the two scale-parameter forecasting methods against the baseline method. It plots the forecast deviation of either method against the forecast deviation for the baseline method using current scale parameter values. Note that the plots have been cropped to exclude some of the outlier values visible on Figure 37. By comparing the plots, one may observe that the η_{naive} method tends to cause more significant overestimation of extreme loads than the $\eta_{perfect}$ method. Nevertheless, many of the worst underestimations remain uncorrected using η_{naive}. $\eta_{perfect}$ tends to ameliorate the forecast deviation, at least somewhat, in all cases of underestimation.

Finally, Figure 39 shows the distribution in forecast deviation for all three forecasting methods. The mean and standard deviation of the distributions indicate that η_{naive} shifts the distribution the furthest into the positive direction, but also increases the spread. By contrast, $\eta_{perfect}$ increases the deviations while also reducing the spread.

Overall, this study demonstrated that extrapolating extreme loads using a forecasted value of the Weibull scale parameter could improve extreme value forecasting accuracy. This was the case even using a naïve forecasting method for the scale parameter. However, the results showed that the forecast deviation could still be on the order of -0.6, even given perfect knowledge of the maximum future scale parameter. Thus, while using even a simple model to predict future scale parameters might ensure more conservative extreme load forecasts, this approach still does not eliminate the risk of underprediction.

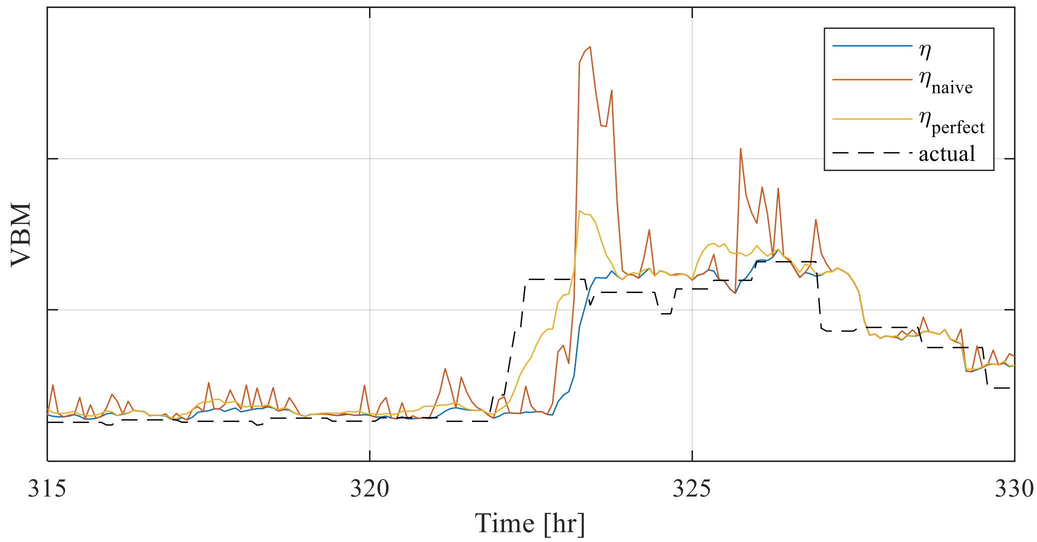


Figure 36. Time history excerpt of forecasted extreme loads using various η-forecasting methods.

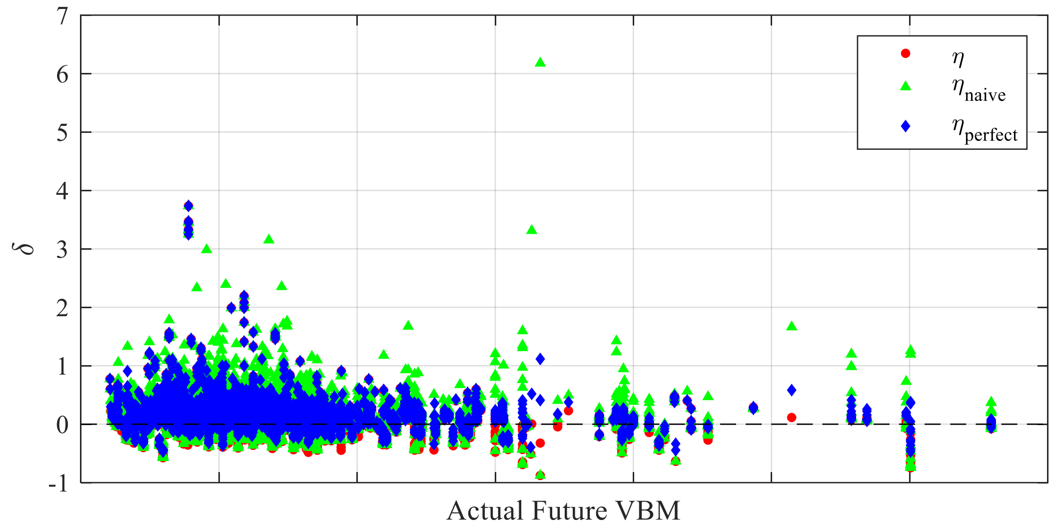


Figure 37. Scatter plot of forecast deviations vs. actual future extreme load

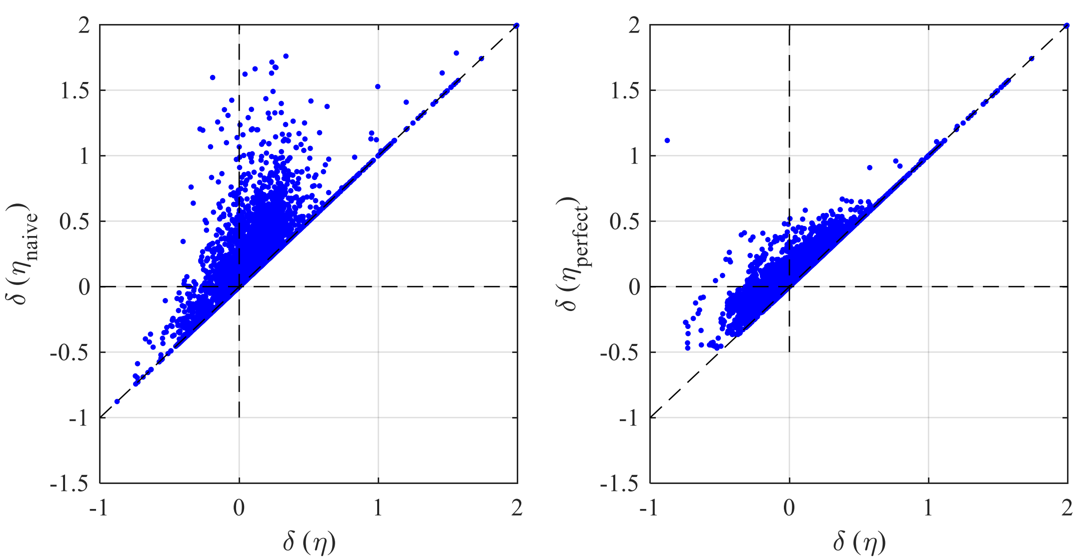


Figure 38. Comparison of forecast deviations of different η-forecasting methods.

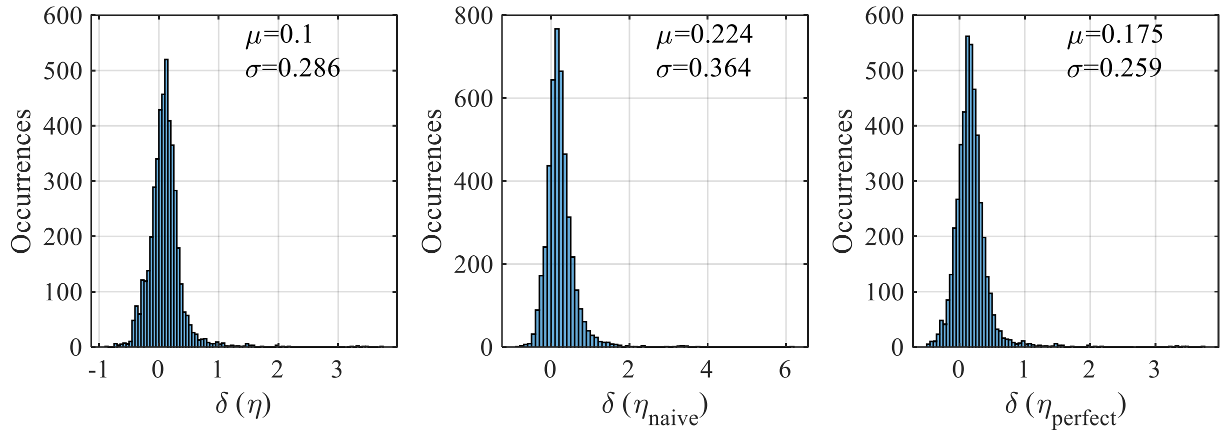


Figure 39. Forecast deviation distribution for different η-forecasting methods.

9.3.6. Extrapolation Forecasting

In addition to attempting to improve forecast accuracy by forecasting the scale parameter time history, a study was conducted to explore the effect of forecasting the extrapolated load time history itself. For an initial benchmarking, it was assumed a perfect model of this time history was available. Thus, a perfect forecasted time history could be generated for the dataset:

$$x_{perfect,i} = \max\left(x_{forecast}(t)\right) for\ t_i \leq t \leq t_i + t_{forecast} \tag{44}$$

The forecast deviations were computed for this new extreme load forecast, and the results were compared against those obtained using perfect knowledge of the future scale parameter. Again, a forecast horizon of 60 minutes and Weibull fitting window of 40 minutes were selected. The original 1/10s frequency of fit data was down sampled to 1/5 minutes. Figure 40 depicts an excerpt of the extreme load forecast time history. It illustrates that the forecast obtained by eliminating the time delay in the original extrapolations better envelops the actual extreme future extreme load curve than does the forecast using perfect scale parameter knowledge. This result suggests that if it were possible to eliminate the time delay from the original extrapolation, then the forecast deviations could be effectively minimized. Moreover, it indicated that forecasting the extreme load time history would be more effective at increasing accuracy than forecasting the scale parameter and recomputing extrapolations. This conclusion is further supported in Figure 41, which shows the forecast deviations associated with the $vbm_{perfect}$ result is concentrated at values above 0. In particular, the deviation for the largest values of future extreme load is all close to 0.

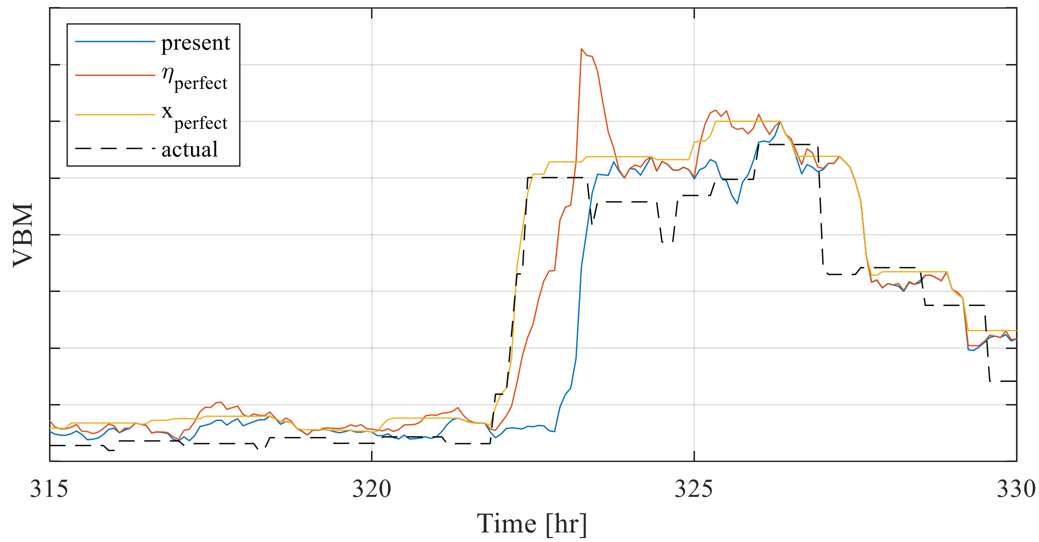


Figure 40. Extreme Value Forecast Time History Excerpt for Perfect Models

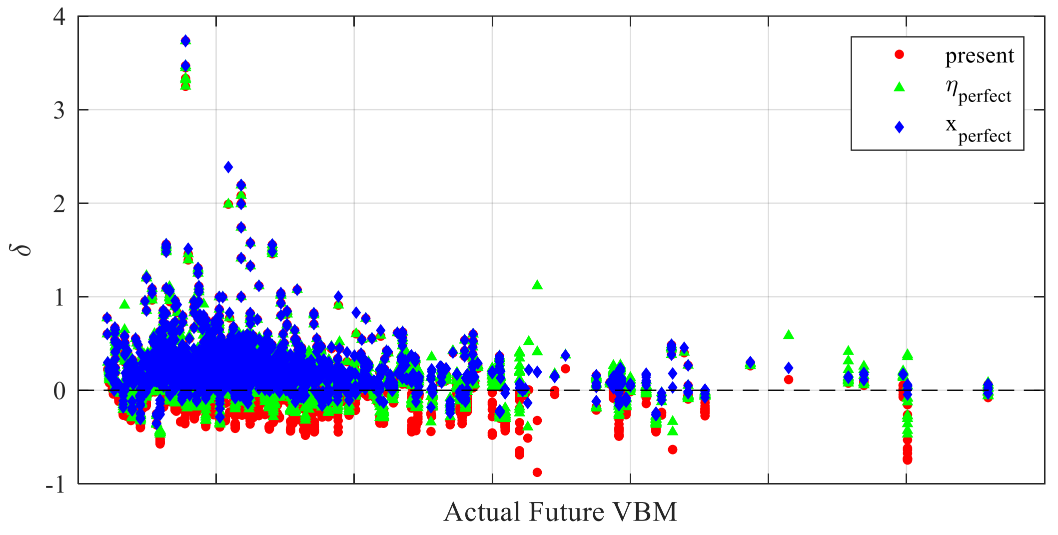


Figure 41. Forecast Deviations vs. Future Load for Perfect Models

As with the scale parameter forecasting, a naïve load forecast was computed using the current (x_i) and last previous (x_{i-1}) extrapolated loads:

$$x_{naive,i} = \begin{cases} x_i,\ if\ x_i < x_{i-1} \\ x_i + (x_i - x_{i-1}) * \frac{t_{forecast}}{t_i - t_{i-1}},\ otherwise \end{cases} \tag{45}$$

The result was assessed visually. Figure 42 shows the time history excerpt, which illustrates the volatile response of this naïve forecast. While the forecast typically overestimates the actual future load, and slightly reduces the time delay, it still does not eliminate it for rapidly changing conditions. As further demonstrated in Figure 43, the naïve forecast has no effect on many of the worst underestimations.

For completeness, Table 20 documents statistical measures of the forecast deviation for the various attempts to improve forecasting accuracy.

Overall, the investigations herein suggest that there is an extreme deficiency with using time series models to forecast the extreme load time history with sufficient accuracy, particularly by using Weibull fit data alone. It is possible that by fusing the data with other environmental and weather information, instances of underestimation during worsening conditions may be mitigated. Regardless, the extrapolation forecasting study did suggest that the extrapolated load time history can provide a conservative envelope to the actual future experienced extreme loads once the time lag in the data is removed. This result thus provides further supporting evidence to the validity of the general methodology.

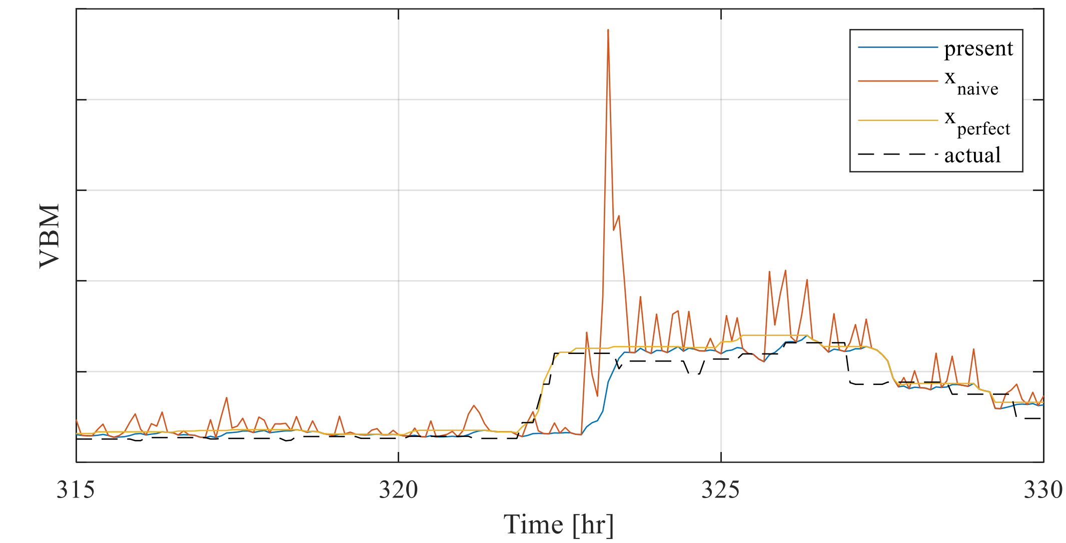


Figure 42. Extreme Value Forecast Time History Excerpt for Naïve Forecast Model

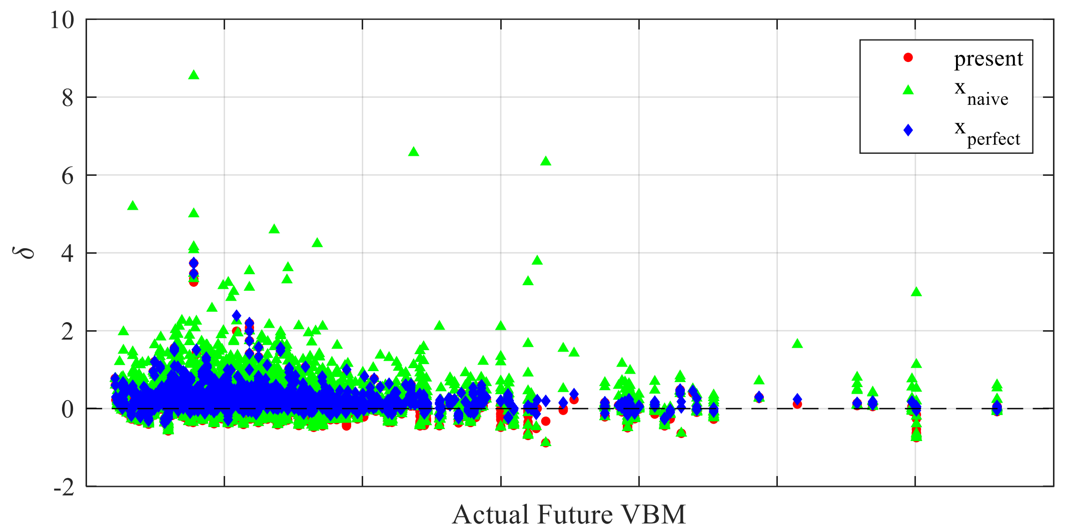


Figure 43. Forecast Deviations vs. Future Load for Naïve Forecast Models

Table 20. Forecast Deviation Statistics for Forecasting Attempts

Forecast Method	Mean	Median	Min	Max	Std. Dev.
present	0.100	0.093	-0.878	3.735	0.286
$\eta_{perfect}$	0.175	0.156	-0.470	3.735	0.259
η_{naive}	0.224	0.178	-0.878	6.176	0.364
$x_{perfect}$	0.208	0.187	-0.356	3.735	0.242
x_{naive}	0.289	0.198	-0.878	8.543	0.487

9.3.7. *Alternative Forecasting Approach 2: Under/Overestimate Classification*

The second approach for improving forecasting accuracy draws upon classification modeling. It was hypothesized that a model could be trained to classify a forecast as either an over- or underprediction, using past values of Weibull fit parameters and extreme value forecasts as predictors. Any forecasts classified as underpredictions could then be multiplied by a safety factor. By extension, forecasts classified as overpredictions could also be scaled down to less conservative values.

To evaluate the hypothesis, the data for the expanded study was transformed into a training data set for classification learning. For every forecast made in time, a feature vector would be constructed consisting of the values of Weibull fit parameters, event rates, and extreme load forecasts taken from the present and the last N time steps. Then, using the knowledge of the actual future extreme load in the forecast window, the feature vector could be labeled as either an over- or underestimate. In the process of creating such a training data set, it was necessary to set several parameters. These are described below:

a) **Predictors:** To repeat, the predictors used for classification were:
   - Weibull scale parameter η
   - Weibull shape parameter β
   - Average event rate in the Weibull sample window
   - Extreme load forecast
b) **Response**: The response or label associated with any feature vector was a binary variable representing whether the extrapolation associated with the feature vector was greater than or less than the actual future extreme event load within the forecast horizon.
c) **Raw Data Parameters:** In assembling the feature vectors for a training data set, it was necessary to fix the following quantities, which were previously introduced when discussing the exploratory study:
   - Forecast window.
   - Weibull fit sample window.

- PNE: The PNE was set at 0.90 and was not altered.
- Different training data sets were developed for different combinations of forecast window and Weibull fit sample window to assess the effect of these parameters on classification accuracy.

d) **Down sampled rate:** The Weibull fit parameters had been determined down to 10-second intervals. However, on an actual vessel, it may be both infeasible and undesirable to make forecasts or even compute fit parameters at such a rapid rate. Weibull fit parameters might only be estimated at frequencies of 1/minute or 1/5minute. Such a down sampling might influence the accuracy of classifying extrapolations as over- or underestimates. Therefore, the training data was generated for several down samplings of the original dataset to compare classification accuracy.

e) **Feature Vector Order: t**he number of past time steps for which to include predictor values in the feature vectors. This value will henceforth be termed the "feature vector order" N. This parameter was varied to study its effect on classification accuracy.

9.3.7.1. Training/Validation Data Split

After computing the feature vectors from the data developed as part of the expanded study, they needed to be divided into training and validation sets. The training data would be used to train a classification model to predict whether the extrapolation associated with each feature vector was an over- or underprediction. The validation data set would be used to test the accuracy of the trained model according to metrics defined subsequently.

In splitting the datasets into training and validation, three considerations came into play:

- Size of training data relative to validation data: A larger training data set would likely lead to a more accurate classification model. However, a model that performs well on a large validation data set would display evidence of more general validity.
- Proportion of positives to negatives in training data: The proportion of positives to negatives in the training data set can affect the classification bias: A model trained on mostly positives may tend to overclassify instances as positive (and similarly for a model trained mostly on negatives).
- Random or chronological selection of training data: If the proposed classification model were to be trained in real time based on real time data, the training data would likely come from a concentrated segment in time. Such a model would be feasible to develop but might not perform as well as one trained on data sampled from a longer time history.

The effect of these considerations on classification accuracy were assessed by generating models for different combinations thereof and comparing the classification metrics.

9.3.7.2. Classification Metrics

Classification accuracy was quantified by the false positive rate (FPR), false negative rate (FNR), and Overall False Rate (OFR). In this context, a positive classification was that an extrapolation was classified as an underestimate. Consequently, extrapolations classified as overestimates were considered negatives. The metrics FPR, FNR, and OFR are defined as:

$$FPR = \frac{FP}{FP+TP} \tag{46}$$

$$FNR = \frac{FN}{FN+TN} \tag{47}$$

$$OFR = \frac{FP+FN}{FP+FN+TP+TN} \tag{48}$$

In the above equations, FP denotes the number of false positive classifications in the validation data set, i.e., instances when negatives were incorrectly classified as positive. FN denotes the number of false negative classifications, i.e., instances when positives were incorrectly classified as negative. TP and TN stand for the number of true positive and true negative classifications respectively. Lower values of FPR, FNR, and OFR indicate higher classification accuracy. For the current application, FNR is a more important metric than FPR, as it is more conservative to have a lower number of underestimates that were falsely classified as overestimates.

9.3.7.3. Parameter Studies

Classification models were developed for many parameter variations to understand the classification sensitivity to each. This section documents these studies, which were conducted by varying a single parameter while keeping all others fixed. Training data was typically randomly sampled from the larger dataset several times to obtain mean classification metrics. Moreover, training data was normalized, such that the mean of each feature was 0.0 and its standard deviation was 1.0.

9.3.7.4. Study 1: kNN vs. Decision Tree

The first study was aimed at identifying the better of two classification model choices: k-Nearest Neighbor or a Decision Tree. In simplistic terms, a k-Nearest Neighbor (kNN) model estimates the response of any feature vector as equal to the majority response of the k nearest feature vectors in the training data. Nearness is typically quantified using Euclidean distance. By contrast, a decision tree issues a response after evaluating a series of conditional statements based on feature vector elements. Table 21 documents the parameters and data characteristics of the dataset used for this study. Default settings in MATLAB's built-in decision tree training function were used. The metrics visualized in Figure 44 are the average of the metrics computed for each random sampling of training data. From this figure, it can be observed that the kNN model performed nearly three times better than the decision tree according to each metric. This result suggests that the dataset can be better modeled using the kNN approach than by other classification algorithms, such as decision trees.

Table 21. Classification parameters used in the kNN vs. decision tree study.

Dataset size (# feature vectors)	17,277	Training Data Size [Fraction of Data]	0.25
Fraction of Underestimates in entire Dataset	0.253	Fraction of Underestimates in Training	0.3
t_{window} [s]	2400	k for kNN	1
$t_{forecast}$ [s]	1800	Number of random samplings	3
Down Sampled Frequency	1/minute	Training Data Selection Method	random
Feature Vector Order	120 (2 hours)		

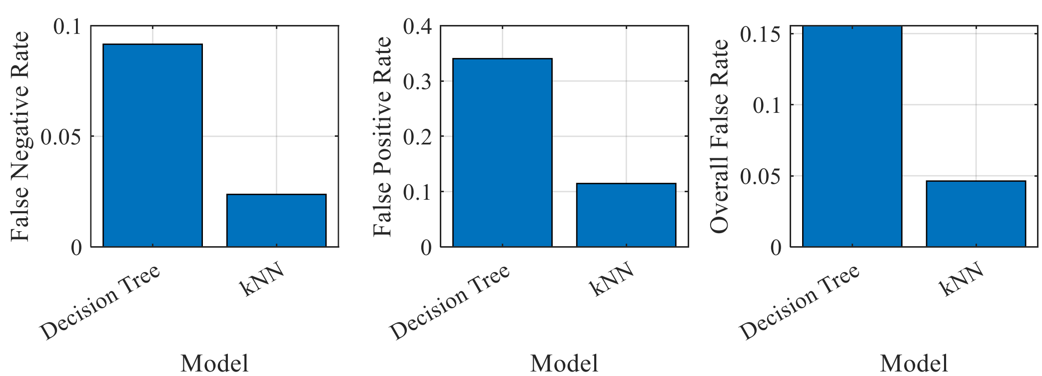


Figure 44. Classification metrics for the kNN vs. decision tree study

9.3.7.5. Study 2: Training Data Selection Study

A second study assessed the differences in kNN classification when training data is selected either randomly or from a single chronologically connected segment of the whole dataset. Table 22 documents the training data parameters, while Figure 45 visualizes the classification metrics. The results indicate models based on chronologically coherent training data are roughly 7 times less accurate than ones based on randomly selected training data of the same size. A possible reason for this is that when training data is randomly selected, each remaining validation feature vector will likely have an associated training data vector taken from a similar moment in time. Feature vectors from moments close together in time will likely also be nearer to each other in feature space and possess the same class label. Hence, a kNN model should be better able to predict the validation data labels given a randomly sampled training dataset.

Table 22. Classification parameters used in the training data selection study.

Dataset size (# feature vectors)	17,277	Training Data Size [Fraction of Data]	0.25
Fraction of Underestimates in entire Dataset	0.253	Fraction of Underestimates in Training	0.3
t_{window} [s]	2400	k for kNN	1
$t_{forecast}$ [s]	1800	Number of random samplings	3
Down Sampled Frequency	1/minute	Training Data Selection Method	varied
Feature Vector Order	120 (2 hours)		

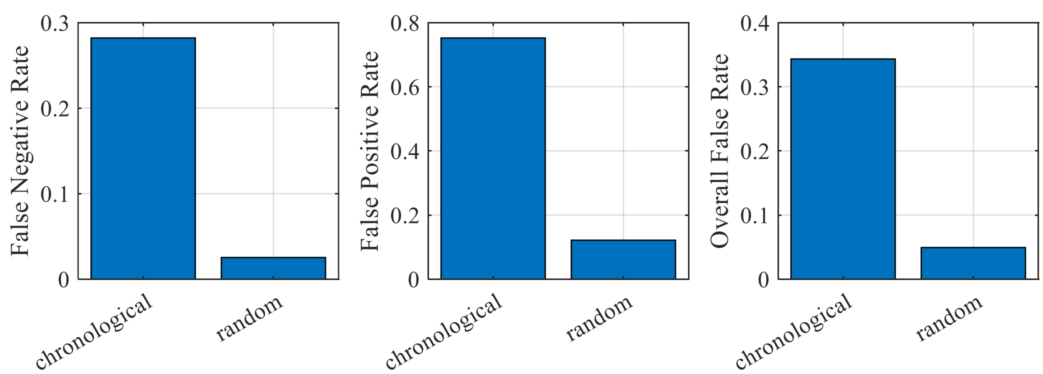


Figure 45. Classification metrics for the training data selection study

9.3.7.6. Study 3 Training Data Size Study

A study was conducted to demonstrate the sensitivity in kNN classification accuracy to the size of the training dataset. Models were trained using training data constituting 10%, 25%, 50%, and 75% of the available data respectively. The other classification parameters were held constant at the values reported in Table 23. Figure 46 plots the classification metrics vs. the training data size as a fraction of the whole. It shows that both the FNR and OFR tend to decrease as the training data size increases. The FPR appears to reach a minimum value at around 50% before increasing again.

Table 23. Classification parameters used in the training data size study.

Dataset size (# feature vectors)	17,277	Training Data Size [Fraction of Data]	varied
Fraction of Underestimates in entire Dataset	0.253	Fraction of Underestimates in Training	0.3
t_{window} [s]	2400	k for kNN	1
$t_{forecast}$ [s]	1800	Number of random samplings	3
Down Sampled Frequency	1/minute	Training Data Selection Method	random
Feature Vector Order	120 (2 hours)		

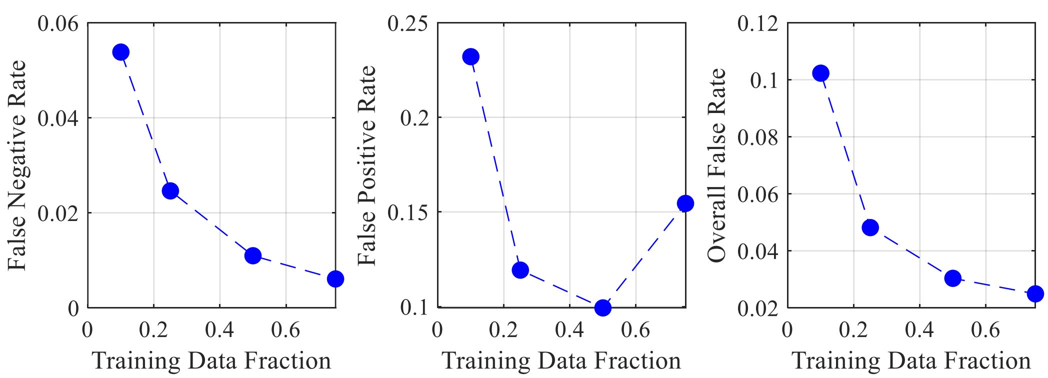


Figure 46. Classification metrics for the training data size study

9.3.7.7. Study 4: K for kNN Study

The kNN classification algorithm estimates the response of a feature vector based on the response of the k nearest neighboring feature vectors in the training data. A study was run to quantify the sensitivity of classification accuracy to the parameter k. Table 24 reports the values of parameters used to generate the training data, while Figure 47 shows the resulting classification metrics for k values of 1, 3, 5, 7, and 9. The plot for the OFR suggests that the rate of false classifications increases almost linearly with the parameter k. Consequently, a k value of 1 is assumed to be optimal for the dataset.

Table 24. Classification Parameters used in k for kNN Study

Dataset size (# feature vectors)	17,277	Training Data Size [Fraction of Data]	0.25
Fraction of Underestimates in entire Dataset	0.253	Fraction of Underestimates in Training	0.3
t_{window} [s]	2400	k for kNN	varied
$t_{forecast}$ [s]	1800	Number of random samplings	3
Down Sampled Frequency	1/minute	Training Data Selection Method	random
Feature Vector Order	120 (2 hours)		

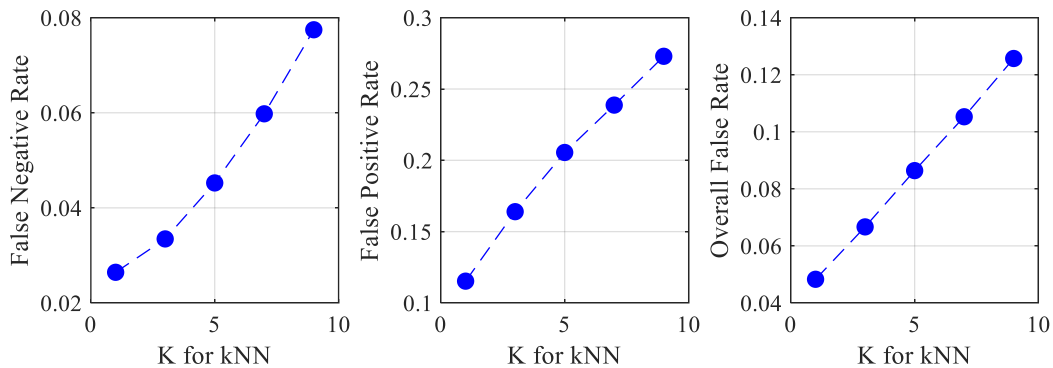


Figure 47. Classification Metrics for the k for kNN Study

9.3.7.8. Study 5: Training Data Composition Study

A study was conducted to determine if there was an optimal proportion of positive to negative class instances in the training data. Using the classification parameters recorded in Table 25, classification metrics were determined for training data sets consisting of 10%, 20%, 25%, 30%, 35%, 40%, and 50% positive classifications (i.e., underestimates of future extreme load). Figure 48 displays the results of this study. One would expect that as the training data consists more and more of positive class instances, the classification model would be more biased in favor of classifying objects as positive. This expectation is borne out by the results, which show that the FNR decreases, and FPR increases with larger positive representation in the training data. The OFR reaches a minimum value when the training data consists of 30% positive class instances. This proportion is similar in value to the actual proportion of positive classes in the entire dataset, which is 25.3% as reported in Table 25. The result indicates that setting the fraction of positive classes in the training data to 30% is close to optimal in terms of classification accuracy.

Table 25. Classification Parameters for the Training Data Composition Study

Dataset size (# feature vectors)	17,277	Training Data Size [Fraction of Data]	0.25
Fraction of Underestimates in entire Dataset	0.253	Fraction of Underestimates in Training	varied
t_{window} [s]	2400	k for kNN	1
$t_{forecast}$ [s]	1800	Number of random samplings	3
Down Sampled Frequency	1/minute	Training Data Selection Method	random
Feature Vector Order	120 (2 hours)		

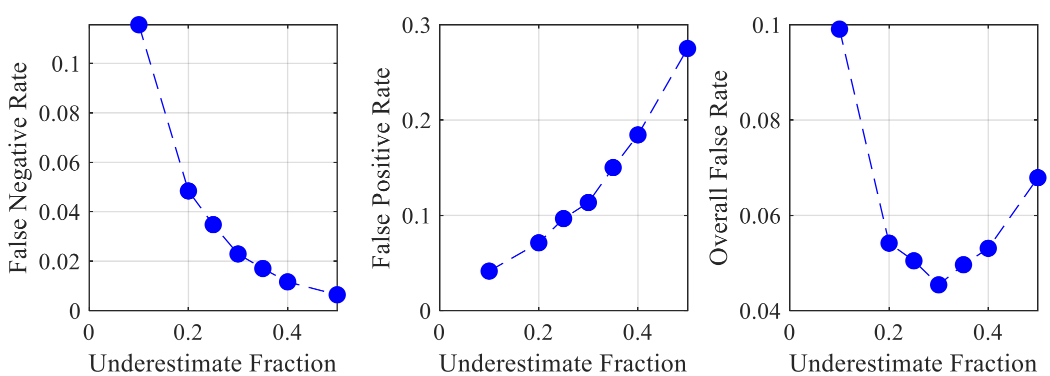


Figure 48. Classification Metrics for the Training Data Composition Study

9.3.7.9. Study 6: Feature Vector Order Study

A study was run to understand the sensitivity of classification accuracy to the feature vector order. Classification parameters were set to the values recorded in Table 26, and classification metrics were computed for training data sets using feature vector orders of 30, 60, 90, 120, 150, 240, and 300. Since the down sampled frequency for the dataset was set to 1/minute, the feature vector orders in this study also equate to the number of minutes into the past from which data was drawn to assemble feature vectors. As Figure 49 shows, classification metrics improve rapidly up until roughly a feature order of 120. After this, gains in classification accuracy are small. Given that a larger feature vector order means longer feature vectors, there is a desire to limit the order so that the classification model does not become overly complex and computationally burdensome. Therefore, the results of this study indicate that, for this dataset, a feature vector order of around 120 strikes an appropriate balance between model accuracy and model simplicity.

Table 26. Classification Parameters for the Feature Vector Order Study

Dataset size (# feature vectors)	17,277	Training Data Size [Fraction of Data]	0.25
Fraction of Underestimates in entire Dataset	0.253	Fraction of Underestimates in Training	0.3
t_{window} [s]	2400	k for kNN	1
$t_{forecast}$ [s]	1800	Number of random samplings	3
Down Sampled Frequency	1/minute	Training Data Selection Method	random
Feature Vector Order	varied		

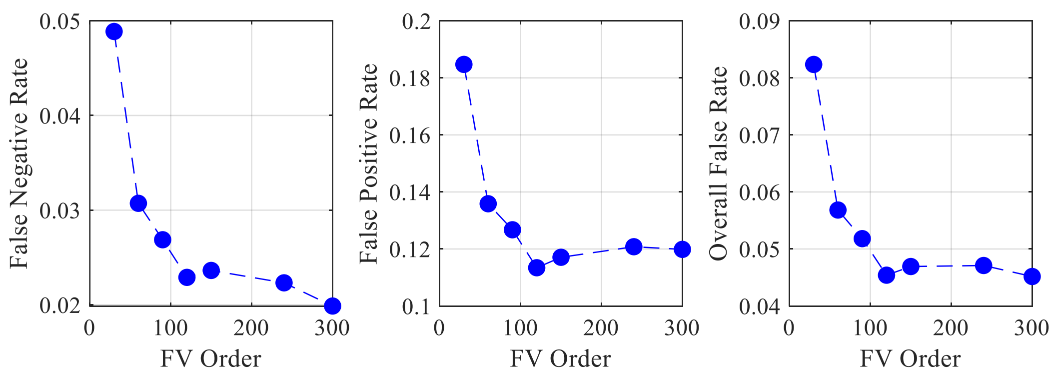


Figure 49. Classification Metrics for the Feature Vector Order Study

9.3.7.10. Study 7: Sensitivity to $\boldsymbol{t_{window}}$

A study was conducted to assess the sensitivity of classification accuracy to the size of the sample window used for determining Weibull distribution fits (t_{window}). Classification metrics were computed for models using the Weibull fit data generated for t_{window} values of 300, 600, 1200, and 2400 seconds. The other classification parameters were set to the values reported in Table 27. Figure 50 plots the classification metrics against t_{window}. Each metric decreases with increasing t_{window}. This result shows that the smoother Weibull fit data associated with larger t_{window} values lead to more reliable classification of over- and underestimates.

Table 27. Classification Parameters used in the $\boldsymbol{t_{window}}$ Sensitivity Study

Dataset size (# feature vectors)	Varies	Training Data Size [Fraction of Data]	0.25
Fraction of Underestimates in entire Dataset	Varies	Fraction of Underestimates in Training	0.3
t_{window} [s]	varied	k for kNN	1
$t_{forecast}$ [s]	1800	Number of random samplings	3
Down Sampled Frequency	1/minute	Training Data Selection Method	random
Feature Vector Order	120 (2 hours)		

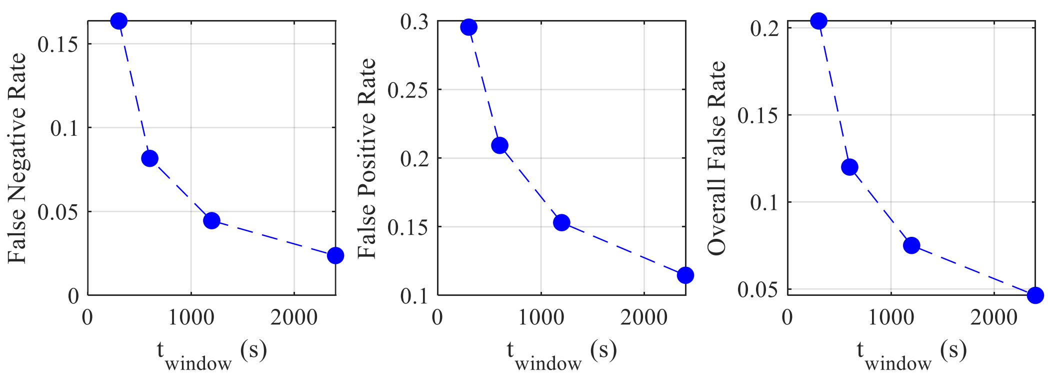


Figure 50. Classification Metrics for the $\boldsymbol{t_{window}}$ Sensitivity Study

9.3.7.11. Study 8: Sensitivity to $\boldsymbol{t_{forecast}}$

A study was conducted to assess the sensitivity of classification accuracy to the duration of the forecast horizon ($t_{forecast}$). Classification metrics were computed for models using for $t_{forecast}$ values of 300, 600, 1200, 1800, 2400, and 3600 seconds. The other classification parameters were set to the values reported in Table 28. Figure 51 plots the classification metrics against $t_{forecast}$. The FNR tends to increase with $t_{forecast}$, while the opposite holds for the FPR. The OFR initially decreases, but then plateaus for $t_{forecast}$ values larger than 1800 seconds. 1800 seconds may therefore be considered an appropriate $t_{forecast}$ value that balances the FPR and FNR.

Table 28. Classification Parameters used in the $\boldsymbol{t_{forecast}}$ Sensitivity Study

Dataset size (# feature vectors)	Varies	Training Data Size [Fraction of Data]	0.25
Fraction of Underestimates in entire Dataset	Varies	Fraction of Underestimates in Training	0.3
t_{window} [s]	2400	k for kNN	1
$t_{forecast}$ [s]	varied	Number of random samplings	3
Down Sampled Frequency	1/minute	Training Data Selection Method	random
Feature Vector Order	120 (2 hours)		

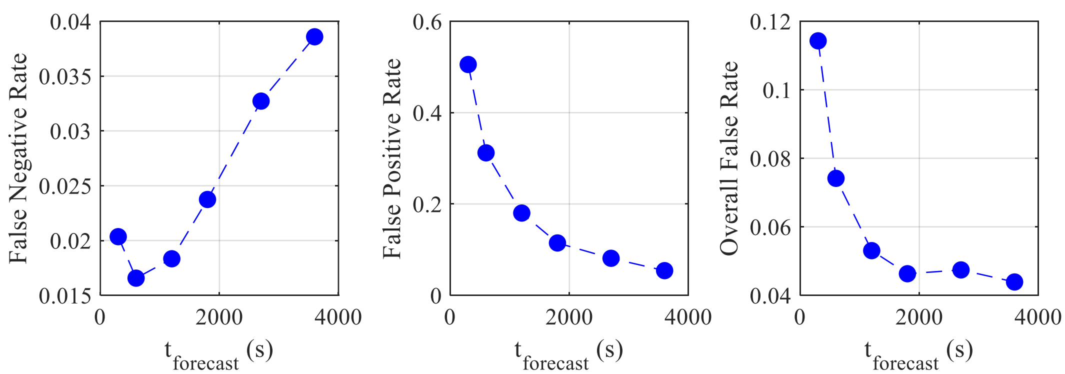


Figure 51. Classification Metrics for the $\boldsymbol{t_{forecast}}$ Sensitivity Study

9.3.7.12. Study 9: Sensitivity to Down Sampled Frequency

A study was conducted to assess the sensitivity of classification accuracy to the down sampled frequency. Classification metrics were computed for models using down sampled frequency values of 1, ½, 1/3, and 1/5 samples per minute. For each down sampled rate, the feature order was adjusted, such that feature vectors contained sampled drawn from the previous 2 hours in the time history. The other classification parameters were set to the values reported in Table 29. Figure 52 plots the classification metrics against the down sampled period, i.e., the inverse of down sampled rate. All metrics tend to increase linearly with down sampled period. An explanation for this is that at higher sampling rates, the chronologically consecutive feature vectors will be more similar and more likely have the same label. A classifier using randomly sampled data would therefore have a stronger validation score, compared to one based on lower frequency data.

Table 29. Classification Parameters for the Down sampled Frequency Study

Dataset size (# feature vectors)	Varies	Training Data Size [Fraction of Data]	0.25
Fraction of Underestimates in entire Dataset	Varies	Fraction of Underestimates in Training	0.3
t_{window} [s]	2400	k for kNN	1
$t_{forecast}$ [s]	1800	Number of random samplings	3
Down Sampled Frequency	varied	Training Data Selection Method	random
Feature Vector Order	2 hours		

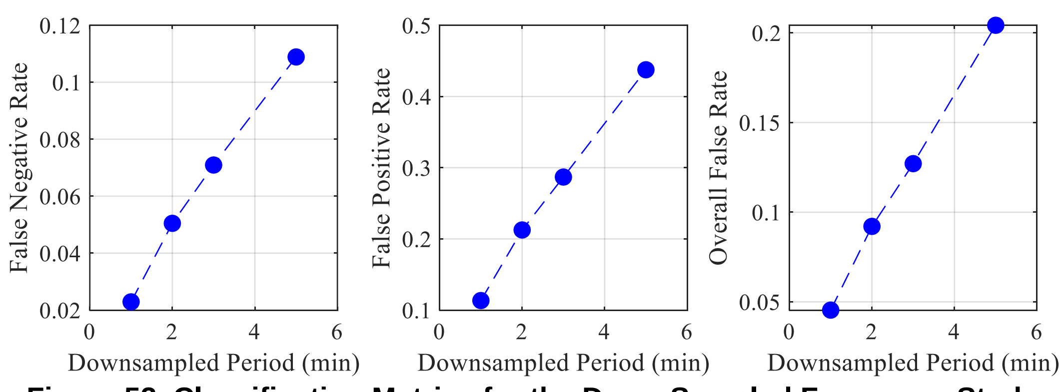


Figure 52. Classification Metrics for the Down Sampled Frequency Study

9.3.7.13. kNN Model Testing: Findings

The single-parameter studies offered insights into the sensitivity of over- and underestimate classification models to the various parameters controlling the form of the training data. They suggested that strongly performing models could be obtained using:

- k-nearest-neighbor algorithm with k=1
- Randomly sampled training data
- 30% positive (underestimate) class representation in the training data
- Data frequency of 1/minute
- $t_{forecast}$=1800 seconds
- t_{window}= 2400 seconds
- Feature Vector Order of 120 (equivalent to 2 hours)

With these parameter settings, the resulting classification model could attain on average a FNR of 2.4%, a FPR of 11.6%, and an OFR of 4.7%.

The studies were conducted using a training and validation data sets drawn from a single month of a vessel's recorded HMS data. To further test the general predictive power of this classification model, additional HMS data from a different month was processed into the form necessary for generation of feature vectors for classification. A model was trained on the entirety of the first month's data. It was then validated against the data from the second month. Table 30 shows the outcome of this validation test. The performance metrics were degraded compared to the validation tests within the original month. This change should have been expected given the results of the training data selection study: Training data taken from chronological chunks resulted in less predictive models than randomly sampled training data. The model has a high false positive rate of over 70%, meaning the model will tend to be conservative in its classifications. Nevertheless, the false negative rate is still over 20%.

Table 30. Confusion Matrix for Classification Model Validation

		Actual Class		
		Negative	Positive	
Predicted Class	Negative	1655	443	FNR=0.211
	Positive	438	155	FPR=0.739
				OFR=0.327

Figure 53 and Figure 54 provide further insight into the classification results. Figure 53 shows that the forecast deviation for the largest forecasted extreme loads was already close to 0. Thus, due to the nature of the dataset, the false negative classifications occurred for lower severity conditions, which would mitigate the risks of false forecasts. However, for a more rigorous assessment, a validation dataset including more examples of underestimates at the upper end of the load distribution should be assessed. Figure 54 shows the distribution of the classified forecast deviations. For a perfect classifier, these distributions would be distinctly separated by the δ=0 line. By contrast, the actual resulting distributions demonstrate significant overlap, thereby highlighting the weakness of the classification model.

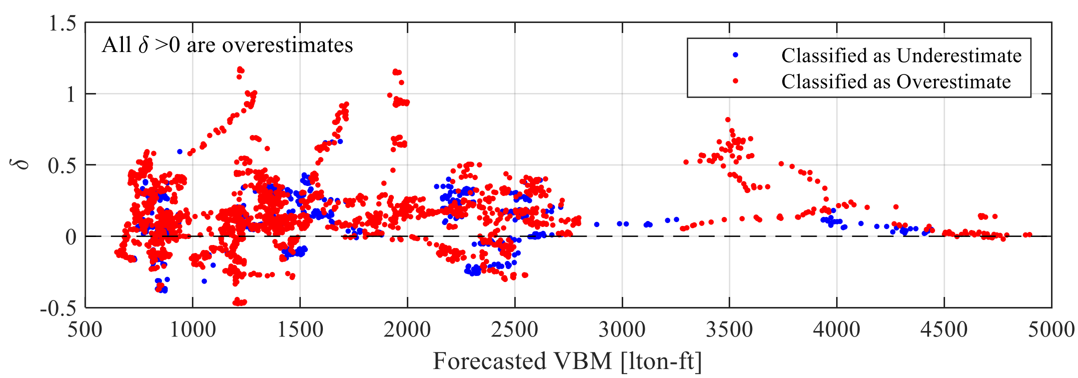


Figure 53. Forecast Deviation vs. Forecasted Extreme Load, Colored by Classification Result

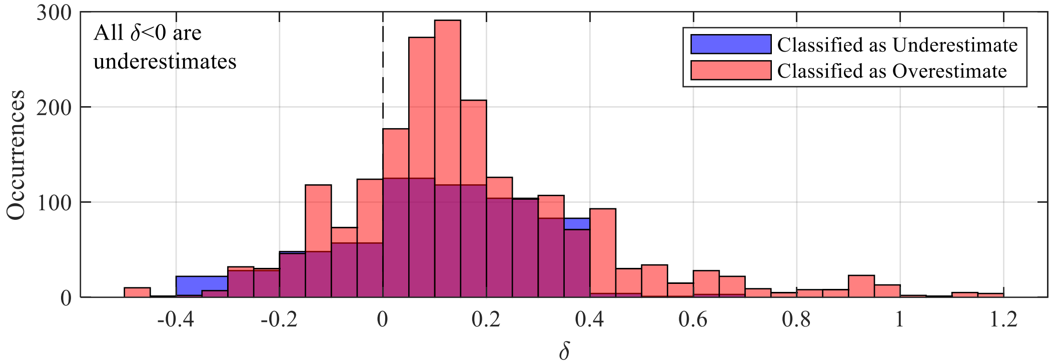


Figure 54. Distribution of classification results

Overall, the validation study showed that although a kNN classifier appeared to work well on the original dataset, it failed to generalize to conditions.

10. Conclusion and Recommendations

This report presents advancements to the digital twin methodologies for integrating data from hull monitoring systems with physics-based models. This report proposes ontology for the design and development of a structural digital twin (SDT) and proposes and investigates functional SDTs through practical examples.

In section 3, this report presents the overarching design considerations for structural digital twins. The broader field recognizes the ample opportunities SDTs provide, such as: reducing maintenance costs, reducing risk, improving efficiency of operations, increasing availability, reliability, and resilience, optimizing re-usability, interoperability, interchangeability, etc. These vast goals, however, have the potential to create an un-tractable solution space, chasing after perfection. To form clear and actionable objectives, this report proposes that the objective for a structural digital twin be defined by a hierarchy of objectives comprised of the following levels: (1) Vision (high level goal), (2) Strategic (targeted mechanism for enabling goal to be met), and (3) Operational (lowest level, scoping of basic requirements and constraints)

In section 4, the design of SDTs is discussed with respect to data and models. The breadth of research formulations and commercial products pertaining to SDTs demonstrates the current gap in defining what a digital twin is (or needs to be). This report identified the need to establish a useful ontology to help guide conversations for digital twins. While seemingly obvious, this report presented the need to clearly articulate what “physics-based models” and “data” requirements are when designing a practical instantiation of a digital twin. To help clarify conversations, this report suggested that one should first identify the type and fidelity of model needed, and the data category. In this way, the required accuracy of the data, accuracy of the model(s), and uncertainty propagation throughout the schema can be established and then designed to.

Section 5 discusses the uncertainties inherent in the evaluations and decision-making processes support by SDTs. The natural variabilities in the loads, strength, and failure mechanisms must be accounted for in an SDT. The errors and uncertainties in the models must also be quantified and tracked. The inherent randomness in environment and ship maneuvers is also an uncertainty that is critical to map into an SDT. This report presents a review of current recommendations in class guidances. The class societies unilaterally identify the need to account for uncertainty but lack details on the specifics. This parallels the state of research for this area, and, as research continues to evolve, so too should the standards. This report identifies a gap in the current standards wherein there is a missing design consideration for inherent randomness (i.e., the ability to incorporate abrupt speed, heading, or environment changes). This omission may lead to SDTs being designed that provide invalid (i.e., false negative) assessments that provides operational guidance that puts the system at risk of being damaged, but relays the guidance to the operator as a safe choice.

Section 6 highlights the current gaps in structural design criteria that supports the performance assessment of in-service assets. That is, most of the class design rules and some of the smart structure guidances invoke criteria that have safety factors on loads, strength, and risk tolerances baked in, with no clear way of isolating and updating or modifying said items based on available data and situational awareness. This report identifies this as a key feature inhibiting the fullest return on investment from the use of digital twins. As such, the report proposes the there is an essential need to establish clear performance-based assessment procedures for the naval and maritime industries. In addition, developing categories for consideration would be useful to help guide the design of digital twins to invoke the appropriate performance-based assessment procedures based on the type of vessel (such as unmanned small boats where there may be a high-risk tolerance to large, manned vessels with lower risk tolerances).

Section 7 of this report proposed a logic tree framework to enable the integration of performance based-assessment techniques with uncertainty considerations into the design of SDTs. First and foremost, the SDT must be developed to be able to place itself in its correct operational state (i.e., branch) considering if its operations are (a) routine or under duress, (b) manned or unmanned, and (c) in normal or heavy weather. This branch identification is used to assign the correct risk tolerance for use. Secondly, the SDT must be developed with a robust probabilistic solver that can account for steady state, gradual changes in state, and abrupt changes in state, so as to avoid false negatives (and putting the vessel at risk). This proposed framework provides the basis for development of an SDT but needs further work to establish how to incorporate system-level assessments. The logic tree formulation also presents a unique opportunity to capitalize on Artificial Intelligence (AI) and Machine Learning (ML) solutions for use in SDTs.

Section 8 takes a step back from the design aspects of SDTs and focuses on the validation process. At the end of the day, the SDT is a product; its primary use is for decision support, and thus, it must be demonstrated that the SDT provides valid decision support. This report emphasis the need to validate on all levels. That is, if an SDT, simply put, is an integrated collection of data driven and physics-based models to support decisions, there are validation needs for all three: the data, the model, and the design support process. The class societies have started to document the expectations for validation and testing. However, this report has identified that further work is needed to clarify the validation requirements and codify the testing methods necessary for validation.

Section 9 does a deep dive into different examples of digital twins developed for surface ships. Three approaches are presented, each targeting a different feature of digital twins. These included surrogate modeling, advanced FE modeling techniques, and predictive assessments.

Approach A (section 9.1) focused on the use of surrogate models in the SDT. Both surrogates were developed using hydrodynamic simulations to capture the physics-based response along the length of the ship. The first method proposed in this report uses surrogate established with the vertical bending moment envelope. This method was found to be computationally fast, explicitly conservative, but potentially overly conservative for broad use. The second uses a lagged regressive model that capitalizes on the temporal and spatial relationships in structural response. This method was also computationally efficient, can be optimized to support use over a broad set of conditions, and is not inherently conservative.

Approach B (section 9.2) focused on advanced FE modeling techniques geared at reducing the epistemic uncertainties that arise from estimating structural loads from strain measurements. This report proposed a method for the systematic formulation of a load basis for conversion matrices. The proposed method capitalized on the natural vibration mode shapes for the structure. It has the advantage of being systematic and repeatable, but requires information on weight, as well as stiffness. Additionally, this investigation identified a viable path to improve loads estimation: invoke principles of structural vibrations and use the developed modes to identify the modal participations factors as a function of time. Further work on this topic is recommended.

Approach C (section 9.3) focused on predictive methods. Three methods were proposed for use. First, a basic approach that statistically characterized the current data and used a prediction window to probabilistically estimate loads in that window. This approach was simple to implement, fast to execute, but (inconsistently) under- or over-predicted future loads during times of rapid changes. Thus, this report identified a critical challenge to predictive metrics: an underlining weakness of extrapolation-based methods in handling non-stationary conditions. A second method attempted to compensate for this weakness by using time-series forecasting methods to anticipate the advent of a condition change and obtain better future loads predictions. This method only achieved modest gains over the first method. Finally, a third method relied on classification modeling with the idea that when forecasts are classified as overpredictions they could also be scaled down. Overall, the study showed that although a kNN classifier appeared to work well on the original dataset, it failed to generalize to conditions.

Again, this report presents advancements to digital twin methodologies through the establishment of SDT ontology for the design and development, and through the practical examples of SDTs developed and applied to real datasets. Whether Digital Twin is viewed as its own topic or as a ‘compilation of capabilities’, this report identifies three things: (1) there is a need to enhance the requirements, design processes, and validation requirements for digital twins, (2) there are technical gaps in each of the fundamental capabilities that need to be addressed, and (3) there are technical gaps in the characterization and propagation of uncertainties throughout all fundamental capabilities in the digital twin. All these items are essential to address before the value of SDTs can be realized.

11. Acknowledgements

The authors are grateful to the Ship Structure Committee (SSC) for taking interest in structural digital twins for surface ships. We would like to specifically thank Dr. Paul Hess and Dr. Jessica Dibelka for chairing the Project Technical Committee (PTC), organizing the meetings, and keeping the project on track. Furthermore, this SSC project leveraged concurrent work under an Office of Naval Research (ONR) program sponsored by Dr. Paul Hess (ONR 331).

We would like to acknowledge the broader Naval Surface Warfare Center, Carderock Division Digital Twin community, the North Atlantic Treaty Organization (NATO) working groups on Digital Twins, and the TTCP working groups that align with Digital Twins. This project both benefited from participation with the vibrant community and added value to it.

12. References

Abernethy, R. B., Breneman, J. E., Medlin, C. H., & Reinman, G. L. (1983). *Weibull analysis handbook*. Wright-Patterson Air Force Base, OH: Air Force Wright Aeronautical Laboratories.

ABS. (2020). *Guide for Hull Condition Monitoring Systems*. Spring, Texas: American Bureau of Shipping.

ABS. (2022). *Guide for Smart Functions for Marine Vessels and Offshore Units*. Spring, Texas: American Bureau of Shipping.

American Institute of Steel Construction. (2016). *Specification for Structural Steel Buildings*. Chicago: Illinois.

Ang, A.H.S., and Tang, W. H. (2007). *Probability Concepts in Engineering: Emphasis on Applications to Civil and Environmental Engineering*, 2nd Edition (2 ed.). Wiley & Sons Inc.

Applied Technology Council. (2018). *Building the Performance You Need: A Guide to State-of-the-Art Tools for Seismic Design and Assessment* (FEMA P-58-7). FEMA.

Assani, N., Matić, P., & Katalinić, M. (2022). Ship's digital twin—a review of modelling challenges and applications. *Applied sciences*, 12(12), 6039.

Barrow, R. S. (2021, June). Do Not Forget Serviceability. *STRUCTURE*.

Baudin, E., Bigot, F., Derbanne, Q., Sireta, F. X., & Quinton, E. (2013, June). Increasing ULCS Structural Response Knowledge Through 3DFEM and a Comprehensive Full-Scale Measurement System. In ISOPE International Ocean and Polar Engineering Conference (pp. ISOPE-I). ISOPE.

Bigot, F., Derbanne, Q., & Baudin, E. (2013). A review of strains to internal loads conversion methods in full scale measurements. *Proceedings of the PRADS2013*, 20(25), 259-266.

Bigot, F., Sireta, F. X., Baudin, E., Derbanne, Q., Tiphine, E., & Malenica, Š. (2015, May). A novel solution to compute stress time series in nonlinear hydro-structure simulations. *In International Conference on Offshore Mechanics and Arctic Engineering* (Vol. 56499, p. V003T02A089). American Society of Mechanical Engineers.

Botín-Sanabria, D. M., Mihaita, A. S., Peimbert-García, R. E., Ramírez-Moreno, M. A., Ramírez-Mendoza, R. A., & Lozoya-Santos, J. D. J. (2022). Digital twin technology challenges and applications: A comprehensive review. *Remote Sensing*, 14(6), 1335.

BV. (2021). *Rule Note NR 675 DT R01 E Additional Service Feature SMART*. France: Bureau Veritas.

BV. (2022, July). *Rules for the Classification of Steel Ships, Part F, Additional Class Notations*. Bureau Veritas.

CCS. (2018). *Guidelines for underwater radiated noise of ships: Section 2 Hull monitoring and assistant decision-making system for operations in ice.* Beijing: CHINA CLASSIFICATION SOCIETY.

CCS. (2020). *Rules for Intelligent Ships*. Beijing: CHINA CLASSIFICATION SOCIETY.

Class NK. (2020). *Rules for Hull Monitoring Systems*. ClassNK.

Danish Maritime Authority. (2018). *Digital Twin Report for DMA: Digital Twins for Blue Denmark*. DNV-GL.

Ditlevsen, O. (1982). Model uncertainty in structural reliability. *Structural safety,* 1(1), 73-86.

DNV GL. (2017*). Rules for classification: Ships* — DNVGL-RU-SHIP Pt.6 Ch.9. DNV GL.

DNV GL. (2020). *Smart vessel* DNVGL-CG-0508. DNV GL.

Dusenberry, D. O. (2019, February). Performance-Based Design is the Future. *STRUCTURE*.

Fonseca, Í. A., & Gaspar, H. M. (2021). Challenges when creating a cohesive digital twin ship: a data modelling perspective. *Ship Technology Research*, 68(2), 70-83.

Galambos, T. V., & Ravindra, M. K. (1981). Load and resistance factor design. *Engineering Journal*, AISC, 18(3), 78-84.

Glaessgen, E. a. (2012). The digital twin paradigm for future NASA and US Air Force vehicles. *In 53rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference 20th AIAA/ASME/AHS Adaptive Structures Conference 14th AIAA.* (p. 1818).

Hess, P. E. (2003). *Reliability-Based Operational Performance Metrics for Ship Structures*. West Bethesda, MD: Naval Surface Warfare Center Carderock Division.

Hughes, O. F., and Paik, J. K. (2010). *Ship Structural Analysis and Design*. Jersey City, NJ: The Society of Naval Architects and Marine Engineers.

International Association of Classification Societies. (2022). *Common Structural Rules for Bulk Carriers and Oil Tankers.* IACS.

Jiang. (2021). *Conversion Matrix Development and Testing* (DRDC-RDDC-2021-C112). Defence Research and Development Canada (DRDC). Contract Report . Prepared by Lloyd's Register Applied Technology Group. CAN Unclassified

Czujko, J., Bayatfar, A., Smith,M., Xu, C., Wang, , D., Lützen, M., Saad-Eldeen, S., Yanagihara,D., Notaro, G., Qian, X., Park, J.S., Broekhuijsen, J., Benson, S., J. Pahos,J., and Boulares, J. (2018). Report of ISSC 2018 Committee III. 1 Ultimate Strength. *In The 20th International Ship and Offshore Structures Congress* (Vol. 1, pp. 335-439).

Kendrick, A., Daley, C., & Pavic, M. (2006). *Comparative Study of Ship Structure Design Standards* Ship Structures Committee Report SR-1444, submitted to US Maritime Administration, by BMT Fleet Technology.

Lee, J. H., Nam, Y. S., Kim, Y., Liu, Y., Lee, J., & Yang, H. (2022). Real-time digital twin for ship operation in waves. *Ocean Engineering*, 266, 112867.

Lewis, R. R. (2005). *A systems-level approach for estimating wave impact design pressures* [Doctoral dissertation, George Washington University].

Lin, W., Berquist, J.R., Collette, M.D., Liut D., Weems, K.M., Weems, M.H.C., and Zhang. S. (2008*) User's Guide to the LAMP System Volume I: Introduction to the LAMP System* (SAIC Report 96/1040). SCIENCE APPLICATIONS INTERNATIONAL CORPORATION, Advanced Systems and Technology Division. Bowie, Maryland, USA.

Llyods Register. (2021, July). *Chapter 16 ShipRight Procedures for the Design, Construction and Lifetime Care of Ships.* Llyods Register.

Merriam-Webster.com. (2022, December 12). Failure. Retrieved from Merriam-Webster: https://www.merriam-webster.com/dictionary/failure

Moyne, J., Qamsane, Y., Balta, E. C., Kovalenko, I., Faris, J., Barton, K., & Tilbury, D. M. (2020). A requirements driven digital twin framework: Specification and opportunities. *IEEE Access*, 8, 107781-107801.

Nielsen, R. E., Papageorgiou, D., Nalpantidis, L., Jensen, B. T., & Blanke, M. (2022). Machine learning enhancement of maneuvering prediction for ship Digital Twin using full-scale recordings. *Ocean Engineering*, 257, 111579.

Nielsen, U. D. (2022). Spatio-temporal variation in sea state parameters along virtual ship route paths. *Journal of Operational Oceanography*, *15*(3), 169-186.

Raju, M., & Anandh, A. (2018). A study on common ship structural failures. *International Journal of Mechanical Engineering and Technology*, 9(7), 746-754.

Rice, S. O. (1944). Mathematical analysis of random noise. *The Bell System Technical Journal*, 23(3), 282-332.

Sudret, B. (2022, November). Surrogate models for uncertainty quantification in computational sciences. *In XLIII Ibero-Latin American Congress on Computational Methods in Engineering (CILAMCE 2022).* Chair of Risk, Safety and Uncertainty Quantification, ETH Zurich.

US Department of Defense . (2018). *Department of Defense Digital Engineering Strategy.* Washington D.C.: Office of the Deputy Assistant Secretary of Defense for Systems Engineering.

VanDerHorn, E., & Mahadevan, S. (2021). Digital Twin: Generalization, characterization and implementation. *Decision support systems*, 145, 113524.

Vukelić, G., & Vizentin, G. (2017). Common case studies of marine structural failures. Failure Analysis and Prevention*. InTech*, 135-51.

Wang, B., Liu, M. L., & Chen, Y. (2021, October). Sensor-Based Structural Health Monitoring for Marine and Offshore Structures*. In SNAME Maritime Convention* (p. D031S012R002). SNAME.

West, S., Stoll, O., Meierhofer, J., & Züst, S. (2021). Digital twin providing new opportunities for value co-creation through supporting decision-making. *Applied Sciences*, 11(9), 3750.

Woolley, A., Mitchell,D., Drazen,D., Tuegel, E., Grisso, B., Mondoro, A., Pegg, N.,, Thompson, I., Greening,I., Guey,J., Magoga, T., Hield,P., Chisholm, J., and Mathys, Z. (2023) Examining User Requirements for a Digital Twin Capability Supporting Naval Platform Management and Operations. *In: AVT-369 Research Symposia on " Digital Twin Technology Development and Application for Tri Service Platforms and Systems"* .