Delft University of Technology

Faculty of Mechanical, Maritime and Materials Engineering

Master of Science

Offshore and Dredging Engineering

Predicting Vessel Motions: A Comparative Analysis of Machine Learning and Conventional Approaches

Thesis

Chairman of Thesis Committee:
Prof. Dr. Andrei Metrikine

Candidate:
Torsten Oosterholt

University Supervisor:
Dr. Hongrui Wang

Company Supervisors:
Ir. Peter Jacobs
Dr. Ir. Onno Peters

Monday, January 15, 2024

"*The only source of knowledge is experience.*"

Dr. Albert Einstein

Abstract

Founded in 1910, Boskalis, a leader in offshore operations, contends with limitations in ABB's Ability Marine Advisory System 'OCTOPUS' in predicting maximum vessel motions for heavy-transport vessels (HTVs). Accurate prediction of these motions, especially roll and pitch, is vital for transporting large, heavy structures, as exceeding predefined limits can jeopardize both vessel and cargo integrity. OCTOPUS's challenges, stemming not only from its reliance on linear theory but also potentially from the quality of its environmental data, underline the need for exploring alternatives, such as Machine Learning (ML) approaches, adept at handling complex, nonlinear phenomena, to ensure operational safety and efficiency.

This thesis presents the development and comparison of three new approaches to predict maximum roll and pitch motions. The approaches are compared and evaluated against OCTOPUS. Two validation strategies are used to test their performance under known and unknown loading conditions (LCs). Known LCs in this context refer to the evaluation of data that incorporate LCs that are included in the training dataset for ML-based approaches. On the other hand, unknown LCs refer to the evaluation of data that incorporate LCs that are not included in the training dataset for ML-based approaches. The approaches are trained and validated using sensor data, LC data, and environmental data from 24 different voyages for a specific HTV. They differ in their design and the type of environmental data they use.

The superior performance of ML-based approaches over OCTOPUS in known LCs is mainly due to two factors. First, ML approaches inherently incorporate nonlinear phenomena, which is particularly effective in accurately predicting maximum roll motion. Second, they are better equipped to handle flaws in environmental data. Although these advantages contribute to a significantly lower mean absolute percentage error (MAPE) compared to OCTOPUS, ML-based approaches face challenges in unknown LCs and extreme motion response scenarios. However, it is noteworthy that ML approaches quickly adapt to unknown LCs when small portions of these LCs are included in the training dataset.

ML shows potential in vessel motion prediction, and this thesis underscores the importance of diverse training data to enhance its reliability in unknown LCs and extreme motion response scenarios. For Boskalis, addressing these challenges with strategies such as adjusting the custom loss function, data augmentation, and implementing ensemble methods could improve the accuracy of these approaches. This progress is significant for Boskalis and the wider maritime industry, paving the way for adaptive and efficient prediction systems. Collaborative efforts between industry and academia, using rich data and expertise, are essential to drive these innovations.

Acknowledgments

I would like to express my gratitude to my graduation committee for their support and guidance. I am especially thankful to Prof. Dr. Andrei Metrikine for his expertise and guidance, which have been influential in the development of this thesis. His insights during the kick-off meeting were crucial in shaping the project. I also appreciate Dr. Hongrui Wang for his assistance in developing the ML-based prediction approaches and for his insightful interpretations of the results. His guidance was essential for the success of this project. Additionally, my thanks go to Ir. Peter Jacobs for his guidance in critically evaluating my work and reminding me to consider Prof. Dr. Metrikine's perspective, which helped me make informed decisions. Peter's advice also improved my presentation skills and aided in crafting a clear narrative for my presentations and thesis. The lessons I have learned from Dr. Hongrui Wang and Ir. Peter Jacobs, particularly in critical thinking and self-evaluation, have been invaluable and will continue to benefit me beyond my academic pursuits. I am grateful to Dr. Ir. Onno Peters for his help in developing robust validation strategies and adding nuance to my conclusions, both of which significantly enhanced the academic value of my research.

I appreciate Ir. Ruud Beindorff, my manager, for his significant role in this project. His enthusiasm in exploring AI and ML was vital in shaping the direction of this research. Our discussions were instrumental in defining the research topic and integrating innovative technologies. I am grateful to the TU Delft for providing essential resources and facilities that have supported this research. Moreover, my time at Boskalis has been greatly enhanced by the collaborative workspace and the expertise shared by many individuals within the organization, which significantly influenced the initial stages of my project. Lastly, I acknowledge the assistance of virtual assistants like ChatGPT and Writefull in improving my coding and writing skills, which were invaluable in completing this thesis.

Contents

List of Figures

List of Tables

1. Introduction

Machine learning (ML) has emerged as a groundbreaking technology in the maritime industry, offering the potential to revolutionize various aspects of maritime operations. It has the ability to process and learn from large datasets, making it a promising tool for improving operational efficiency and safety, especially in complex and nonlinear environments (Black et al., 2021; Feng et al., 2022; Jiang et al., 2021; Li et al., 2017; Khan et al., 2005; Skulstad et al., 2021; Vuttipittayamongkol et al., 2021; Yeter et al., 2023). Boskalis, a major player in the offshore energy sector, has expressed interest in this technological shift toward data-driven approaches. The prediction of vessel motion is a crucial aspect in Boskalis's offshore transport operations. It is necessary to ensure that the operations are carried out within predetermined motion limits. This is important to avoid dangerous situations such as excessive acceleration of the cargo, equipment damage, and to ensure the safety of the crew.

In the challenge of managing vessel motion limits, the focus on roll (rotational motion about the ship's longitudinal axis) and pitch (rotational motion about its transverse axis) becomes particularly significant. The criticality of accurately predicting these motions in transport operations lies in the unique impact these motions have on the loads and stresses experienced by both the vessel and its cargo. Although other motions, such as surge, sway, and yaw, influence overall stability, their impact on the direction of cargo loads is relatively moderate. In contrast, roll and pitch can lead to significant shifts in the direction of the forces exerted on the cargo and vessel structure. This is due to the dynamic nature of these motions, which can alter the orientation of the vessel relative to the sea state, subsequently changing the distribution of forces. Such changes are particularly critical when transporting large sensitive structures such as monopiles, drilling rigs, oil or gas platforms, and transformer substations, as they can substantially affect the structural integrity and operational functionality of the cargo. The stresses induced by roll and pitch motions are not just a matter of magnitude, but also of the changing direction, which can exert uneven pressures and lead to potential failures in the joints and supports of the cargo. This makes an accurate prediction of these motions essential to maintain the safety and integrity of the vessel and its cargo, underlining the need for precise and reliable prediction tools.

To fulfill this objective, Boskalis currently depends on ABB's Ability Marine Advisory System 'OCTOPUS'. This system is used to predict and monitor motions on heavy-transport vessels (HTVs) such as the "Target" (see Figure 1). The extensive data available for the HTV Target, which has a length of 216.79 [m] and a deadweight of 54,000 [t], along with the high quality of these data, make it an ideal choice for this thesis. The construction of the HTV Target was converted in 2007 by Cosco Nantong Shipyard in China and belongs to the t-class HTV family, which originally consisted of six

vessels, all with nearly identical dimensions. Currently, the t-class vessel family still comprises three operational vessels: Trustee, Triumph, and Target.

Figure 1: The Heavy-Transport Vessel "Target" During the Transport of Four Monopiles in 2023

OCTOPUS, widely adopted by Boskalis and others in the maritime industry, offers universal applicability across various vessel types, ensuring versatility in operations. Differing from ML-based predictions, OCTOPUS operates independently of historical data, enabling immediate application in varied scenarios. It functions primarily on linear theory, using spectral analysis where a wave spectrum, derived from environmental data like nowcast data from SPOS, is matched with the vessel's response amplitude operator (RAO). This generates a response spectrum to infer the vessel's maximum responses. Despite its industry-wide use, OCTOPUS struggles with accurately predicting maximum roll and pitch motions. This could be partly due to its conventional approach, which overlooks nonlinear phenomena such as roll damping and adheres to the superposition principle, thereby neglecting arising nonlinear elements. Additionally, the accuracy of OCTOPUS predictions might be compromised by the quality of its environmental data sources, as it relies on nowcast data that, if flawed, could lead to inaccuracies.

This thesis aims to address these limitations by exploring the potential of ML-based approaches to predict maximum roll and pitch motions for the HTV Target. In addition, the objective is to investigate the influence of the type of environmental data on the predictive accuracy. Unlike OCTOPUS, ML-based approaches are not bound by linear theories or fixed rules. Instead, they can learn from real-world motion measurements. By analyzing actual vessel motion data, ML can potentially identify nonlinear patterns that OCTOPUS cannot incorporate, leading to improved reliability. Therefore, the adoption of ML-based approaches goes beyond academic interest; it aims to find an alternative

to existing limitations. The research focuses on comparing ML-based prediction approaches with OCTOPUS and examines the impact of using both hindcast and nowcast environmental data on the performance of these approaches. The distinction between hindcast and nowcast data may be important; while nowcasting, as defined by the WMO, provides forecasts for up to six hours ahead based on current weather details, hindcast data relies on reanalysis, combining model data, including a previous nowcast, with observations into a consistent dataset, as defined by Copernicus (Hersbach et al., 2023). This difference in data quality, with hindcast data arguably of higher quality due to extensive data assimilation and historical coverage, is hypothesized to influence prediction accuracy. The ultimate goal of this comparative study is to help improve future vessel motion prediction systems, providing practical and innovative solutions for future operational improvements at Boskalis.

The main research question that guides this thesis is:

How do ML-based prediction approaches compare with (simulated) conventional prediction approaches when predicting maximum roll and pitch motions of an HTV?

To comprehensively address this question, the following sub-research questions are formulated:

1. What are the key factors, including the role of environmental data types (nowcast vs. hindcast), in predicting the maximum roll and pitch motions using the different prediction approaches, and what are their respective limitations?

2. Which type of ML-based approach is optimal and how does its design and optimization contribute to accuracy?

3. How should the predictive performance of the prediction approaches be validated and evaluated?

4. How does the performance of (simulated) conventional prediction approaches compare to that of ML approaches in vessel motion prediction?

5. Under which conditions and assumptions do ML approaches demonstrate predictive reliability and at what point do their limitations emerge in predicting the maximum roll and pitch motions?

The structure of this thesis is designed to methodically address the objective of this thesis and the main research question with its subquestions. Chapter 2, 'Background,' lays the foundational knowledge in vessel motion dynamics, elaborates further on OCTOPUS, and discusses relevant ML techniques, equipping the reader to fully grasp the subsequent chapters. Chapter 3, 'Methodology,' presents the methodological framework, detailing the comparative analysis of OCTOPUS with newly developed prediction approaches and explaining the validation strategies and adaptability of ML approaches to unknown loading conditions (LCs). Chapter 4, 'Execution,' delves into the practical application of these methods, detailing the specific decisions and actions taken during the research process. In Chapter 5, 'Results,' the outcomes of the predictive performance evaluation of prediction

approaches are presented, offering insights into the effectiveness of ML-based and (simulated) conventional approaches. Chapter 6, 'Discussion,' provides an in-depth analysis of the input data, model development, and validation approaches, further elucidating the findings from the results chapter. Finally, Chapter 7, 'Conclusion and Recommendations,' synthesizes all findings, draws conclusions, and offers recommendations for Boskalis, the broader maritime industry, and academia. The appendices supplement this thesis with additional details on feature descriptions, prediction plots, and supplementary information.

2. Background

This chapter provides the theoretical background required for this research. At the end of this chapter, the reader will be equipped to comprehend the methodology, execution, results, discussion, and conclusions presented in subsequent chapters. However, the reader is expected to possess a basic understanding of mathematics and physics at the Master of Science level to understand the concepts in this chapter. This chapter introduces the concepts of vessel motion response, with a particular focus on roll and pitch motions. It discusses the factors influencing these motions and explores the methods for their prediction. Section 2.1 delves into the realm of vessel motion. Section 2.2 presents OCTOPUS, a conventional approach to vessel motion prediction. Section 2.3 introduces the ML and regression models that could be applied in this context. Finally, Section 2.4 presents an overview of the literature review conducted for this research and highlights a research gap that this study aims to address.

2.1 Vessel Motion: Roll and Pitch

This section serves as a foundation for understanding conventional prediction approaches. It begins by defining the vessel motions and their degrees of freedom in Subsection 2.1.1, followed by an exploration of static floating stability in 2.1.2. Subsection 2.1.3 introduces the vessel motion response to regular waves, leading to an analysis using linear theory in Subsection 2.1.4. The response to irregular waves is discussed in Subsection 2.1.5. The section concludes with an examination of the challenges faced by OCTOPUS in the prediction of vessel motion in Subsection 2.2.6. Collectively, these aspects aim to provide a comprehensive understanding of the interplay between vessel motions, environmental conditions, and operational specifications in offshore environments. Much of the content in this section is derived from the work of Journée and Massie (2008), particularly on the derivation of the Response Amplitude Operator and the discussion of linear and potential theory.

2.1.1 Degrees of Freedom in Vessel Motion

Figure 2 illustrates the six degrees of freedom in vessel motion, each of which corresponds to a specific type of periodic motion. For instance, roll motion, represented by ϕ, is a periodic angular rotation on the longitudinal x-axis of the vessel, while pitch motion, denoted by θ, is a periodic angular rotation around the y-axis, which runs throughout the width of the vessel. All six motions together determine the response of a vessel in an offshore environment, but their impact on heavy transport operations varies. Ribadeneira et al. (2022) point out that roll motion is a significant limiting factor for floating

vessels experiencing multimodal wave spectra. A captain or superintendent usually tries to head the vessel into the waves to reduce roll (and thus increase pitch). However, most offshore sites are multimodal, meaning that there is no single wave direction. This means that the vessel will be affected in the roll direction (Ribadeneira et al., 2022). This view is supported by Ibrahim and Grace (2010), who discuss the importance of roll motion in static vessel stability, taking into account factors such as relative motion, local wave slopes, wind, cargo shifts, and accumulation of water or ice on deck. Sayed and Hamed (2010) demonstrate that roll and pitch motions are interdependent, with an increase in one motion often leading to increased motions in the other. Both roll and pitch motions are limiting factors during Boskalis' transport operations. This is because the large and heavy cargo welded on deck significantly affects the vessel's stability and center of gravity. High accelerations in roll and pitch motions create substantial dynamic stresses. In the roll, side-to-side movement leads to uneven force distribution, which strains the welds with asymmetric loads. In pitch, the bow and stern's up-and-down movement subjects the welds to varying tensile and compressive forces, particularly at high velocities. These stresses are intensified when roll and pitch occur simultaneously, imposing complex, multidirectional forces on the weld joints, making these motions critically impactful for the vessel and cargo's safety and integrity.

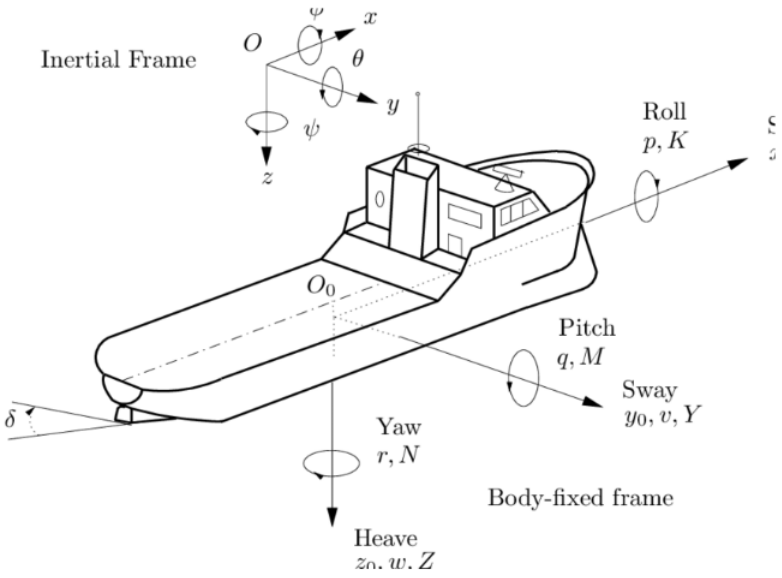


Figure 2: Illustration of Vessel Motions in the Six Degrees of Freedom. Adapted from Seyffert, H. (2022). Lecture 1: Introduction, hydrostatics, & buoyancy. *TU Delft Lecture - Introduction to Ship and Offshore Hydromechanics (MT44045).* https://brightspace.tudelft.nl/d2l/le/content/500870/viewContent/2718364/View

2.1.2 Static Floating Stability

It is essential to understand the nuances of static stability when comprehending the roll and pitch responses of a vessel. According to Journée and Massie (2008), when in still water, a vessel is exposed to a variety of hydrostatic forces and moments that attempt to bring it to a state of equilibrium. This equilibrium can be neutral, stable, or unstable. In a neutral equilibrium, the vessel will remain in its position when undisturbed, but will not return to it if moved. On the other hand, a stable equilibrium will not only keep the vessel in its initial position when undisturbed, but will also bring it back to that position after being displaced. This stability is usually determined by the vessel's center of gravity and center of buoyancy; a lower center of gravity in comparison to the center of buoyancy leads to greater stability. In contrast, an unstable equilibrium will cause the vessel to move further away from its original position when disturbed, usually due to a higher center of gravity, as Journée and Massie (2008) points out. Knowing these equilibrium states is essential for predicting how a vessel will react to dynamic forces, such as waves and wind, which cause roll and pitch motions. Having established the fundamentals of equilibrium and its various forms in the context of a vessel's static stability, it is essential to explore the details. According to Journée and Massie (2008), this comprehension can be further refined by dividing equilibrium into three essential components, all of which are necessary for the vessel's overall stability and reaction to external forces. These components are:

1. Horizontal equilibrium: achieved when the sum of all horizontal forces equals zero.

2. Vertical equilibrium: achieved when the total of all vertical forces is zero. This is crucial for maintaining the buoyancy of the vessel and preventing it from sinking or rising excessively, as outlined by Journée and Massie (2008).

3. Rotational equilibrium: achieved when the moments around the vessel's center of gravity (CoG) - or any other point - are balanced. This equilibrium is the key to preventing unwanted rotational movements that can lead to instability in roll and pitch, according to Journée and Massie (2008).

Each of these equilibrium states plays an important role in the vessel's ability to maintain stability and effectively respond to dynamic marine conditions.

Righting Moments and Metacentric Height

Figure 3 demonstrates a case where the structure experiences an external heeling moment M_H, modifying the underwater shape of the structure, leading to a CoB displacement. This shift generates a righting stability moment M_S due to the distance $\overline{GZ}$. Static stability is achieved when the righting moment balances the external heeling moment. The distance $\overline{GZ}$ is expressed as $\overline{GN_\phi} \cdot \sin\phi$, resulting in:

$$M_s = \rho g \nabla \cdot \overline{GN_\phi} \cdot \sin\phi \tag{2.1}$$

Here, N_ϕ represents the structure's metacentre, where buoyancy force verticals intersect (Journée and Massie, 2008). Generally, the initial metacentre is referred to as M, and therefore the metacentric height is written as $\overline{GM}$. A larger $\overline{GM}$ increases $\overline{GZ}$, enhancing the righting moment M_S. This results in a structure that is more able to handle larger external heeling moments.

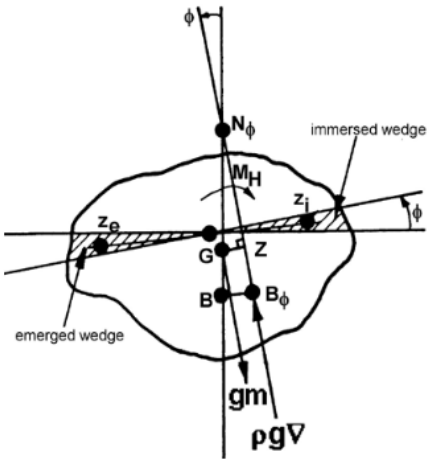


Figure 3: Illustration of Rotational Equilibrium with Heel Angle ϕ. Adapted from Journée, J., & Massie, W. (2008). *Offshore hydromechanics*. Delft University of Technology.

Free Surface Correction

Thus far, the discussion on static floating stability has focused primarily on solid and homogeneous structures. However, the introduction of liquids within a floating structure, particularly those with a free surface, alters the dynamics. Although fully filled tanks, lacking a free surface, behave similarly to a homogeneous and stationary load, Journée and Massie (2008) emphasize that partially filled tanks with a free surface notably affect static stability. This influence originates from the mass change of the liquid in the direction of rotation, leading to a decrease in the righting moment and subsequently a reduction in the metacentric height $\overline{GM}$. This effect is known as the free surface correction. Figure 4 visually illustrates this phenomenon.

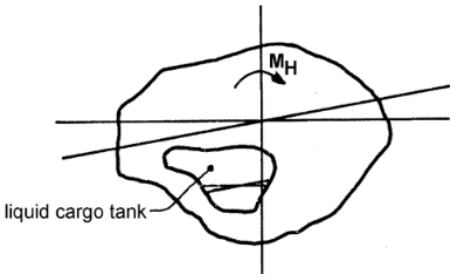


Figure 4: Illustration of Free Surface Impact on a Heeled Structure. Adapted from Journée, J., & Massie, W. (2008). *Offshore hydromechanics*. Delft University of Technology.

2.1.3 Vessel Motion Response to Regular Waves

This subsection delves into the dynamic behavior of a vessel in response to regular waves. Understanding this behavior will later help in understanding the Response Amplitude Operator of a vessel and eventually helps understanding the conventional prediction approach.

Journée and Massie (2008) describe a situation where a vessel faces a regular wave characterized by specific frequency and amplitude. The wave function at the origin, representing the elevation of a regular wave, is formulated in Equation 2.2:

$$\zeta = \zeta_a \cos(\omega_e t) \tag{2.2}$$

Here, ζ signifies the wave function, ζ_a the wave amplitude in meters, and $\cos(\omega_e t)$ the oscillation pattern. The term ω_e denotes the encounter frequency in radians per second, influenced by both the frequency of the wave and the forward velocity of the vessel (Journée and Massie, 2008). The variable t represents time in seconds.

The interaction with a regular wave initiates a steady and periodic oscillation in the vessel. This motion is captured by Equation 2.3, which demonstrates that the vessel's oscillation frequency is equal to that of the incoming wave, although they can differ in phase, indicated by the phase shift ϵ.

$$\begin{aligned}
\text{Surge:}\ & x = x_a \cos(\omega_e t + \epsilon_{x\zeta}) \\
\text{Sway:}\ & y = y_a \cos(\omega_e t + \epsilon_{y\zeta}) \\
\text{Heave:}\ & z = z_a \cos(\omega_e t + \epsilon_{z\zeta}) \\
\text{Roll:}\ & \phi = \phi_a \cos(\omega_e t + \epsilon_{\phi\zeta}) \\
\text{Pitch:}\ & \theta = \theta_a \cos(\omega_e t + \epsilon_{\theta\zeta}) \\
\text{Yaw:}\ & \gamma = \gamma_a \cos(\omega_e t + \epsilon_{\gamma\zeta})
\end{aligned} \tag{2.3}$$

By differentiating with respect to time, the velocity and acceleration for these movements are derived. Equation 2.4 describes the roll motion, including its displacement, velocity (in radians per second), and acceleration (in radians per second squared):

$$\begin{aligned}
\text{Displacement:}\ & \phi = \phi_a \cos(\omega_e t + \epsilon_{\phi\zeta}) \\
\text{Velocity:}\ & \dot{\phi} = -\omega_e \phi_a \sin(\omega_e t + \epsilon_{\phi\zeta}) \\
\text{Acceleration:}\ & \ddot{\phi} = -\omega_e^2 \phi_a \cos(\omega_e t + \epsilon_{\phi\zeta})
\end{aligned} \tag{2.4}$$

In these equations, ϕ represents the displacement of the roll motion, $\dot{\phi}$ indicates the velocity of the roll motion, and $\ddot{\phi}$ signifies the acceleration of the roll motion. ϕ_a is the roll amplitude in radians, ω_e

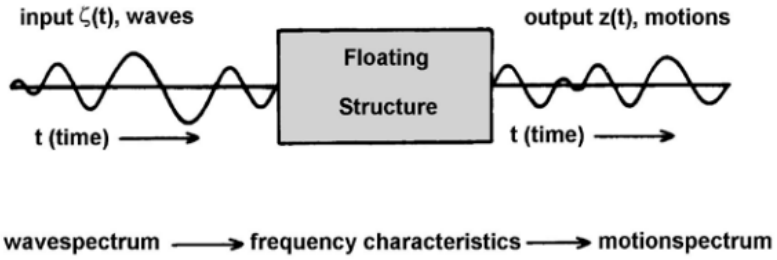


Figure 5: Linear System Describing Relation between Motion and Waves

again represents the frequency of encounter in radians per second, t represents the time in seconds, and finally $\epsilon_{\phi\zeta}$ denotes the phase shift in radians.

2.1.4 Linear Theory and External Loads

Journée and Massie (2008) further note that for an offshore vessel subjected to irregular waves with a known energy distribution across wave frequencies, it is assumed that linear theory can be used to calculate the response of the vessel to waves by deriving the response amplitude operator (RAO). This presupposes the vessel as part of a linear system in which both the input wave spectrum and the frequency characteristics are required for the derivation of the response spectrum. Figure 5 presents a schematic overview of such a system. In this subsection, the derivation of the RAO will be explained step by step, based on the work of Journée and Massie (2008). The RAO is commonly used in conventional prediction approaches and is a necessity to determine the response spectrum.

The derivation starts with the introduction of Equation 2.5, based on Newton's second law.

$$\sum_{j=1}^{6} m_{i,j} \cdot \ddot{x_j} = F_i \text{ for } i = 1, ...6 \tag{2.5}$$

Here, $m_{i,j}$ resembles a 6x6 matrix which includes the solid mass and inertia of the body, $\ddot{x_j}$ denotes the acceleration in direction j (see Figure 2), and F_i represents the sum of forces or moments exerted in direction i.

Linear theory encompasses two primary load types: hydrodynamic loads and wave-exciting loads. Employing the superposition principle, one can add the hydrodynamic loads, which consider the vessel's oscillation in still water, with the wave exciting loads, which account for the response in regular waves.

Hydrodynamic Loads

Focusing on the roll motion, yet acknowledging that it applies to all six vessel motions, the equation of roll motion, considering solely hydrodynamic loads, is formulated in Equation 2.6.

$$m\ddot{\phi} = F_h \quad \text{with} \quad F_h = -a\ddot{\phi} - b\dot{\phi} - c\phi \tag{2.6}$$

In this equation, m refers to the cylinder's total mass, a represents the hydrodynamic mass coefficient or added mass (in $\frac{Ns^2}{m} = kg$), b is the hydrodynamic damping coefficient ($\frac{Ns}{m} = \frac{kg}{s}$), and c is the restoring spring coefficient ($\frac{N}{m} = \frac{kg}{s^2}$).

The hydromechanic reaction force, $a\ddot{\phi}$, originates from accelerations transmitted to the water particles near the cylinder. This force does not dissipate energy from the system and resembles a standing wave adjacent to the cylinder. The coefficient a is mass-dimensional and intuitively symbolizes the additional mass of water displaced during sailing. The hydrodynamic damping coefficient, b, describes the energy dissipation due to wave generation and propagation away from the cylinder. Other forms of damping, such as viscous damping (discussed in Section 2.2.6), are not considered here. The restoring spring coefficient, c, originates from the buoyancy of the cylinder and functions as a restoring force.

Revising the Equation of motion for roll motion of a cylinder in decaying motion in still water from Equation 2.6 yields:

$$(m + a)\ddot{\phi} + b\dot{\phi} + c\phi = 0 \tag{2.7}$$

Wave Loads

Linear theory dictates that the total wave force is hypothesized as a combination of the Froude-Krylov force, a diffraction force originating from the acceleration of the water particles, and a force proportional to the velocity of the particles. According to Journée and Massie (2008), the total wave force in a linear system facing a regular incoming wave can be described by the following equation:

$$F_w = a\ddot{\zeta^*} + b\dot{\zeta^*} + c\zeta^* \tag{2.8}$$

The expression $a\ddot{\zeta^*} + b\dot{\zeta^*} + c\zeta^*$ includes corrections to the Froude-Krylov force due to wave diffraction, as well as the Froude-Krylov force itself. The star annotation on ζ^* indicates that it is the reduced wave elevation, which is a modified form of the wave elevation that takes into account the effects of depth and other modifying factors. This reduced wave elevation is expressed as $e^{-kT}\zeta_a \cos(\omega t)$, where e^{-kT} is a depth-related reduction factor. Here, k is the wave number, defined as $k = \frac{2\pi}{\lambda}$ (with λ being the wavelength), and T describes the wave period in seconds. ζ_a is the amplitude of the wave elevation, and ωt is the frequency and time of the wave, respectively. The use of the reduced wave elevation is

important in this context, as it provides a more accurate representation of the wave forces acting on a structure such as a cylinder in deep water, taking into account the depth-related modifications to the wave characteristics (Journée and Massie, 2008).

Journée and Massie (2008) describe the Froude-Krylov force as a pressure force acting on the submerged vessel due to incident waves. The force can be calculated by integrating the hydrodynamic pressure from the undisturbed incident wave along the wetted surface of the vessel.

Adding the wave force to the previously derived Equation of motion (Equation 2.7) in accordance with the superposition principle, the equation of motion for an oscillating cylinder becomes:

$$(m+a)\ddot{\phi} + b\dot{\phi} + c\phi = a\ddot{\zeta^*} + b\dot{\zeta^*} + c\zeta^* \tag{2.9}$$

Determining the response to a regular wave involves combining the roll displacement, velocity, and acceleration according to Equation 2.4, and the reduced wave elevation equation. Differentiation of the reduced wave elevation proceeds as follows:

$$\zeta^* = \zeta_a e^{-kT}\cos(\omega t),\ \ \dot{\zeta^*} = -\zeta_a e^{-kT}\omega\sin(\omega t), \text{ and } \ \ddot{\zeta^*} = -\zeta_a e^{-kT}\omega^2\cos(\omega t) \tag{2.10}$$

Journée and Massie (2008) demonstrate that substituting equations 2.10 and 2.7 into Equation 2.9, and utilizing $\phi = e^{-i\omega t}$, ultimately yields the RAO for roll of the cylinder to the regular incoming wave at frequency ω:

$$\frac{\phi_a}{\zeta_a} = e^{-kT}\sqrt{\frac{(c-a\omega^2)^2 + (b\omega)^2}{(c-(m+a)\omega^2)^2 + (b\omega)^2}} \tag{2.11}$$

The RAO acts as a transfer function, accurately depicting the connection between the wave spectrum and the resulting motion response of the vessel, such as roll and pitch. The natural frequency of these motions, calculated using $\omega_n = \sqrt{\frac{c}{(m+a)}}$, is a key factor in this context. It is determined by the system's stiffness and the combined mass of the vessel, including the added mass due to the surrounding water. This stiffness is quantified by the restoring spring term $c_{\phi\phi} = \rho g \nabla \overline{GM}$, where $c_{\phi\phi}$ is the restoring spring coefficient for the roll, ρ is the water density, ∇ the volume displaced, and $\overline{GM}$ the initial metacentric height, as outlined in Journée and Massie, 2008.

The significance of the natural period, which is the inverse of the natural frequency, in the context of RAO is profound. It essentially represents the vessel's inherent tendency to oscillate in roll or pitch in response to wave excitation. When the frequency of the incoming waves approaches the natural frequency of the vessel, resonance can occur, leading to large-amplitude motions. The RAO peaks at or near this resonant frequency, indicating that the vessel's response to wave excitation is maximized. This phenomenon underscores the importance of understanding and calculating the natural period. By doing so, designers and operators can predict and mitigate scenarios where the vessel's response to the

wave spectrum could be dangerously amplified, ensuring safer and more efficient maritime operations.

Potential Theory

Transitioning to potential theory, its role is instrumental in the computation of hydrodynamic coefficients a, b, and c, and in deriving first-order hydrodynamic wave loads. This theory assumes that a rigid body oscillates in or beneath a free surface. The velocity potentials derived from the linearized Bernoulli equation are utilized to calculate the hydrodynamic loads on the body. According to Journée and Massie (2008), a potential function within a potential flow field is a mathematical construct that allows the determination of velocity components at any point in the flow by differentiation. Integration of these potentials into the body surface is essential for a comprehensive understanding of the dynamic response of the vessel to wave loads. In this context, two main theoretical approaches are addressed: 2D Strip Theory and 3D Diffraction Theory.

2D Strip Theory: In the context of potential theory, 2D Strip Theory provides a simplified but effective approach to calculate the hydrodynamic coefficients and exciting wave loads on vessels. This method considers the vessel as a series of 2D cross sections along its length (see Figure 6) and calculates hydrodynamic coefficients by integrating potential solutions across these sections. A key assumption of strip theory is that all motions of the vessel in waves are linear, and only the external loads on the vessel's wetted surface are considered. The shape of the vessel's hull, along with its forward and aft drafts, defines this wetted surface. Despite its effectiveness for slender vessels (length-to-breadth ratios of 3.0 or more), strip theory has limitations due to its 2D nature, notably neglecting cross-sectional interactions and assuming a steady-wave system. Therefore, its accuracy decreases for high-speed vessels and significant vessel motions where these assumptions are invalid. Moreover, strip theory assumes minor vessel motions compared to the vessel's dimensions, focusing only on the hydrodynamic effects below the still water level.

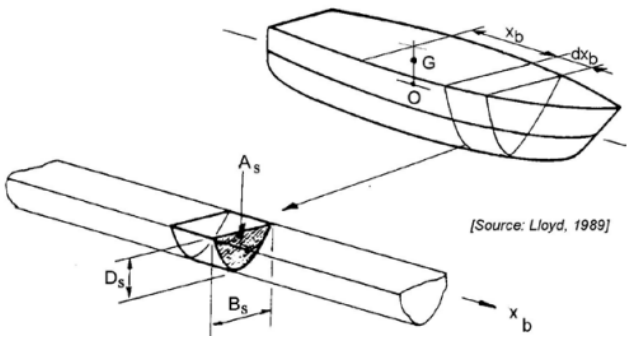


Figure 6: Strip Theory Representation by Cross Sections. Adapted from Journée, J., & Massie, W. (2008). *Offshore hydromechanics.* Delft University of Technology.

3D Diffraction Theory: The 3D diffraction theory, an alternative to strip theory, involves the 3D

panel method to calculate the exciting forces of the first-order waves and the oscillating hydrodynamic forces on offshore vessels. This method segments the vessel hull into multiple panels (see Figure 7), with hydrodynamic loads calculated by integrating pressure over the area of each panel. The overall hydrodynamic load is determined by summing these contributions. The 3D panel method offers a more comprehensive approach, accounting for the three-dimensional nature of vessel and wave interactions, and is especially suitable for vessels with complex geometries or in highly irregular wave conditions.

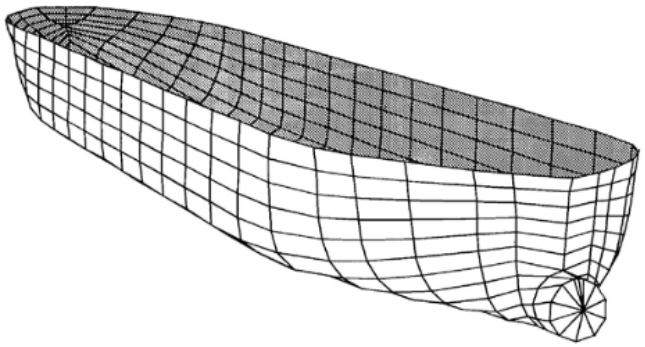


Figure 7: 3-D Representation of Hull Divided in Panels. Adapted from Journée, J., & Massie, W. (2008). *Offshore hydromechanics.* Delft University of Technology

.

2.1.5 Vessel Motion Response to Irregular Waves

The response of a vessel to irregular waves is intricately linked to the concept of a wave energy spectrum, which quantifies the distribution of energy across different wave frequencies. Irregular waves, unlike regular waves, do not have a single frequency or amplitude, but are a superposition of many regular waves with varying frequencies and amplitudes. This complexity is captured in the wave energy spectrum $S_\zeta(\omega)$, which is dependent on environmental factors such as significant wave height and wave period, as noted by J.M.J. Journée (2012).

Different wave spectra, such as the Neumann, Bretschneider, and JONSWAP spectra, model this distribution with varying degrees of peakedness and energy distribution across frequencies. The JONSWAP spectrum, for example, is characterized by its concentration on specific frequencies, particularly in the North Sea region (Hasselmann et al., 1973). This contrasts with the more evenly distributed energy across various wave frequencies seen in the Neumann spectrum. The Bretschneider spectrum offers an intermediate profile, representing an average spectrum. These spectra are essential to understand the interaction of waves with vessels, as shown in Figure 8.

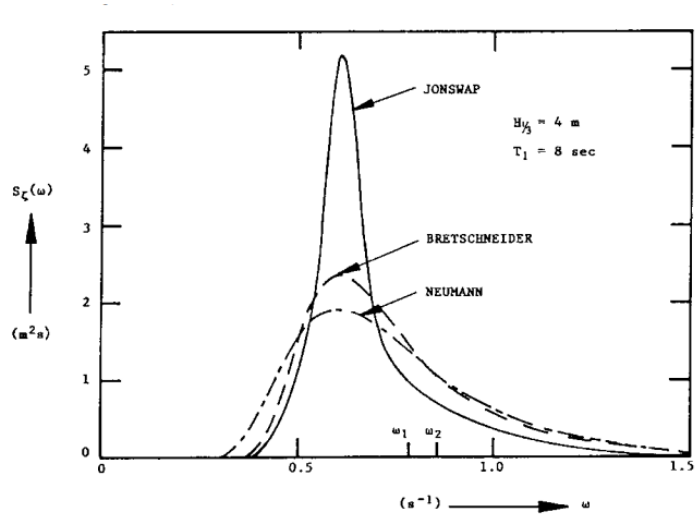


Figure 8: Three Energy Wave Spectra with $H_s = 4$ [m] and $T_1 = 8$ [s], adapted from J.M.J. Journée, L. A. (2012). Octopus technical note (1st ed.).

The spectral analysis approach decomposes irregular waves into their constituent regular wave components through superposition. This allows for the calculation of the vessel's response spectrum $S_\phi(\omega)$, which is a function of both the wave energy spectrum $S_\zeta(\omega)$ and the vessel's Response Amplitude Operator (RAO) $\frac{\phi_a}{\zeta_a}(\omega)$ as shown in Equation 2.12:

$$S_\phi(\omega) = \left|\frac{\phi_a}{\zeta_a}(\omega)\right|^2 \cdot S_\zeta(\omega) \tag{2.12}$$

Here, the RAO captures how the vessel's motion amplitude relates to the wave amplitude, varying across different frequencies. This relationship is important in predicting vessel motion responses in the maritime engineering domain.

The graphical representation of the spectral analysis in Figure 9 shows the transformation of the wave energy spectrum into the response spectrum of the vessel. It highlights how the energy response spectrum peaks at frequencies where both the energy wave spectrum and the RAO are heightened, thus allowing for a nuanced understanding of vessel responses to different wave conditions.

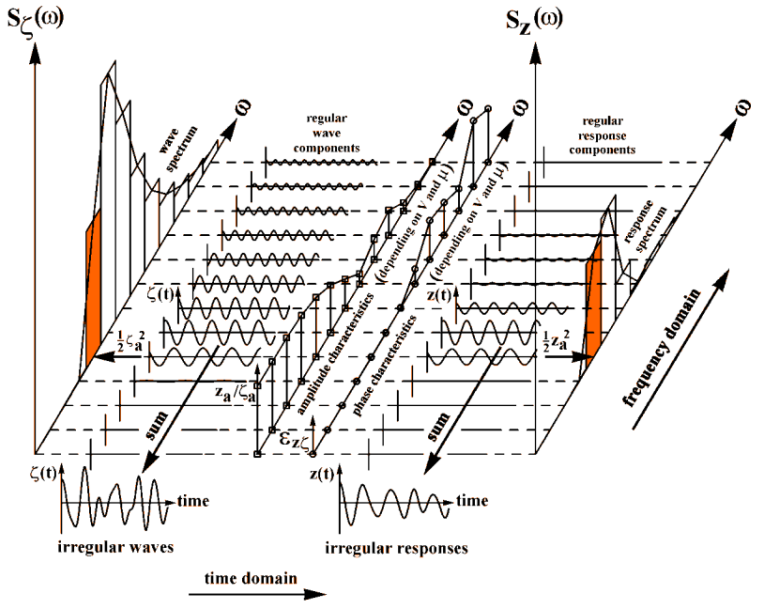


Figure 9: Spectral Analysis, adapted from J.M.J. Journée, L. A. (2012). Octopus technical note (1st ed.).

2.2 Conventional Prediction Approach: OCTOPUS

OCTOPUS, an integral component of the ABB Ability Marine Advisory System used by Boskalis, is pivotal in managing motion limits on HTVs such as the Target. This section links the theoretical aspects of vessel motion response, specifically RAO and spectral analysis, with their practical application in OCTOPUS.

2.2.1 Overview of OCTOPUS

OCTOPUS integrates various methodologies related to vessel motion prediction, focusing on RAO calculations, selection of wave energy spectrum, and spectral analysis. Figure 10 provides an overview of the approach of the system in determining the motion responses of the vessel.

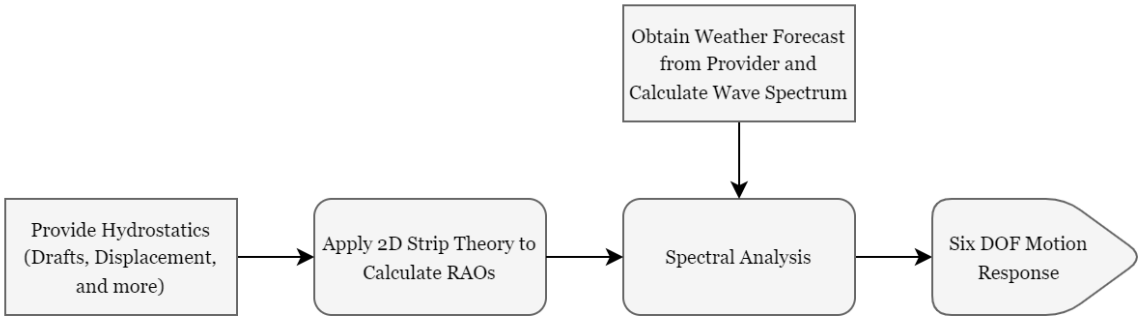


Figure 10: High Level Flow Diagram of OCTOPUS, based on descriptions in J.M.J. Journée, L. A. (2012). Octopus technical note (1st ed.).

2.2.2 RAO Calculations in OCTOPUS

The 2D Strip Theory, as seen in Figure 6 from Section 2.1, is a key component of OCTOPUS. This theory is adapted to the specific conditions of maritime operations, taking into account linear vessel motions and external loads on the wetted surface, as well as the vessel's forward speed. This adaptation provides a more realistic representation of the dynamic LCs at sea. OCTOPUS uses the 2D Strip Theory to calculate hydrodynamic coefficients and wave exciting loads by integrating these 2D loads over the vessel's length. These coefficients are then used to compute the RAOs, which is a crucial connection between wave spectra and vessel motion responses.

2.2.3 Wave Energy Spectrum in OCTOPUS

OCTOPUS provides a range of wave energy spectra, including Neumann, Bretschneider, and JONSWAP, which are tailored to environmental parameters such as the significant wave height and average wave period, which can be obtained from weather forecasts. The JONSWAP spectrum also has a variable peakedness factor as an additional parameter. When these two or three predicted components are inputted, the wave spectrum can be adjusted for use in spectral analysis. According to J.M.J. Journée (2012), the Bretschneider spectrum, which has been widely accepted since the Second International Ship Structures Congress in 1967, is the most commonly used for seakeeping calculations and model experiments.

2.2.4 Spectral Analysis in OCTOPUS

Applying spectral analysis, OCTOPUS transforms the incoming wave energy spectrum through the vessel's RAO to produce a response spectrum, as described in Figure 9. With this response spectrum, maximum roll and pitch motions can be derived.

2.2.5 Predicting Maximum Roll and Pitch Motions in OCTOPUS

OCTOPUS is designed to predict the maximum roll and pitch motions, rather than providing a continuous series of motion values. This is especially important for HTVs, as extreme roll and pitch motions can cause significant dynamic stresses on welded cargo and structural components, which can have a major impact on the vessel's structural integrity and operational safety. The focus on maximum motions is also due to the nature of the available forecast data, which usually consists of statistical values such as significant wave height and peak wave period, rather than continuous wave data. Therefore, predicting a continuous response is not feasible and is not as useful for operational decision making. OCTOPUS uses the energy response spectrum in combination with the Rayleigh distribution to create a probability density function that allows it to extract the statistical maximum and minimum values for vessel motions. This method involves determining the standard deviation of the probability density function to calculate the significant wave height ($H_{1/3} = 4\sigma$). The maximum roll motion, which is essential to assess the stability and safety of the vessel, is estimated using the relationship described in Equation 2.13.

$$\exp\left\{-2\left(\frac{\phi_{\max}}{\phi_{1/3}}\right)^2\right\} = \frac{1}{1000} \quad \text{or} \quad \phi_{\max} = 1.86 \cdot \phi_{1/3} \tag{2.13}$$

2.2.6 Challenges with OCTOPUS

The main reliance of OCTOPUS is on linear theory, which restricts its capability to simulate intricate real-world maritime conditions. This constraint becomes especially apparent when examining phenomena like roll damping and the nonlinear interconnection of roll and pitch movements. These aspects, crucial for comprehending the behavior of vessels in maritime environments, are mostly nonlinear and, therefore, inadequately accounted for by linear theories. However, it is important to first understand the inherent differences between roll and pitch motions before delving into the specifics.

Differential Dynamics of Roll and Pitch Motions

Understanding the inherent differences between roll and pitch motions is essential to understand the dynamics of maritime vessels. Pitch motion is heavily influenced by radiation damping. Due to this substantial damping influence, pitch motions are typically well regulated and exhibit less susceptibility to dynamic amplification. Conversely, roll motion is characterized by a distinct sensitivity to resonance. In roll motions, radiation damping plays a less significant role compared to pitch. As a result, other factors, such as viscous damping, which involves energy dissipation due to internal fluid friction and boundary layer effects around the vessel's hull, become more prominent. However, even with these damping forces, the overall resistance to roll motion is considerably lower compared to pitch. This lower resistance to roll motion makes it particularly susceptible to resonance effects. When the natural

roll periods of vessels align with certain wave frequency ranges, it can lead to a phenomenon known as dynamic amplification. This effect is notably more pronounced in roll motions than in pitch due to the inherent differences in damping characteristics. Dynamic amplification in roll can result in significantly increased motion amplitudes, posing challenges for linear theoretical approaches such as OCTOPUS that may not fully capture these nonlinear behaviors (Journée and Massie, 2008).

In summary, while pitch motion is predominantly governed by a damping-dominated regime, leading to relatively stable behavior, roll motion experiences a more complex interplay of damping forces and is highly sensitive to resonance. This fundamental difference underlines the need for advanced, often nonlinear modeling techniques to accurately predict and manage the behavior of maritime vessels in varying sea conditions.

Roll Damping

According to Chakrabarti (2001), the damping of large floating structures in waves is calculated using linear diffraction/radiation theory. In the analysis of vessel dynamics, damping is an important factor and comprises various components. For most degrees of freedom of a vessel, such as heave, sway, surge, and pitch, radiation damping, which is one component of the total damping, plays a dominant role and is often sufficient to describe the vessel's response. However, this is not particularly true for the roll motion of a long floating structure, where the roll damping exhibits highly nonlinear characteristics (Chakrabarti, 2001).

In the case of roll motion, the radiation damping is generally quite small compared to the total damping in the system. Viscous damping becomes an important factor in the restoration of equilibrium, particularly in roll motion, where wave damping is comparatively lower (Journée and Massie, 2008). The Ikeda method, proposed by Ikeda et al. (1978), is essential in estimating viscous roll damping. This method takes into account variables such as forward vessel speed, eddy making, and bilge keels, and approximates these nonlinear forces through an equivalent linear damping coefficient. This integration improves the accuracy of roll motion predictions using OCTOPUS.

Roll Damping Coefficients and Superposition

The roll-damping coefficients for the form of the hull of a vessel have several contributions, each reflecting a different aspect of the interaction between the vessel and the water. These components are:

1. **Skin Friction**: This arises from the frictional resistance between the hull surface and the water as the vessel rolls. Skin friction contributes to damping, as it dissipates energy through the viscous shear stresses developed on the hull surface. In the case of forward speed, the skin friction increases slightly, affecting the overall damping.

2. **Eddy Shedding**: Eddy shedding is a form of damping that occurs due to vortices generated by flow separation, particularly at sharp corners of the hull. This type of damping is more pronounced near the bow and stern of slender vessels, where the flow separation is more significant. Eddy shedding contributes to roll damping by dissipating energy through the formation and dissipation of these vortices.

3. **Wave Damping**: Wave damping is associated with the energy lost in creating free surface waves as the vessel rolls. This component, often referred to as radiation damping, is vital in the overall damping mechanism. Although it is more prominent in other motions, in the roll motion, it forms a smaller yet significant part of the total damping. Figure 11 below provides a quantitative visualization of wave damping as a function of the width-draft ratio of a vessel. The curve illustrates the tendency for wave damping to vary inversely with the vessel's breadth-draft ratio, showcasing the nuanced interplay between vessel geometry and damping characteristics.

4. **Lift Effect Damping**: Lift damping in roll is analogous to the lift force experienced by a ship moving forward with sway motion. It occurs in the form of a lift moment that increases with the ship's forward speed. At higher speeds, the contribution of the lift effect damping to the total roll damping becomes more substantial.

5. **Bilge Keel Damping**: The presence of bilge keels on a vessel introduces additional damping. This damping arises from the normal forces acting on the keel and the pressure variations on the hull surface as the vessel rolls. The bilge keel damping is nonlinear and can significantly contribute to the overall damping, especially in heavy seas.

Derived from model testing data, primarily from 2-dimensional models, these expressions isolate components of damping and later integrate them using the principle of superposition. Although methodical, this approach often overlooks the interdependent interactions, which Chakrabarti (2001) emphasizes as potentially significant. Thus, while OCTOPUS incorporates methodologies like those of Ikeda et al. (1978) for estimating roll damping, it continues to grapple with inherent assumptions and limitations. Reflecting on the overarching reliance on the superposition principle in linear and potential theory, a consistent theme is observed: Both the damping terms and the total wave force are assumed to be the cumulative sum of individual components. This principle, effective in linear systems, becomes questionable amidst nonlinear interactions, as evident in the phenomena that are discussed. In the real maritime world, the wave forces on a vessel are not merely a linear summation, but engage in complex interdependent interactions.

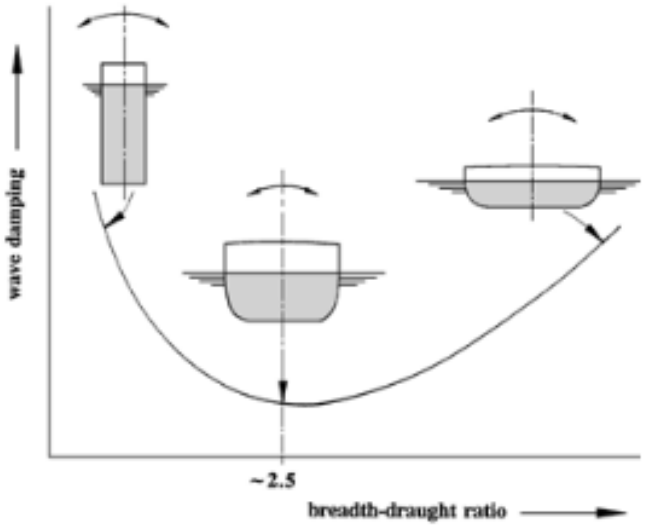


Figure 11: Relation Between the Roll Damping and Breadth-Draft Ratio of a Vessel's Cross-section. Adapted from Journée, J., & Massie, W. (2008). *Offshore hydromechanics*. Delft University of Technology

.

Nonlinear Coupling of Roll and Pitch

The nonlinear coupling of roll and pitch motions also impacts vessel behavior in waves, as indicated by the foundational work of Nayfeh et al. (1973). This phenomenon can result in large-amplitude responses in one mode due to excitation in another, particularly under conditions of two-to-one internal or autoparametric resonance, where the natural frequency of pitch is twice that of roll Nayfeh et al., 1973. Such a resonance can lead to complex dynamic behaviors, including the saturation phenomenon, where increases in pitch motion eventually stabilize, and roll motion begins to dominate.

2.3 Machine Learning for Regression Problems

As discussed in the Introduction (see Chapter 1), ML could potentially offer key advantages over conventional approaches like OCTOPUS in dealing with nonlinear phenomena. Unlike conventional approaches, ML-based approaches are not bound by linear assumptions, theoretical models, or fixed rules and relationships. Instead, they learn from data that reflect the complexities of real-world maritime environments, naturally encompassing nonlinear aspects. This practical approach enables ML-based approaches to adjust and potentially provide more accurate predictions in scenarios filled with complex nonlinear interactions. By analyzing actual vessel motion data, ML-based approaches have the potential to uncover patterns and connections that conventional approaches might overlook, possibly leading to improved predictions in terms of accuracy and reliability.

This section provides an overview of the fundamentals of ML, beginning with an exploration of ML and regressions in Subsections 2.3.1 and 2.3.2. Subsection 2.3.3 then delves into the architecture of the plain-vanilla neural network, followed by an explanation of the train, validation and test data in Subsection 2.3.4. Subsection 2.3.5 discusses learning methodologies and techniques, and Subsection 2.3.6 introduces multitask deep learning.

2.3.1 Introduction to Machine Learning

ML has a significant impact on both daily activities and complex tasks. It allows computer systems to identify patterns in the data and make decisions with minimal human involvement. According to the Oxford Dictionary, ML can be defined as "the use and development of computer systems that are able to learn and adapt without following explicit instructions, by using algorithms and statistical models to analyze and draw inferences from patterns in data". In the context of vessel motion, ML could potentially play a role in predicting such motion.

Rebala et al. (2019) categorize ML into two primary types: supervised and unsupervised learning. Supervised learning, with its ability to utilize labeled data for model training, is directly applicable to the prediction of vessel motion. This approach involves models that learn the relationship between input features and a known output, the target variable (for example, a maximum roll or pitch motion). Once trained, these models can then predict the target for new, unseen inputs. In the realm of supervised learning, there are various models for regression tasks that predict continuous values. While logistic regression is used for classification, more complex models, such as neural networks, are used for their ability to capture nonlinear relationships, making them better suited for continuous outcomes like vessel motion. This chapter introduces the application of sophisticated supervised learning models, including neural networks, in the prediction of vessel motion. These models are crucial for generating accurate and reliable predictions necessary for safe maritime operations.

2.3.2 Introduction to Regressions

Understanding regression in the context of ML is critical to understanding the complexities of predicting vessel motion. Classical regression models, typically used for inference, contrast with ML models, which emphasize prediction. ML regression is particularly suitable for large datasets where nonlinear interactions are present, as in the case of vessel motions (Bzdok et al., 2018).

Basic Linear Regression

Basic linear regression provides a basic understanding of how models can be trained to predict outcomes. As described by Rebala et al. (2019), it assumes a linear relationship between the features and the target variable. However, in the context of vessel motion, where the goal is to predict multiple

related outcomes, such as roll and pitch, the simplicity of linear regression may not suffice due to the complex and nonlinear interactions between these motions.

Multi-Target Regression

In the domain of vessel motion prediction, multitarget regression models (MTRs) are valuable. These models are designed to predict multiple related outcomes simultaneously. For example, predicting roll and pitch motions of a vessel at the same time is a multi-target problem. By considering the correlation between these motions, MTRs can provide a more holistic understanding of vessel behavior in waves.

Artificial neural networks (ANNs), which are capable of capturing complex, nonlinear relationships, offer a potent solution to MTR problems. Unlike linear regression models that may be limited to individual target predictions, ANNs excel in scenarios where multiple outcomes are interconnected, as is often the case with vessel dynamics (Nielsen, 2015, Bishop, 2016).

2.3.3 Plain Vanilla Neural Network

Delving into plain vanilla neural networks is the next step in exploring the capabilities of ANNs for vessel motion prediction. These networks form the basis for the more advanced neural network architectures used in this research. With their ability to learn from data and approximate complex functions, ANNs are particularly adept at dealing with the nonlinearity and multifaceted nature of vessel motions (Abiodun et al., 2018).

Architecture of a Plain Vanilla Neural Network

Nielsen (2015) offers an insightful exposition on the architecture of plain vanilla neural networks. This section draws upon Nielsen (2015)'s work to introduce the neural network's architecture and parameters, while avoiding deep mathematical derivations. For a detailed mathematical explanation, the reader may refer to Nielsen (2015).

Imagine each neuron as a node within a network. The columns of these nodes represent different layers in the network. These layers are interconnected through links known as 'weights', as shown in Figure 12. The input layer comprises neurons that represent the values of the input features. For instance, in a neural network for vessel motion prediction, these features might include the vessel's mass, natural frequency, significant wave height, peak period, and more. These feature values are also termed neuron activation values. The output layer, situated at the opposite end of the network, may represent, for example, the vessel's roll motion. In between the input and output layers are the hidden layers, processing initial activation values from the input layer and transforming them into the predicted output values. The activation value of each neuron in a layer is represented as $a_i^{(j)}$, where i and j denote the neuron index within the layer (ranging from 0 to $n-1$) and the layer index (ranging from 1 to L), respectively.

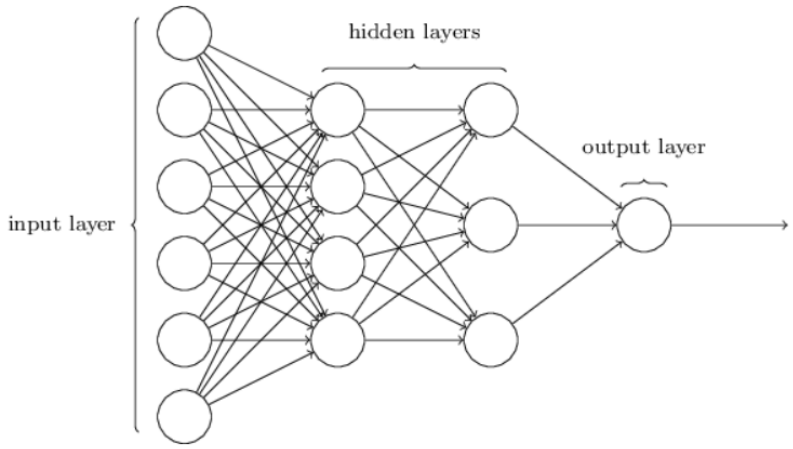


Figure 12: Architecture of a Plain Vanilla Neural Network containing Multiple Hidden Layers. Adapted from Nielsen, M. A. (2015, January). *Neural networks and deep learning.* Determination Press. http://neuralnetworksanddeeplearning.com/

While the design of input and output layers is relatively straightforward, the formulation of hidden layers is more complex. Nielsen (2015) notes that numerous design approaches have been developed for these layers, aiming to optimize the neural network's performance. There is no one fits all solution. Every problem or task requires its specific architecture.

Two crucial concepts in neural network design are the bias and the activation function of neurons. In the architecture depicted in Figure 12, some neurons might be inactive and therefore do not influence the subsequent layer. This is governed by the specific formulation of the activation function.

Bias and Activation Functions

In neural networks, the bias functions similarly to the intercept in a linear equation, enabling neurons to output an activation value even when all input values are zero. This feature increases the adaptability of the network in learning (Nielsen, 2015).

Activation functions transform the sum of weighted inputs and bias into the neuron's output signal, introducing non-linearity. Different activation functions are available, with the sigmoid function, as used by Nielsen (2015), being an example. It maps any real value to a range between 0 and 1.

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{2.14}$$

Rectified Linear Unit (ReLU) is another commonly used activation function. It sets negative input values to zero, thus activating only for positive inputs.

$$\text{ReLU}(x) = \max(0, x) \tag{2.15}$$

ReLU is appreciated for its simplicity and efficiency while still offering non-linear output.

Calculation of Activation Values

The activation of neurons in the input layer is indicated by the vector $\vec{a^0} = [a_0^{(0)}, a_1^{(0)}, \ldots, a_n^{(0)}]^T$. For the first neuron in the first hidden layer, its activation, $a_0^{(1)}$, is computed by summing the products of the neuron activations of the previous layer and their respective weights. Adding the bias and applying the sigmoid function leads to the following.

$$a_0^{(1)} = \sigma(\sum_{i=0}^{n-1} w_{0,i}^{(1)} \cdot a_i^{(0)} + b_0^{(1)}) \tag{2.16}$$

Nielsen (2015) presents a vectorized version for more efficient computation:

$$a^l = \sigma(\bar{W}^l a^{l-1} + b^l) \tag{2.17}$$

Here, a^l and b^l are the activation and bias vectors for each layer's neurons, and $\bar{W}^l$ is the weight matrix for layer l.

Cost Function

In a neural network, the final output is determined by the activation values of the neurons in the last layer. As shown in Figure 12, the network consists of an input layer, two hidden layers, and an output layer, which in this case is a single neuron layer. The number of neurons in the output layer and, hence, the dimension of the output, depends on the nature of the problem. For the illustrated network, the activation of the output layer is represented by $a^{(3)}$, also referred to as y^p in this thesis, indicating the predicted value of the network. The actual desired outcome, known as the target variable, is denoted by y^t, representing the ideal prediction of the network. The objective is to reduce the discrepancy between the predicted values y^p and the target values y^t. One common way to measure this discrepancy is by using the Mean Squared Error (MSE) function, as discussed by Nielsen (2015):

$$C(w, b) = \frac{1}{N} \sum_x \|y^t(x) - y^p(x, w, b)\|^2 \tag{2.18}$$

The function is inherently nonnegative due to the squaring of terms and aims to minimize the difference between predicted and target values. Custom loss functions, tailored to specific problems, can also be effective. For example, Masud et al. (2022) increased the accuracy of their model in identifying

the stages of diabetic retinopathy by modifying the loss function (Masud et al., 2022). Fernando and Tsokos (2022) developed a dynamic loss function that adjusted weights based on class frequency and prediction certainty, helping to balance learning between classes (Fernando and Tsokos, 2022). Furthermore, Liu et al. (2022) demonstrated that customizing the loss function in physics-informed neural networks leads to better performance in solving complex equations (Liu et al., 2022). Additionally, Oktavian et al. (2022) found that a specialized loss function improved Alzheimer's diagnosis from brain scans, particularly with unbalanced data (Oktavian et al., 2022).

Gradient Descent

Nielsen (2015) emphasizes in his book the significant complexity and computational expense of finding the minimum of a function in multivariable problems using calculus alone, as the number of input variables increases. This challenge is particularly present in artificial neural networks, which often involve numerous variables, necessitating an alternative method, namely *gradient descent*. To deconstruct this term, the *gradient* represents a vector indicating the direction of the greatest change of a variable. This gradient vector is obtained by calculating the partial derivatives with respect to all parameters. For example, in Equation 2.18, summing the terms for each training example x eliminates the dependency on x, leaving the cost function C dependent solely on the weights w and biases b. The gradient vector for C is expressed as:

$$\nabla C = \left[\frac{\partial C}{\partial w_{i,j}^{(1)}}, \frac{\partial C}{\partial b_i^{(1)}}, \ldots, \frac{\partial C}{\partial w_{i,j}^{(L)}}, \frac{\partial C}{\partial b_i^{(L)}} \right]^T . \tag{2.19}$$

To illustrate the concept of *descent*, Nielsen (2015) employs a graphical representation (see Figure 13). Consider a cost function dependent on variables v_1 and v_2, with a green ball marking an initial position in this plot. This position is defined by the vector v that extends from point O to the green ball. The objective of gradient descent is to find the minimum of the cost function. At the location of the green ball, the gradient vector is calculated to identify the direction of the greatest change. Reversing this vector by multiplying it by -1 indicates the direction of the steepest descent. The size of each step taken in this direction is determined by the learning rate η. A high η may cause overshooting and prevent convergence, while a very low η results in slow progress. Equation 2.20 demonstrates one iteration of the gradient descent algorithm:

$$v \to v' = v - \eta \nabla C. \tag{2.20}$$

Nielsen (2015) further elaborates on applying gradient descent in neural networks, suggesting its

use in minimizing the cost function through adjustment of weights w_k and biases b_l (Equation 2.21):

$$\begin{aligned} w_k \rightarrow w_k' &= w_k - \eta \frac{\partial C}{\partial w_k} \\ b_l \rightarrow b_l' &= b_l - \eta \frac{\partial C}{\partial b_l}. \end{aligned} \tag{2.21}$$

By cyclically applying this update rule and processing data through the network (referred to as epochs, where one epoch equates to processing the entire data set once), a (local) minimum of the cost function can eventually be achieved.

The gradient descent algorithm, particularly with an increasing number of input features and training examples, may demonstrate slow learning due to lengthy gradient computation over the entire data set per epoch. An evolved version, *stochastic gradient descent*, mitigates this by using mini-batches of randomly selected training examples instead of the full data set, as discussed by Nielsen (2015). Although this method may occasionally deviate from the optimal path, its overall efficiency in approaching the minimum is generally faster than traditional gradient descent.

The key parameters in this process - learning rate, minibatch size, and epoch number – are known as *hyperparameters* of the neural network (Nielsen, 2015). The gradient of the cost function ∇C and subsequent adjustment of the model parameters are facilitated by a technique called *backpropagation*. Figure 13 provides a visual representation of gradient descent, adapted from Nielsen (2015).

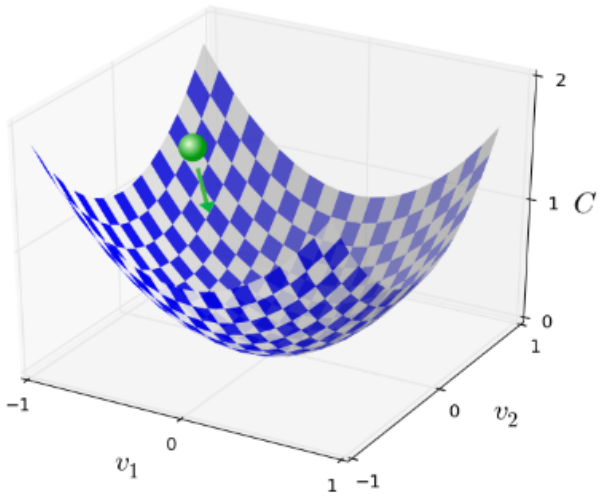


Figure 13: Gradient Descent Illustration. Adapted from Nielsen, M. A. (2015, January). *Neural networks and deep learning*. Determination Press. http://neuralnetworksanddeeplearning.com/

Backpropagation

According to Nielsen (2015), the backpropagation algorithm is vital for learning in neural networks. It efficiently computes the gradient of the cost function, revealing how changes in weights and biases affect the cost. The algorithm's foundation involves an 'error' variable, defined in Equation 2.22:

$$\delta_j^l = \frac{\partial C}{\partial z_j^l}. \tag{2.22}$$

Nielsen (2015) demonstrates that with this error formulation, the gradient of the cost function relative to weights and biases can be determined (Equation 2.23):

$$\frac{\partial C}{\partial b_j^l} = \delta_j^l, \quad \frac{\partial C}{\partial w_{jk}^l} = a_k^{l-1}\delta_j^l. \tag{2.23}$$

The backpropagation algorithm unfolds in the following steps:

1. **Input:** Introduce the network to a mini-batch from a specific epoch, containing a set of training examples.
2. **Loop:** For each training example x:
   (a) **Feedforward:** Use Equation 2.17 to calculate activation values across layers.
   (b) **Output error:** Determine the error in the final layer L.
   (c) **Backpropagate error:** Apply this error to propagate backwards through the network (see Nielsen (2015) for more details).
3. **Gradient descent:** Utilize the gradients derived from Equations 2.23 to update weights and biases.

This algorithm has earned the name 'backpropagation' because the error is calculated in a reverse manner - it begins at the final layer (the output) of the network and then proceeds backward through the preceding layers. Upon completion of the last step in this algorithm, a new mini batch containing several training examples can be initiated, starting once more from the initial step. This method effectively links minor changes in weights or biases with the corresponding shifts in the cost function (Nielsen, 2015). After having processed all the mini batches for one epoch, the next epoch starts with the improved weights and biases. The point at which the algorithm stops will be further discussed in Subsection 2.3.5, providing deeper insight into a ML model's learning process. Before that, the necessity of data splitting will be addressed.

2.3.4 Train, Validation, and Test Data

Utilizing all data exclusively for training is not feasible, as it leaves no means to evaluate the model's performance on novel, real-world data. The standard approach involves dividing the data into three distinct segments:

1. **Training data:** Used for updating the model's weights and biases, this dataset is central to the training process, and used in the backpropagation algorithm.

2. **Validation data:** Used to assess the performance of the model during its training phase. The model is evaluated on 'unseen' data at this stage, enhancing its ability to perform on new data post-training. Adjustments to the model's hyperparameters are made based on its performance with this dataset.

3. **Test data:** Used to evaluate the actual performance of the model on entirely unseen data. Since the hyperparameters are fine-tuned based on the validation set, there is a potential bias toward these data. Testing the model on a completely new dataset, the test dataset, is thus crucial.

Cross Validation

The test data is extracted before the model training and excluded from the training process. However, the division of the remaining data into training and validation sets can be approached in various ways. Cross-validation (CV) is a widely used method that partitions the data into k distinct folds. In each iteration, one fold is withheld from training, serving as the validation set, while the remaining folds are included. This process is cycled through until each fold has functioned as the validation set. The performance of the model is typically assessed based on the average results across all folds, offering a comprehensive view of its effectiveness on k different 'unseen' datasets. Cross-validation, a longstanding method in data division, continues to be a preferred option among researchers (Xu and Goodacre, 2018).

2.3.5 Learning Methodologies and Techniques

As introduced in Section 2.3.3, the stochastic gradient descent and backpropagation processes are important in neural networks for pattern recognition. Specifically, during the training phase, the backpropagation algorithm (see Subsection 2.3.3) iterates numerous times, each identified as an epoch. After each epoch, the model evaluates its performance using the loss function. The gradient descent method then adjusts the weights and biases to minimize the loss function for the training data. However, to avoid overfitting, it is crucial to validate the model with unseen data.

Addressing Overfitting

Nielsen (2015) defines overfitting as a situation where a model demonstrates high accuracy in training data but performs poorly on new, unseen data. As neural networks are typically applied to new data, overfitting is undesirable and undermines their practical value. According to Nielsen (2015), overfitting is particularly problematic in modern neural networks, which often have numerous parameters. To mitigate overfitting, he recommends techniques that either stop the learning process early or prevent the model from easily learning from the training data. This section discusses two such techniques: early stopping and regularization.

Early Stopping The early stopping method ceases the learning process at a predetermined point. It involves monitoring both training and validation losses: the former is based on the training data, and the latter on the validation data. The validation data informs the decision on when to stop training by comparing the model's predictions against the actual values. Early stopping is activated when the validation loss ceases to decrease. However, to avoid premature stopping, a patience parameter is used, defined as the number of epochs to wait before stopping. Additionally, an early stopping threshold can be set, imposing a minimum required improvement in the validation loss to prevent early termination.

Regularization Techniques Nielsen (2015) observes that regularized neural networks generally generalize better than unregularized ones, although he notes that this understanding is based on heuristic principles rather than systematic analysis. This section highlights two commonly used regularization techniques.

- **Weight Decay**: Also known as L2 regularization, this technique adds a term, $\frac{\lambda}{2n} \sum_w w^2$, to the cost function (see Equation 2.18). It encourages the network to learn using smaller weights, thus reducing the sensitivity to data noise and helping in generalization (Nielsen, 2015).
- **Dropout**: Unlike weight decay, dropout adjusts the structure of the neural network by randomly omitting neurons during training. This reduction in dependency on specific neurons mitigates overfitting and essentially trains multiple models on the same dataset, averaging out individual model biases (Nielsen, 2015).

Adaptive Learning Techniques

Another approach to enhance generalization is the use of adaptive learning techniques, which adjust the learning rate based on progress. As Moreira and Fiesler (1995) states, a large initial learning rate facilitates rapid learning but can lead to overshooting the local minimum. To avoid this, the learning rate should decrease as the model approaches the minimum. This adjustment is typically based on a set patience in epochs, after which the learning rate reduces if there is no decrease in validation loss, with a common reduction factor of 0.1 (Moreira and Fiesler, 1995).

2.3.6 Multi-Task Deep Learning

The previous discussion about the plain vanilla neural network, as depicted in Figure 12, focused primarily on a network structure with a single output neuron. However, in multitask learning, the structure evolves to accommodate the prediction of multiple related tasks, such as the roll and pitch motions of a vessel. This approach diverges from the single-output architecture by incorporating multiple neurons in the output layer, each neuron representing a distinct task.

In the context of vessel motion prediction, multitask learning is beneficial. The shared layers in the multitask network architecture, while handling the same input data, can learn common patterns relevant to both roll and pitch motions. The task-specific layers following these shared layers then focus on the unique characteristics and nuances of each motion. This design allows for a more comprehensive analysis that captures the intricate relationship between roll and pitch, which could be important for accurate maritime predictions.

Deep neural networks in multitask learning retain the basic structure of input and output layers, but are 'deep' due to their multiple hidden layers, as mentioned by Thung and Wee (2018). These deep structures are more adept at handling complex relationships, such as those found in vessel motion. Figure 14 presents an abstract model of multitask deep learning. Thung and Wee (2018) emphasize that the success of multitask deep learning models hinges on the relatedness of the tasks. The simultaneous learning of roll and pitch motions, which are inherently interconnected but distinct aspects of vessel dynamics, may exemplify an ideal application for this architecture.

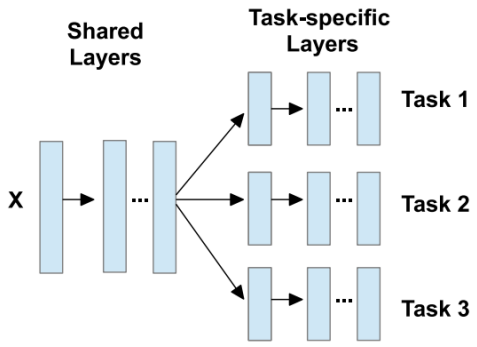


Figure 14: Abstract representation of a multi-task deep learning model, suitable for tasks like vessel motion prediction. Adapted from Thung, K.-H., & Wee, C.-Y. (2018). A brief review on multi-task learning. *Multimedia Tools and Applications*, *77*, 29705–29725. https://doi.org/10.1007/s11042-018-6463-x

2.4 Comparative Review of Vessel Motion Prediction Approaches

This section provides an overview of the literature on applications of vessel motion prediction approaches. It starts with a look at conventional prediction approaches and their background in Subsection 2.4.1. It then moves on to ML based approaches in the offshore industry, particularly focusing on vessel motion prediction in Subsection 2.4.2. Finally, this section concludes by discussing the implications of these studies for the current research in Subsection 2.4.3.

2.4.1 Overview of Conventional Prediction Approaches

Historically, various approaches for predicting vessel motion based on conventional linear theories have been used. Introduced by Salvesen et al. (1970) in 1970, the strip theory used by OCTOPUS (see Section 2.2) has been widely adopted by numerous institutes (K. A. McTaggart, 2011). In 1997, an updated strip theory program, SHIPMO7, was launched, enhancing the original work of Salvesen by including viscous forces in lateral plane motion evaluations and introducing novel boundary methods to compute hydrodynamic coefficients, addressing irregular frequency problems (K. McTaggart et al., 1997). A decade later, in 2007, K. McTaggart (2007) unveiled ShipMo3D, a new vessel motion prediction system. Unlike its predecessors that relied on strip theory, ShipMo3D employed the 3D panel method, allowing problem solving in both frequency and time domains (K. McTaggart, 2007). In 2011, further advances in ShipMo3D were acknowledged, particularly in its ability to predict vessel motions using the Green function for zero forward speed. However, the challenges in predicting roll motion due to viscous effects remained a concern (K. A. McTaggart, 2011). The most recent iteration, ShipMo3D version 4.3, continues to use diffraction theory, with improvements such as a more representative effective roll damping factor and improved modeling of bilge keels, skegs, and hull eddy damping (K. McTaggart, 2021). In 2018, MO4, a maritime engineering office, claimed to offer the most advanced motion forecasting system available (MO4, 2018), and is currently used by Boskalis' heavy lifting department. Similarly to ShipMo3D, MO4 employs a 3D diffraction approach for the calculation of hydrodynamic coefficients, crucial for the calculation of RAO (see Section 2.2).

These conventional approaches predominantly rely on strip theory or the panel method for the determination of hydrodynamic coefficients. As early as 1991, thung and Huang (1991) suggested that the 3D panel method, despite being computationally less efficient, offered superior results to strip theory, especially at higher Froude numbers. Contrasting this view, a more recent study by Karola et al. (2022) in 2022 found minimal differences between the strip theory and the panel method on a global scale, with notable variations in small roll motions and local loads near the bow. They concluded that incorporating forward speed effects did not significantly alter the results in either method (Karola et al., 2022).

2.4.2 Overview of Machine Learning Approaches

The offshore community has increasingly adopted ML techniques and applies them in various domains. Examples include conditional monitoring of offshore wind farms, life extension assessments for wind support structures, prediction of significant wave heights in China, and help to decommission oil and gas platforms (Black et al., 2021; Feng et al., 2022; Vuttipittayamongkol et al., 2021; Yeter et al., 2023).

In the realm of vessel motion prediction, neural networks have been the focus of several research initiatives. Skulstad et al. (2021) investigated a hybrid approach, combining ML with conventional approaches, to improve vessel motion prediction during docking. The integration of an ML-based approach showed improved performance compared to the conventional approach. The goal was to predict vessel movements 30 seconds ahead, augmenting, rather than replacing, the existing onboard approach. Other researchers, like Khan et al. (2005), focused solely on ML-based approaches, assessing the efficacy of Artificial Neural Networks (ANNs) in real-time vessel motion prediction. Their findings highlighted the ability of ANNs to accurately predict vessel movements up to 10 seconds in advance. Additionally, Jiang et al. (2021) proposed using a long-short-term memory (LSTM) neural network for ship maneuvering predictions, emphasizing its proficiency in handling non-linear relationships, although recommending further testing with real data. Li et al. (2017) developed a two-layer recurrent neural network (RNN) model, demonstrating high accuracy in vessel motion predictions, with the intention of including environmental factors in future enhancements.

Comparing conventional and ML-based approaches reveals both similarities and differences. Both methods face challenges in accurately predicting roll motions compared to pitch motions (K. A. McTaggart, 2011; Silva and Maki, 2022). In addition, the accuracy of both approaches is heavily influenced by the quality of weather forecasts. However, a significant difference lies in the handling of nonlinear effects. Neural networks excel at processing nonlinear phenomena. Traditional methods rely on linear and potential theories, incorporating nonlinearities through equivalent terms.

Rebala et al. (2019) describe the fundamental distinction between these approaches as follows: conventional approaches hypothesize relationships based on domain knowledge, which are tested experimentally, whereas ML-based approaches discover relationships through data analysis without predefined assumptions. This is exemplified in Equation 2.24:

$$y = f(x) \tag{2.24}$$

In conventional modeling, a specific function $f(x)$ is assumed to transform x to y, with this assumption tested against measured output. On the contrary, ML regards the function $f(x)$ as an unknown, learned through large data sets that represent y with corresponding input parameters x. ML provides

a fresh perspective in predictive modeling, particularly in fields where processes and outcomes are not fully understood (Willard et al., 2022).

2.4.3 Implications for Study

This review of the literature suggests several implications for the current study. First, the successful application of ML in predicting statistical values, such as the standard deviation of vessel responses (Silva and Maki, 2022), implies the feasibility of predicting maximum roll and pitch motions. Second, the prevalent use of neural networks in prior vessel motion prediction studies guides this research toward designing a neural network-based architecture. Lastly, given that both conventional and ML-based methods show difficulty in accurately predicting roll motion, this study may benefit from employing multitask learning, allowing for varied complexity between tasks.

In addition, the gaps identified in the literature include the absence of comparative studies between ML-based and conventional approaches. Although previous studies have validated ML methods with real-world data, this study aims to compare their error rates with those of conventional approaches. Moreover, most previous studies focused on nowcasting, predicting short-term future scenarios. On the contrary, this study seeks to develop an ML-based model capable of forecasting several hours into the future. This aligns with the operational relevance of predicting statistical maximum values for Boskalis, the primary focus of the study. The next chapter will detail the applied methodology.

3. Methodology

This chapter outlines the methodological framework employed in this research to evaluate and compare the effectiveness of ML approaches against (simulated) conventional approaches in predicting the maximum roll and pitch motions of the Target. Three new prediction approaches are developed and compared with OCTOPUS using two distinct validation strategies to validate both known LCs and unknown LCs. The new approaches are trained and validated using sensor data and precalculated LC data from 24 different voyages that the HTV Target undertook, and differ in their design and the type of environmental data they use. Moreover, the adaptability of ML-based approaches to the encounter of unknown LCs is explored. Two datasets are used to evaluate the validation approaches and adaptability: the significant dataset, which includes the highest one third of the observed response data points in terms of magnitude, and the general dataset, which includes the entire range of data points. The methodology chapter first outlines the prediction approaches—NauticalNet, NauticalNet*, and HindHold—followed by sections on validation strategies and the adaptability analysis of ML approaches to unknown LCs. In general, this chapter aims to answer the question of how this research was conducted. The subsequent chapter, 'Execution,' delves deeper into several aspects addressed in this chapter by focusing on the question of what was done in this research (see Chapter 4).

3.1 Prediction Approaches

As stated in Chapter 1, this thesis employs a dual data strategy to predict vessel motion. It uses both nowcast and hindcast environmental data. Additionally, a key objective of this research is to assess the effectiveness of ML approaches compared to (simulated) conventional approaches. To achieve this, four potential prediction approaches were initially identified. These approaches were chosen to ensure a comprehensive comparison across all aspects. Specifically, two conventional and two ML-based approaches were developed, each utilizing nowcast or hindcast data, in alignment with the thesis objectives.

OCTOPUS, the preexisting conventional approach, utilizes nowcast environmental data and serves as a baseline for comparison. HindHold, an innovative approach developed during this research, employs ML techniques to simulate how a conventional approach might perform using hindcast data. This was necessary because only the outputs of OCTOPUS were available, not the system itself, which prevented direct application of the hindcast data. Alongside these, two purely ML-based approaches were developed: NauticalNet, which uses hindcast data, and NauticalNet*, which uses nowcast environmen-

tal data. It is important to note that all four approaches, while varying in their use of environmental data and underlying methodologies, rely on the same navigational data, including vessel speed and location, and LC data, which encompass general hydrostatic data per voyage. Figure 15 provides a schematic overview of how these approaches interact with the observed vessel motions, setting the stage for the comparative analysis. Additionally, Table 1 summarizes these four approaches, describing their design methodologies, the types of environmental data they use, and their predictive purposes. For a detailed explanation of the ML architectures used, including NauticalNet, NauticalNet*, and HindHold, please refer to Chapter 4, 'Execution,' where the configuration and implementation are discussed in depth.

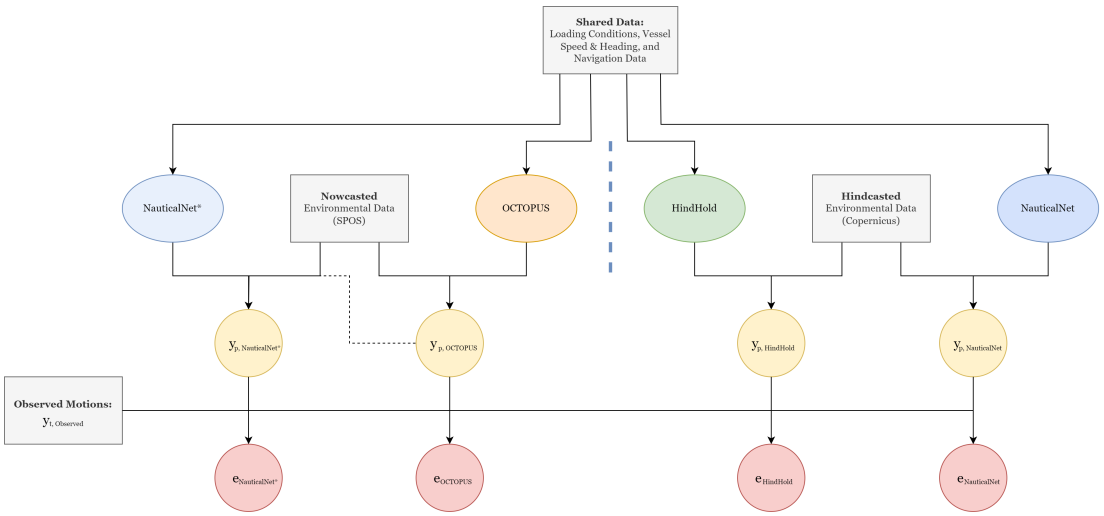


Figure 15: Schematic representation of the prediction approaches and data flow

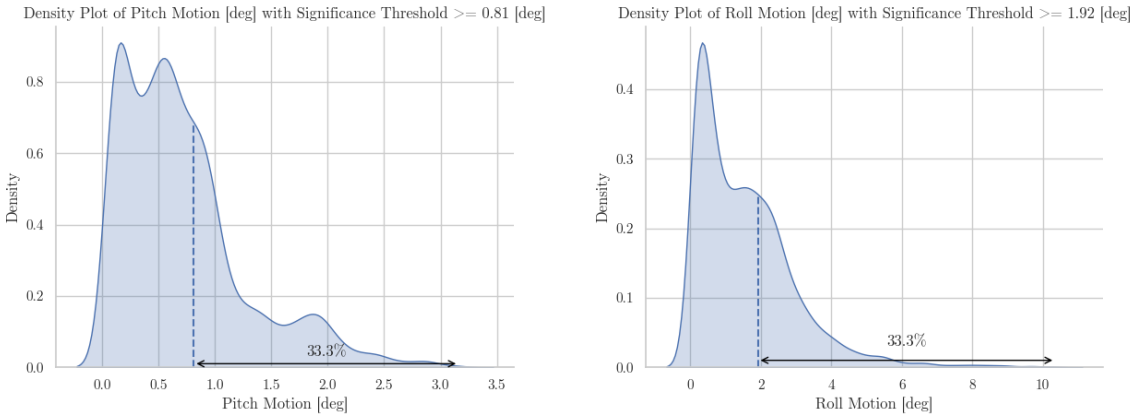
Approach	Design Approach	Data Type	Prediction Goal
OCTOPUS	Conventional (RAO and wave spectrum)	Nowcast Environmental Data	Real World Roll and Pitch Motions
HindHold	Machine Learning trained on OCTOPUS values	Hindcast Environmental Data	OCTOPUS's projections with hindcast data.
NauticalNet	Machine Learning	Hindcast Environmental Data	Real World Roll and Pitch Motions
*NauticalNet**	Machine Learning	Nowcast Environmental Data	Real World Roll and Pitch Motions

Table 1: Comparison of Prediction Approaches

3.2 Validation

Following the delineation of prediction approaches, this section proceeds to detail the validation strategies employed to assess the predictive performance of the compared prediction approaches: OCTOPUS, NauticalNet, NauticalNet*, and HindHold. The dataset comprising the 24 voyages undertaken by the HTV Target is central, as it encompasses all the data used in this study. The predictive accuracy of the four approaches is evaluated based on their mean absolute error (MAE), root mean square error (RMSE), and mean absolute percentage error (MAPE). Furthermore, separately reported overestimations and underestimations indicate tendencies toward over- or underprediction, each with different practical implications. An error exceeding one standard deviation from the actual value, derived from the entire dataset, is classified as an over- or underprediction.

These metrics are applied to two interrelated but distinct datasets: the significant dataset, which includes the highest one-third of the observed response data points in magnitude, and the general dataset, which encompasses the full spectrum of data points (see Figure 16). This distinction is imperative because an error within the most significant one-third of the data points could be more impactful than a general error across the entire dataset. Specifically, for the significant data points, high underpredictions in roll and pitch could unexpectedly increase the stresses and loads on the cargo's joints and within the cargo itself, thereby jeopardizing operational safety and Boskalis's commitment to delivering cargo safely. Conversely, overpredictions within this critical subset may prompt overly cautious decisions by vessel operators, such as route alterations or operational halts, diminishing efficiency, and incurring additional costs for Boskalis. These concerns are primarily relevant for significant data points, as errors associated with lower response magnitudes in less significant data points are unlikely to exert substantial effects on the joints. Consequently, such errors are less likely to affect operational decisions. Therefore, it is essential to evaluate the dataset in this divided manner, with 'significant' data points being those within the top one-third in terms of magnitude, as per the standard practice in the offshore industry.



(a) **(b)**

Figure 16: Data Density and Significant Threshold of the Maximum Roll and Pitch Motions in the Observed Data

In addition, the validation of the prediction approaches includes two distinct strategies. Each of the 24 voyages in the dataset is associated with a unique LC. These strategies are designed to assess the approaches on both known LCs (included in the training of the ML-based approaches) and unknown LCs (excluded from the training of the ML-based approaches), providing a comprehensive assessment of the prediction approaches' performance.

1. **Validation Strategy I:** In this strategy, the first 90% of data from each voyage is used for training the ML-based approaches. As each voyage contributes to the training, each LC is also included in the training data. The remaining 10% of each voyage's data is then used to evaluate all four approaches. The aim is to assess the accuracy of the approaches under known LCs, revealing their potential maximum prediction accuracy.

2. **Validation Strategy II:** This strategy trains ML-based approaches with all available data, except that of a specific voyage and its unique LC. The dataset from the excluded voyage is then used to evaluate the performance of the four prediction approaches. This strategy tests the generalizability of ML-based approaches to an unknown LC, specifically focusing on their predictive capabilities under untrained conditions.

By utilizing both these validation strategies, the goal is to gain deeper insight into the performance of the ML-based approaches and address potential limitations. The second validation strategy adds practical value to this research, as it partially indicates how the developed ML-based approaches would perform in new transport operations for Boskalis. In contrast, the first validation strategy aims to demonstrate the maximum potential of ML by removing one problem from the equation (the generalization across LCs), potentially showcasing what can be achieved in an ideal scenario.

3.3 Adaptability of Machine Learning-Based Prediction Approaches

The adaptability of ML-based prediction approaches to unknown LCs is assessed to explore their potential further. The assumption underpinning this methodology is that the performance of ML-based approaches decreases when encountering unknown LCs, attributed to the limited diversity of unique LCs available for training, which may not be sufficient for effective generalization.

The adaptability is investigated using an iterative training process, beginning with the scenario in which all the data from the voyage, along with its corresponding LC, is excluded, reflecting the second validation strategy. This process is incrementally advanced through subsequent iterations until the framework mirrors the first validation strategy, where 90% of the excluded voyage data is incorporated into the training set, thus connecting both validation strategies. The iterations proceed as follows:

- Iteration 1: NauticalNet and NauticalNet* are trained on data from 23 voyages.
- Iteration 2: Training includes 23 voyages plus a random 5% subset of the excluded 24th voyage.
- Iteration 3: Training encompasses 23 voyages and a random subset of 10% of the excluded 24th voyage.
- ...
- Iteration 19: Training involves 23 voyages plus a random 90% subset of the excluded 24th voyage.

Each iteration incrementally integrates a random subset of the LC data from the excluded voyage. For example, the third iteration incorporates 10% of the data, including 5% utilized in the second iteration. Figure 31 in Chapter 4 'Execution', section 4.3.3, illustrates this methodology, with red data points denoting the training set and yellow points the test data.

4. Execution

This chapter delves into the specific decisions and actions taken during the progress of this research project. While the preceding chapter, 'Methodology' (Chapter 3), outlined the overarching framework and strategies for achieving the thesis objectives, 'Execution' focuses on the practical application of these methods. It explains the specific choices made during the project, thus bridging the gap between theoretical methodology and practical implementation. The importance of this chapter lies in its detailed exposition of the research process. Although not essential for general understanding of the subsequent chapters, it provides valuable information on design decisions and execution steps. This detail is crucial for those interested in replicating or building on this research.

The chapter is organized into several key sections for ease of understanding. Section 4.1 offers a comprehensive look at data preprocessing, including the sources of data and the preparation steps required for the application of ML techniques. Section 4.2 provides an in-depth examination of the development process for the NauticalNet, NauticalNet*, and HindHold approaches. Finally, Section 4.3 outlines the rationale behind the chosen validation choices. In essence, this chapter not only offers a detailed roadmap of the execution process but also ensures that the research is accessible and reproducible, adhering to the standards of thorough academic inquiry.

4.1 Data preprocessing

Data preprocessing is an important initial step in ML techniques, which involves the curation, unification, preparation, and cleaning of data (Ilyas and Chu, 2019). These steps are crucial to improving the efficiency of model learning by simplifying the data (Kotsiantis et al., 2006). This section discusses the data preprocessing procedures for NauticalNet, NauticalNet*, and HindHold, and outlines the methods used to prepare the datasets for training. Data preprocessing in this study can be divided into three main areas: selection, extraction, and merging of data sources; feature selection; and data cleaning. The selection, extraction, and merging of data sources play an important role in determining the volume and variety of data available for this research. Feature selection aims to reduce the dimensionality of the dataset through various steps. Data cleaning ensures that the dataset is ready for training by addressing issues such as outliers, missing data, and scaling. The following steps provide a summary of the data preprocessing process:

1. Data source selection, extraction, and merging: Identify relevant data sources, extracting the necessary data and combining different datasets to form a comprehensive dataset for analysis.

2. Feature selection: Selecting the most relevant features (variables) for the model.
    (a) Transforming features: Modify or create features to improve the interpretability or performance of the model.
    (b) Preventing data leakage: Ensuring that the model does not have access to information it would not have during real-world operation, to avoid overestimating the model's performance.
    (c) Eliminating irrelevant and redundant features: Removing features that do not contribute to the model's predictive power or that duplicate other features' information.
3. Data cleaning: Preparation of the data for analysis by addressing quality issues.
    (a) Managing outliers: Identifying and handling extreme values that could skew the results.
    (b) Handling missing values: Dealing with gaps in the data, either by imputing values or removing incomplete records.
    (c) Scaling the data: Adjusting the range of data features for better model performance and interpretability.

4.1.1 Selection, Extraction, and Merging of Data Sources

Between June 4, 2012, and April 13, 2023, HTV Target embarked on 37 voyages that covered various global seas. These voyages, totaling 1967 days at sea, are shown in Figure 17. These voyages have a rich diversity of data, offering a wide range of navigational, environmental, and motion response variables. In these data, time periods were assigned to certain states in which the main operation performed by the HTV Target was described. In this thesis, only data points corresponding to the 'sailing' state were retained, excluding other states like 'repair and maintenance', 'detachment', and more. This criterion ensured the consistency of LCs on the HTV Target and relevance of data for the study, since the most practical value is achievable during the sailing state.

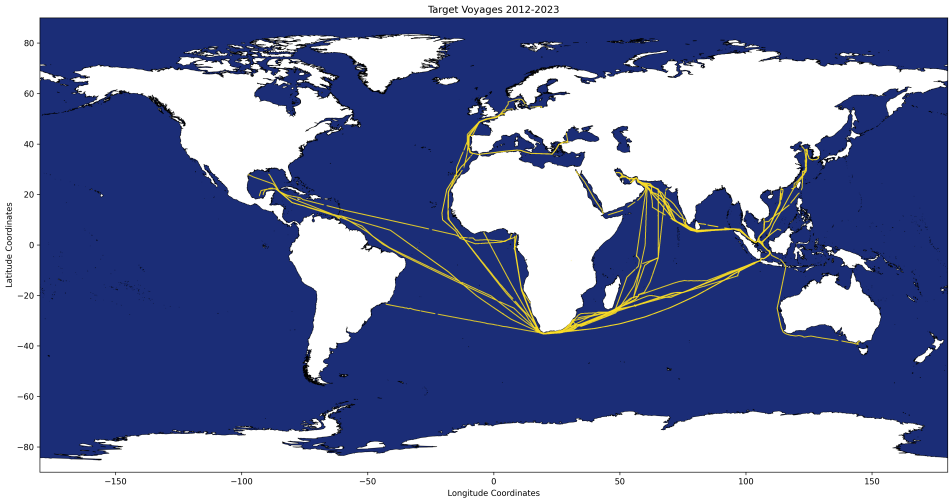
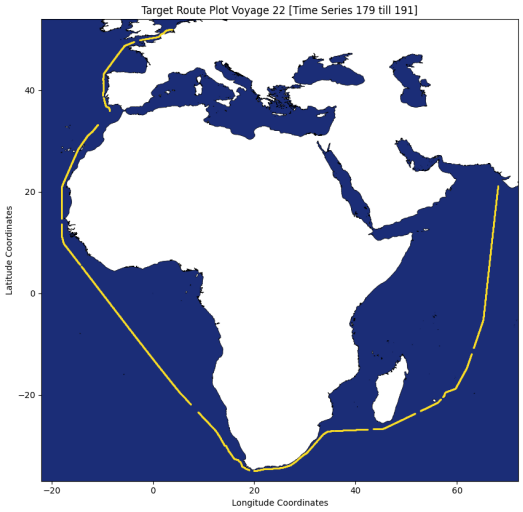


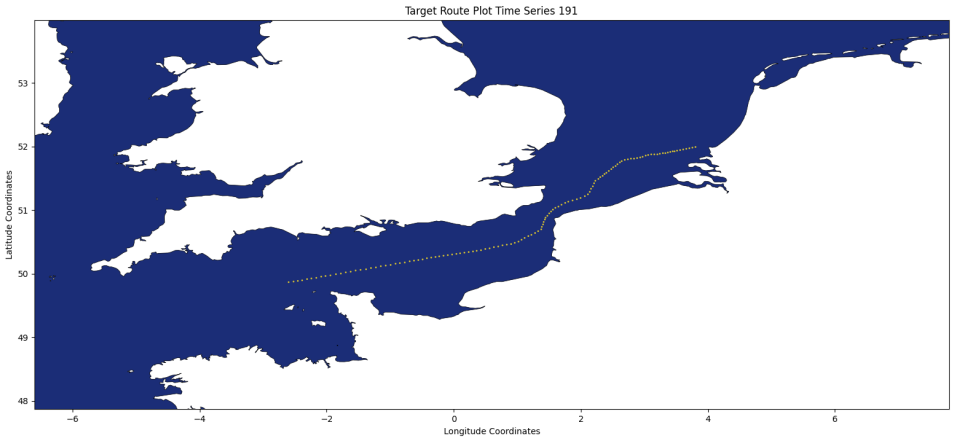
Figure 17: Voyages of the HTV Target since 2012

Navigation

The vessel's navigational data, recorded every 15 minutes, includes its location in Degrees Minutes Seconds (DMS) format, speed, and heading. These data, together with date and time data, are essential for extracting environmental data belonging to these data points. Furthermore, the voyages can be illustrated using these data points. For example, Figure 18(a) illustrates voyage 22, where HTV Target traveled from Sikki, India, to Rotterdam, the Netherlands. In particular, this depiction reveals gaps between some data points, a common occurrence in most voyages. To address these gaps, voyages are divided into smaller segments called 'time series', characterized by data points not exceeding 30 minutes in separation. This division allows for in-depth analysis and comparison of different time series. The data set comprises 191 distinct time series in total. Time series 191, the last segment of voyage 22, is presented in Figure 18(b). This time series traces the HTV Target's journey from the French coast, through the English Channel, to Rotterdam, with each yellow dot marking a data point. Note that this time series belongs to the voyage shown in Figure 18(a).



(a)



(b)

Figure 18: Data Characterization and Structure of Navigation Data: (A) Voyage 22, and (B) Time Series 191

Observed Vessel Response Data

The data obtained from the vessel response in all 22 voyages are considered as statistical maximum values. These values are estimated using the Rayleigh distribution to account for rare events, specifically the highest motion amplitude within a 3-hour period. To determine these maximum values, the standard deviation of a 15-minute interval is calculated (refer to Chapter 2, Section 2.2.5). Figures 16(b)

and 16(a) display the distribution of these statistical maximum roll and pitch motions, respectively. The pitch motion data show a higher density, indicating a narrower range of values. On the other hand, the maximum roll motion data have a wider spread, as indicated by their higher values. For instance, Figure 18(b) shows the time series 191, which illustrates the development of the statistical maximum roll and pitch motions over time (Figure 19).

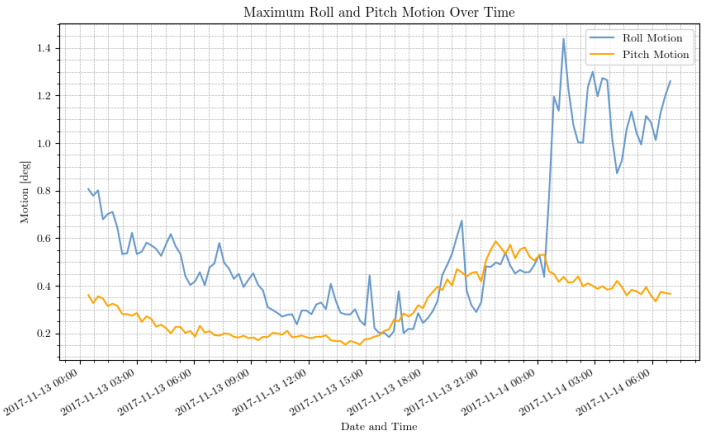


Figure 19: Maximum Roll and Pitch Motions over Date/Time for Time Series 191

Loading Conditions on Vessel

The parameters covered by the general hydrostatic LCs include displacement, metacentric height, free surface moments, and center of gravity. These parameters, which are discussed in Chapter 2, Section 2.1, are assumed to remain constant throughout a voyage, which is generally true in practice. Each voyage involves the transportation of a single heavy cargo, and therefore the displacement during a voyage only changes slightly due to factors such as the use of liquids on board or fuel consumption. The main component of the displacement parameter, the deadweight of the vessel and its cargo, remains constant, so the influence of the decrease in fluid volume is minimal. This also applies to other parameters. In some cases, the LCs were recalibrated during a voyage, resulting in a total of 24 LCs instead of the expected 22. This indicates that the recalibration took place during two voyages.

Environmental Data

In this comparative study, two approaches use nowcast environmental data, namely OCTOPUS and NauticalNet*, while the other two approaches use hindcast environmental data. These environmental data sources are obtained from different providers. Hindcast environmental data for NauticalNet and HindHold are sourced from the European-funded Copernicus Climate Change Service, specifically the fifth European Centre for Medium-Range Weather Forecasts Reanalysis (ERA5) (Hersbach et

al., 2020). ERA5 is known for its high-resolution data on global ocean waves and other weather conditions, offering hourly data on a 0.36-degree grid. For NauticalNet and HindHold, a regridded subset of ERA5 is used, resulting in a grid size of 0.5° by 0.5°. These data are also obtained from Copernicus (Hersbach et al., 2023). Copernicus provides API capabilities, enabling efficient linkage of environmental data with sensor data. However, it is important to handle the ERA5 data with caution due to the possibility of outliers at near-surface wind speeds (Hersbach et al. (2020)). 20 illustrates the ERA5 grid, showcasing the spatial resolution near the Dutch coast as an example. Environmental parameters such as the significant wave height or peak wave period are available at each grid point. The nowcast environmental data for OCTOPUS and NauticalNet* are extracted from the OCTOPUS Onboard system, which relies on a provider called SPOS. The nowcast data set contains fewer environmental variables compared to the hindcast environmental dataset. More details on this difference will be provided in the subsequent section 4.1.2 on feature selection.

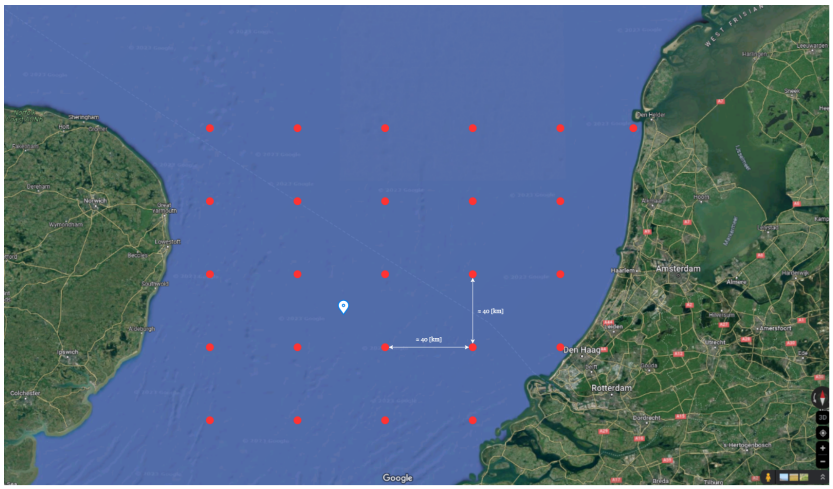


Figure 20: ERA5 HRES Grid Representation

Data Merging

In this study, four distinct data sources are merged into a single dataset for subsequent ML analysis. This process requires the alignment of navigational data, which includes vessel speed and location, with observed statistical maximum response values, LCs, and environmental data. Additionally, the OCTOPUS predictions are aligned with this dataset for comparative analysis.

The vessel response data are matched to the navigational data based on date and time, with a maximum allowable difference of 15 minutes. Navigational data, which contain precise information, lead this process. The observed maximum motions, being statistical in nature and not strictly tied to specific timestamps, are aligned accordingly. Instances where matching within the fifteen-minute window is not feasible result in data exclusion. The LCs, assumed to be constant throughout a voyage, are straightforwardly associated with corresponding data points using date and time information.

Environmental data from nowcast and hindcast sources are integrated differently, each. Nowcast data, available in the OCTOPUS onboard system, are matched in the same manner as the observed maximum response motions. For hindcast data, the ERA5 dataset is utilized. Using API calls, the environmental data for each navigational data point is extracted based on the nearest interpolation method. This method selects environmental data from the closest grid point to the vessel's location. The linear interpolation approach, initially considered, was discarded due to significant data loss in near-shore scenarios, as detailed in Appendix A (Figure 67).

Handling Missing Data

The final dataset, post-merging, includes data from 2012 to 2018. The strict 15-minute matching criterion between navigational data, observed maximum response data, and OCTOPUS predictions, combined with the focus on sailing conditions, influenced this date range. In particular, data for 2020 and 2021 were unavailable. Furthermore, for 2022 and 2023, annual reports or detailed voyage information was lacking, leading to their exclusion. Additionally, certain observed maximum response data points were missing prior to the merging process, likely attributable to sensor failures or errors in raw data post-processing. Consequently, these points were omitted from the dataset.

4.1.2 Feature Selection and Data Cleaning

This section addresses the selection and cleaning of data for three distinct ML-based prediction approaches: NauticalNet, NauticalNet*, and HindHold. Each approach utilizes a different dataset: a hindcast environmental dataset for NauticalNet, a nowcast environmental dataset for NauticalNet*, and a specialized hindcast dataset for HindHold, which aligns with OCTOPUS's available features and predictions. This differentiation is crucial, as each dataset influences the approach's performance in predicting the maximum roll and pitch motion of vessels.

The initial dataset, derived from the merged data, comprises various features categorized into five main groups. A detailed description of these features is provided in Appendix B. Data cleaning involves the detection and handling of outliers, the management of missing values, the scaling of data, and the maintenance of the data that are adequately prepared for the ML training process.

1. **Temporal data:** Features indicating time-related aspects.
2. **Geospatial data:** Features related to geographical locations and geophysical attributes.
3. **Environmental data:** Features representing environmental conditions.
4. **Loading conditions:** Features reflecting the vessel's general hydrostatic LCs.
5. **Response parameters:** Variables measuring the vessel's motion responses and accelerations.

Feature Transformation Feature transformation, as defined by Dong and Liu (2018), involves creating new features from existing ones through mathematical mappings. This step helps in establishing potential (non)linear relationships between features, aiding the model in learning real-world vessel responses. Key features generated in this step include:

- *Distance to Shore:* Calculated using geospatial data to assess the vessel's proximity to the shoreline, influencing wave dynamics.
- *Total Wind Speed and Direction:* Derived from components of the wind speed U and V, which affects the statics of the vessel due to wind loads.
- *Mean Wavelength and Velocity of Wind Waves and Swell:* Based on the relationship between sea depth and wave period, could be important for predicting vessel pitch motions.
- *Relative Directional Features:* Transforms absolute directions into relative orientations with respect to the heading of the vessel, important for understanding the impact of external forces.
- *Windowed Loading Conditions:* Each LC feature is categorized into intervals, with each interval assigned an integer label. The number of intervals is determined by the range (maximum and minimum values) and the count of unique values for each feature. This categorization aims to improve the generalization of the model under diverse conditions.

Preventing Leakage To avoid data leakage, which occurs when training data inadvertently includes additional information, features unavailable at the prediction time or directly derived from target variables are excluded. The only two features removed at this stage are:

- *10 Metre Wind Gust Since Previous Post-Processing:* Excluded due to reliance on unavailable post-processing data at prediction time.
- *Arrival Date:* Excluded as it implies future information.

Dropping Irrelevant and Redundant Features As suggested by Kotsiantis et al. (2006), eliminating irrelevant and redundant features enhances the efficiency of learning algorithms and reduces noise. The following features are removed:

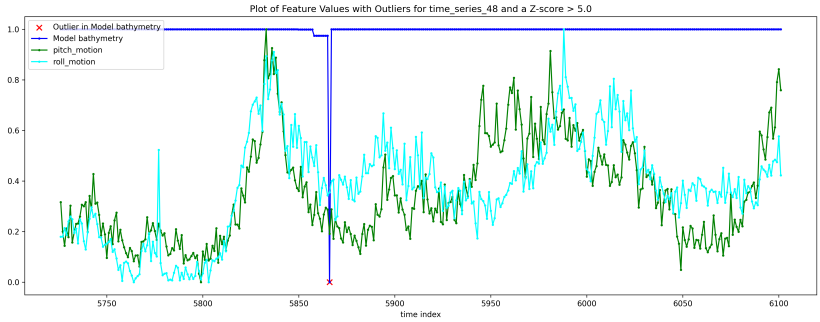
- *Unique Identifier Features:* Not relevant for predicting vessel response.
- *Date and Time Features:* Irrelevant as natural laws are date/time-invariant.
- *Longitude and Latitude:* Excluded, as environmental conditions, not locations, affect vessel response.
- *Total Precipitation:* Removed due to minimal impact on vessel motions.

- *Instantaneous 10m Wind Gust:* Dropped due to its highly localized nature.
- *Air Density over Oceans:* Minimal impact on vessel drag, therefore excluded.
- *Coefficient of Drag with Waves:* Excluded due to minimal correlation with the sea state.
- *Free Convective Velocity over Oceans:* Assumed negligible impact on vessel motion.
- *Initial Wind Features:* Replaced by transformed wind speed and direction features.
- *Normalized Energy Fluxes and Stresses:* Removed due to lack of direct influence on vessel response.
- *Mean Wave Periods from First and Second Order Moments:* Dropped due to strong correlation with regular mean wave period features making these features redundant.
- *Change in Metacentric Height and Free Surface Moment:* Removed due to redundancy and faulty data, respectively.

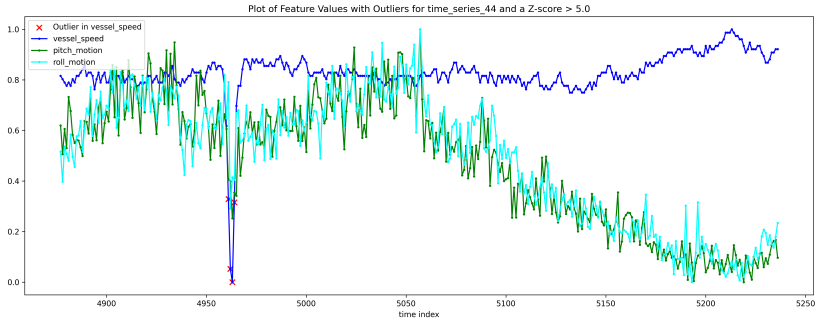
Outlier Detection and Handling Outliers, representing unusual or physically illogical data points, often arise from malfunctioning or damaged sensors. Eliminating these outliers is essential to prevent the model from training on inaccurate data. A common method for outlier removal involves discarding the outlier data points. Outlier detection utilizes the calculation of a Z-score for each data point, indicating how many standard deviations a point's value is from the feature's mean. This calculation is performed individually for each time series, as feature values can vary significantly between different series. The Z-score is calculated using $Z = \frac{x-\mu}{\sigma}$, where x is the data point's value, μ is the feature's mean for the time series, and σ is the standard deviation of that feature within the series. A Z-score threshold of 5 was established after lower thresholds led to frequent false detections. Detected outliers are manually assessed based on domain knowledge to ensure accuracy (see Figure 21). It is important to note that the values depicted in the graphs are scaled. For instance, Figure 21(a) illustrates a correctly identified outlier, while Figure 21(b) demonstrates an incorrectly marked outlier. Through this detailed process, 84 data points were removed from the dataset after reviewing all 191 time series.

Handling Missing Values Data points with missing values are excluded from the dataset. Such omissions are important, as these incomplete data points could disrupt the training process. Predominantly, missing values were observed in environmental features and were frequently found in near-shore locations.

Scaling the Data Scaling the data is a critical step in enhancing the efficiency of the training process and ensuring consistent behavior of the ML model. This procedure is commonly recommended for many ML applications (Pedregosa et al., 2011). Scaling is performed independently for each feature,



(a) Outliers detected in time series 48.



(b) Incorrectly identified outliers in time series 44.

Figure 21: Graphs illustrating outlier detection.

applying the standard score transformation across all time series data collectively. Each data point value for every feature is replaced with its standard score:

$$z = \frac{x - \mu}{\sigma} \tag{4.1}$$

NauticalNet's Hindcast Data Set

The resulting hindcast dataset for NauticalNet consists of 47 features and 51,256 data points. These features are segmented into categories that describe their relevance to the prediction of vessel motion. The categorization includes LCs, wave characteristics, and other environmental factors.

Loading Conditions

- *Displacement.* Displacement significantly influences the buoyancy forces acting on the vessel, integral to its response to external forces. It correlates with both the mass and added mass of the system, affecting vessel accelerations.
- *Metacentric height.* GM and GMFLUID denote the metacentric height, a key indicator of vessel

stability. A larger GM implies a more substantial righting arm, offering greater resistance to external moments. GMFLUID, representing the corrected GM for onboard fluid surfaces, is crucial for similar reasons.

- *CoG.* The CoG location impacts the external moment arms and influences GM. While it also affects vessel trim or list, the minimal variance in the CoG's y-direction led to its exclusion from the dataset.
- *Radii of gyration.* These indicate the vessel's moment of inertia, critical in resisting angular accelerations and affecting the dynamics of rigid bodies and fluid motions. A higher moment of inertia implies slower vessel responses to external forces.
- *Draft features.* Forward and aft drafts are considered important, influencing hydrodynamic pressures and potential flow theory. The drafts' difference provides insights into the vessel's trim.
- *Natural frequencies.* The vessel's natural frequency, indicating its oscillation in the absence of waves, also highlights its sensitivity to external frequencies, and is directly related to the vessel's RAO.

Wave Characteristics

- *(Significant) wave heights.* Wave heights are crucial, affecting the waves' potential and kinetic energy. Generally, higher waves ensure stronger responses.
- *Relative wave direction.* This influences the encountering frequency and vessel response, with beam waves predominantly affecting roll motion and head waves affecting pitch.
- *Wave periods.* Related to the waves' potential energy, certain frequencies near the natural frequency render the vessel more responsive, as indicated by the vessel's RAO.
- *Mean square slope of waves.* This slope is indicative of wave breaking likelihood. Steeper waves, with potentially higher dynamic pressures, are more prone to break.
- *Wavelength and velocity of ocean waves.* Wavelength affects how the vessel interacts with waves, with potential resonance at certain lengths. It also factors into calculating wave energy. Wave velocity simplifies the relationship between vessel speed and wave interaction.

Other

- *Vessel speed.* This affects the encountering frequency of waves, especially when moving against them, influencing the vessel's response as described by its RAO.

- *Velocity and relative direction of wind and current.* While these may not significantly impact roll and pitch motions, they affect the vessel's static stability and, in turn, can influence drag when moving against currents. The features for the current are sourced differently from the wave and wind characteristics, utilizing the same provider as OCTOPUS.
- *Sea depth and distance to shore.* Being near the shore and sea depth considerably affect wave behavior. Near the shore, breaking waves, shoaling, and refraction are more evident, and in shallower depths, these effects tend to intensify.

NauticalNet*'s Nowcast Dataset

The nowcast dataset for NauticalNet* is derived similarly to the hindcast dataset but with distinct environmental features. This dataset shares several input features with NauticalNet, including all LC features and most features from the 'other' category. These features will not be further addressed here. However, it differs in 13 features that are present in both approaches but sourced differently and lacks 15 features that are exclusive to NauticalNet. As a result, NauticalNet* incorporates a total of 32 input features, including OCTOPUS predictions for maximum roll and pitch motions.

Shared Environmental Features: Both NauticalNet and NauticalNet* share the following features, although sourced differently:

- Significant and maximum wave heights (including wind and swell waves, maximum, and combined) [4 features].
- Relative wave directions (covering wind and swell waves, and combined) [3 features].
- Wave periods (pertaining to wind and swell waves, peak, and combined) [4 features].
- Wind speed and direction [2 features].

Unshared Environmental Features The following 15 features, present in NauticalNet, are absent in NauticalNet*:

- Mean square slope of waves [1 feature].
- Wavelength and ocean wave velocity [2 features].
- First, second, and third swell component partitions regarding wave height, direction, and period [9 features].
- Mean zero-crossing wave period and period for maximum individual wave height [2 features].
- Wave spectral directional widths [3 features].

HindHold's Hindcast Dataset

HindHold's input layer contains 32 features. It shares 30 features with NauticalNet *, excluding OCTOPUS predictions. Instead of having the observed maximum roll and pitch motions as target variables, HindHold has the OCTOPUS predictions for these motions as target variables. Moreover, unlike NauticalNet and NauticalNet*, which use relative vessel heading, HindHold employs absolute vessel heading to align with OCTOPUS's input, resulting in one more feature with respect to NauticalNet*. Also, the free surface moment is included in this approach. The center of gravity on the y-axis is omitted due to minimal variance, aligning with ML principles that require variance in features for effective learning. The resulting 32 input features include:

- Displacement [1 feature].
- Metacentric height [2 features].
- Free surface moment [1 feature].
- Center of Gravity [2 features].
- Radii of gyration [3 features].
- Draft features [2 features].
- Natural frequencies [2 features].
- Vessel speed and heading [2 features].
- Current speed and direction [2 features].
- Significant and maximum wave height (of wind and swell waves) [2 features].
- Wave direction (of wind and swell waves) [2 features].
- Wave periods (mean, peak, normal period of wind and swell waves, mean, peak, and all combined) [9 features].
- Wind speed and direction [2 features].

4.2 Model Development

Following the comprehensive preparation of datasets and feature selection, this section delves into the development of the prediction approaches: NauticalNet, NauticalNet*, and HindHold. The focus here is on detailing the architecture, training configurations, and the training process for each approach, emphasizing their distinct design elements and the type of environmental data utilized.

For NauticalNet and NauticalNet*, the architectures and training configurations are described in Subsections 4.2.1 and 4.2.2, respectively. These approaches are trained and validated using sensor data and precalculated LC data from the HTV Target's voyages. The training configurations and methodologies used are specifically tailored to exploit the unique characteristics of the hindcast and nowcast environmental datasets. HindHold, discussed in Subsection 4.2.3, deserves additional focus. In addition to its architecture and training, the development of this approach includes an exploration of its mirroring capabilities with respect to OCTOPUS. This is vital for validating HindHold as a useful addition in this comparative analysis, particularly in its adaptability to encountering unknown LCs. The approach's capability to replicate and potentially enhance OCTOPUS's predictions forms a crucial part of its evaluation. Collectively, this section aims to provide a clear understanding of how the prediction approaches were developed.

4.2.1 NauticalNet

This section provides a comprehensive exploration of the architecture of NauticalNet, which includes the process of selecting hyperparameters and the use of custom loss and validation functions. This discussion culminates in an examination of the training configuration and process, providing a comprehensive overview of NauticalNet's development.

Architecture

The architecture of NauticalNet, depicted in Figure 22, embodies a multi-task deep learning neural network, as outlined in Subsection 14. At the outset, NauticalNet incorporates 47 input features, detailed in the preceding subsection. The development of NauticalNet followed an iterative process, alternating between manual adjustments and automated grid searches, which are iterative training sessions with varying training setups, culminating in the current architecture. NauticalNet's architecture features a combination of shared and task-specific hidden layers. The shared layers contribute to predictions for both maximum roll and pitch motions, while the task-specific layers are tailored for each task. Notably, the roll task incorporates an additional layer and a larger number of neurons, weights, and biases compared to the pitch task. This difference, a byproduct of the grid search, suggests that the neural network benefits from a more complex architecture for roll motion prediction, aligning with findings in the literature (Section 2.4). NauticalNet is developed using the PyTorch library (Paszke et al., 2019).

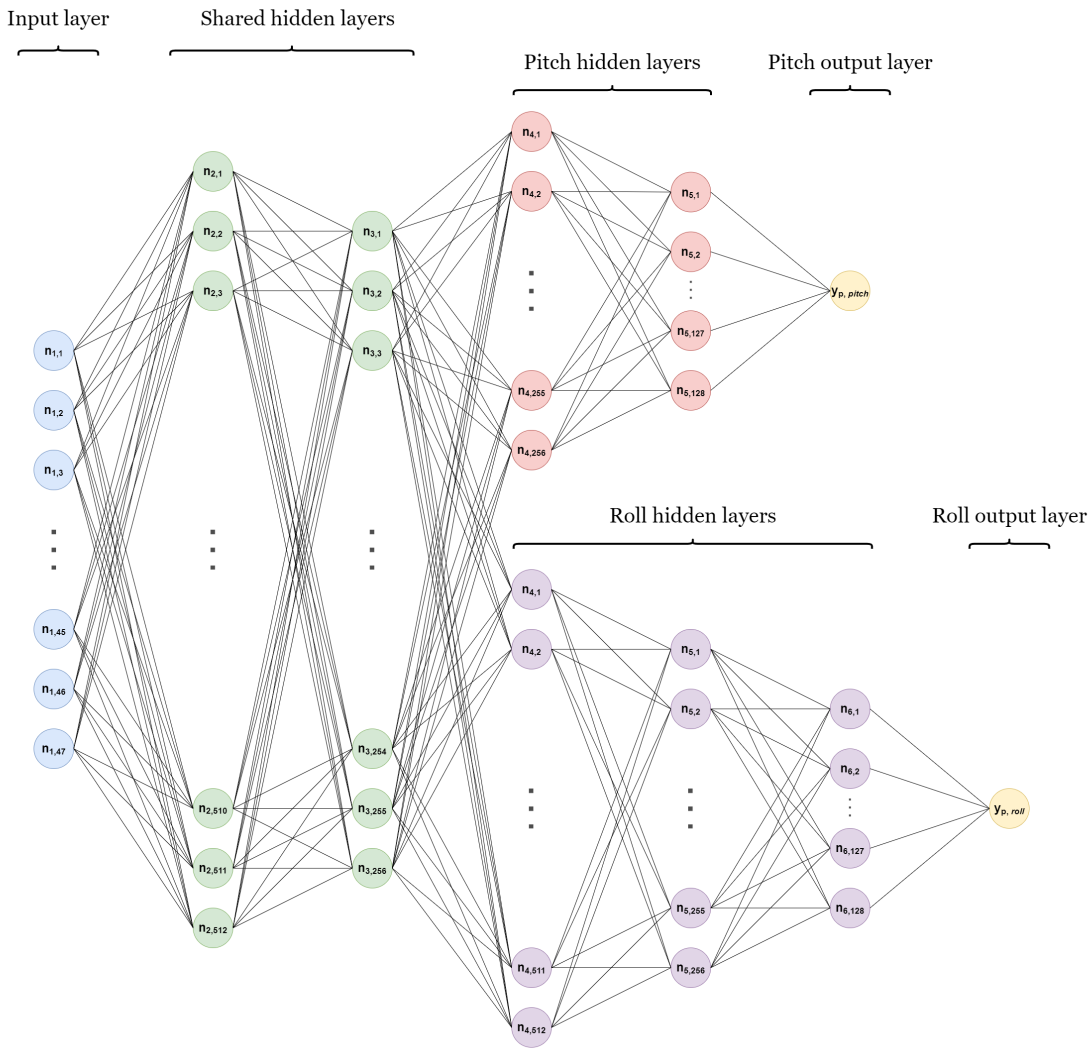


Figure 22: Schematic View of NauticalNet's Architecture

Hyperparameters

Hyperparameters, crucial in shaping the training process of a ML-based approach, are outlined for NauticalNet below. Their selection originated from a mix of manual optimization and grid searches. More information on the specific hyperparameters are provided in Chapter 2, Section 2.3.

- **Number of epochs:** 1250
- **Early stopping patience:** 45 [epochs]
- **Mini batch size:** 512.
- **Initial learning rate:** $1.5 \cdot 10^{-5}$

- **Learning rate patience:** 20 [epochs]
- **Learning rate factor:** 0.1
- **Dropout rate:** 0.2
- **Weight decay:** $1.125 \cdot 10^{-3}$

Custom Cross-Validation Strategy

NauticalNet's cross-validation strategy customizes data allocation for testing and training. As described in the methodology chapter, in validation strategy I, the latter 10% of each time series is reserved for testing, allocating the initial 90% for training and validation. The strategy classifies the top third (33.33%) of values in each series as 'significant', with the remainder deemed 'insignificant', adhering to industry standards and the validation cases used in this study. For validation and training, 1/8th of data from both categories is randomly allocated for validation, with the rest designated for training. This results in a training dataset comprising 78.75% of the original data and a validation set of 11.25%.

Custom Loss Function

NauticalNet's specialized loss function gives more importance to errors in higher values when making predictions for maximum roll and pitch motions. By default, the function uses a value of $\beta = 1.0$, which increases the penalty for errors in larger target values. This helps to address the inaccuracies that have been observed in high-value predictions. The loss function modifies the mean squared error (MSE) by assigning weights to errors based on the magnitude of the true values. This allows for a more precise adjustment of the model's performance. The formulation of the function is as follows:

$$L(y_{\text{pred}}, y_{\text{true}}, \beta = 1.0) = \frac{1}{N} \sum_{i=1}^{N} \left(\left(y_{\text{pred}_i} - y_{\text{true}_i} \right)^2 \cdot \left(\log(|y_{\text{true}_i}| + 1) \cdot \beta + 1 \right) \right) \tag{4.2}$$

- y_{pred} and y_{true} are the vectors of predicted and true values, respectively.
- β is a parameter to control the impact of the weighting scheme on the loss function. When $\beta = 1.0$, the weights are determined solely by the logarithm of the absolute value of the true values plus one.
- The term $\log(|y_{\text{true}_i}| + 1) \cdot \beta + 1$ is used to compute the weights for each element in the dataset. This weighting scheme scales the contribution of each sample to the loss based on the magnitude of its true value.
- The expression $(y_{\text{pred}_i} - y_{\text{true}_i})^2$ represents the mean squared error (MSE) loss for each element.

- The weighted loss for each sample is obtained by multiplying the MSE loss by the computed weights.
- Finally, the expression $\frac{1}{N}\sum_{i=1}^{N}$ represents the aggregation of the loss across all N samples in the dataset, resulting in a single scalar value that quantifies the performance of the model.

Model Training

NauticalNet went through multiple iterations, starting with a plain vanilla neural network (see Chapter 2, Section 2.3.3) and gradually increasing its complexity by adding layers and refining hyperparameters. Once manual adjustments reached a plateau in reducing validation losses, different grid search techniques were used to optimize the architecture. Once the architecture of NauticalNet was established, further manual tuning of the hyperparameters was performed. Although it would have been ideal to conduct a comprehensive grid search optimization for hyperparameters, limited computing resources prioritized architectural grid searches. The training of NauticalNet showed stable trends in both validation and training loss (i.e., no rapid fluctuations were observed), as shown in Figure 23. Early stopping was triggered at approximately 310 epochs, with validation losses slightly lower than training losses, possibly due to the implementation of regularization techniques such as weight decay and dropout.

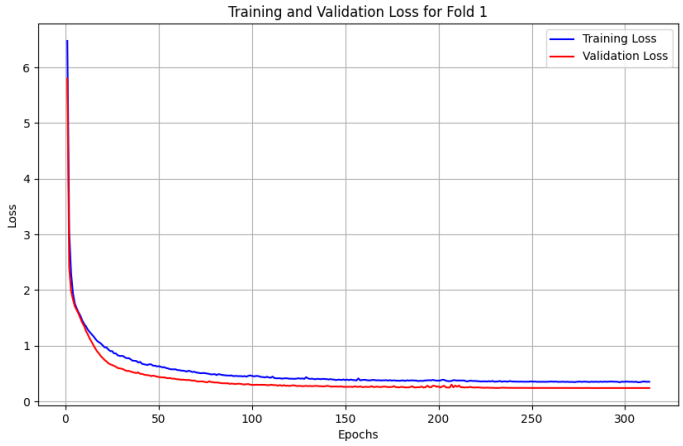


Figure 23: Training of NauticalNet's Fold 1

4.2.2 NauticalNet*

NauticalNet* has a reduced input feature set, comprising 32 features compared to 47 of NauticalNet. Except for the input layer, NauticalNet* maintains the same architecture as NauticalNet, as depicted in Figure 24. The hyperparameters remain unchanged, as well as the custom loss function and the

custom cross-validation approach. This subsection will focus predominantly on the differences that arise from the altered data source.

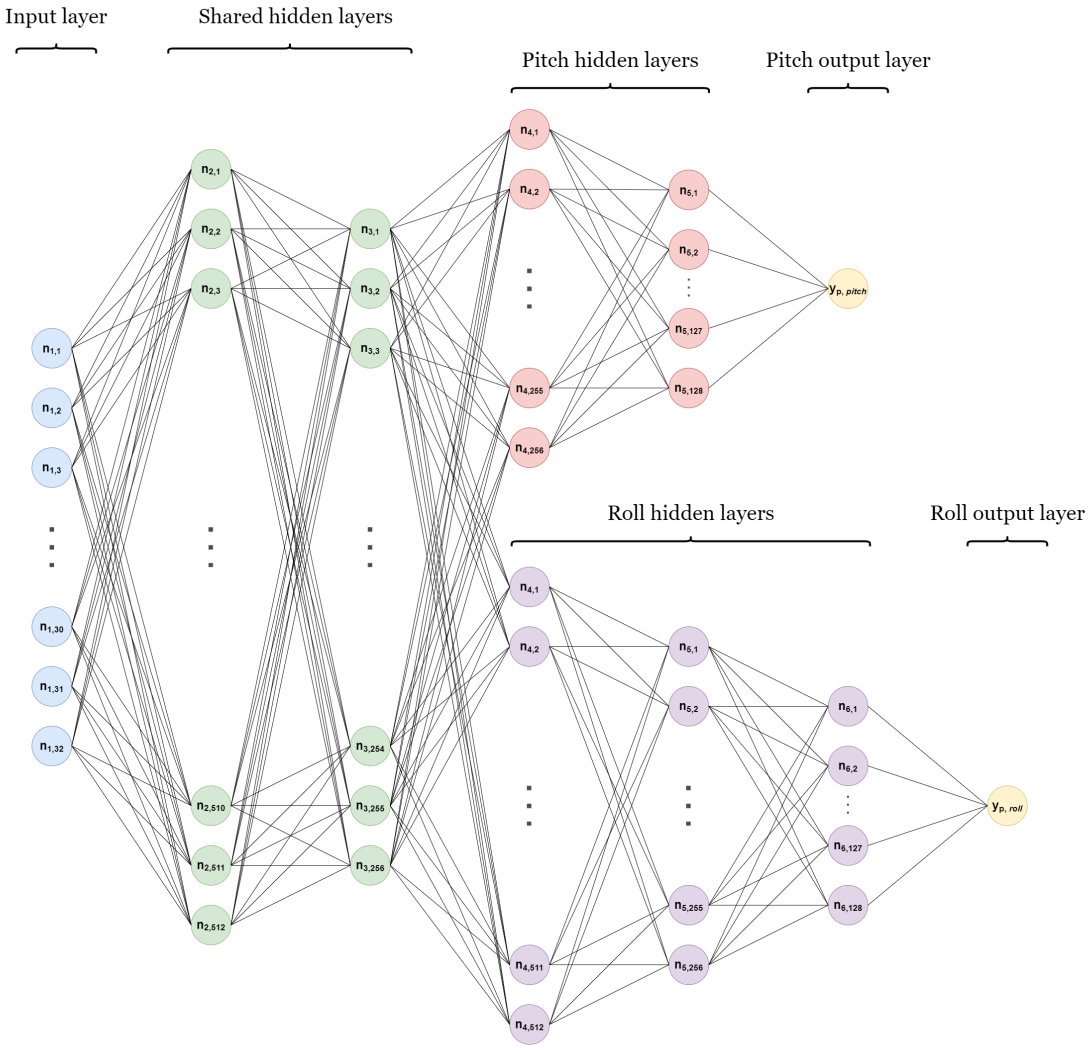


Figure 24: Schematic View of NauticalNet*'s Architecture

Model Training

NauticalNet*'s development involved minimal unique steps; the primary focus was adapting the input structure. The model reuses the hyperparameters, custom loss function, cross-validation strategy, and architecture of NauticalNet. NauticalNet*'s training progressed smoothly, marked by stable validation and training losses, as illustrated in Figure 25. A notable difference from NauticalNet is the difference in training and validation losses during training. When training NauticalNet* the training loss is lower compared to the validation loss, potentially indicating overfitting. This could result from applying a

complex architecture to a reduced set of input features. Notably, the training spanned approximately 440 epochs, significantly more than NauticalNet.

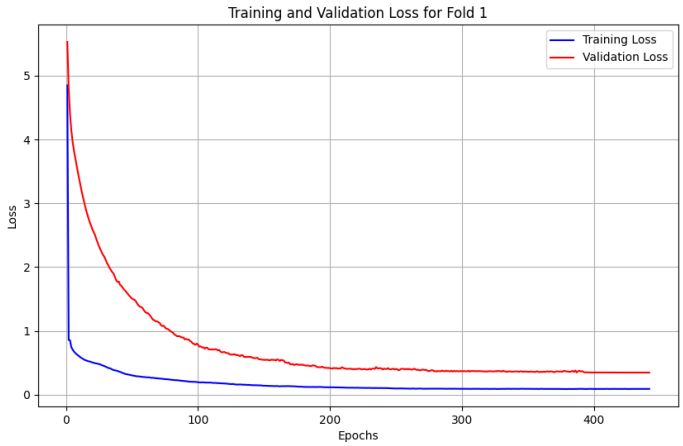


Figure 25: Training of NauticalNet*'s Fold 1

4.2.3 HindHold

HindHold represents a unique approach in the comparative analysis of ML-based and conventional prediction approaches for vessel motions. This subsection details the development of HindHold, focusing on its architecture, hyperparameters, model development, and notably, its mirroring capabilities with respect to OCTOPUS. Emphasizing these aspects underscores HindHold's role in this study, particularly its ability to replicate OCTOPUS predictions and potentially extend their utility.

Architecture

HindHold's architecture, depicted in Figure 26, consists of four hidden layers forming a pyramid-like structure, decreasing from 512 to 64 neurons, and culminating in a two-neuron output layer. This architecture, unlike NauticalNet and NauticalNet*, does not specialize in specific tasks but predicts both roll and pitch using the same layers. Task-specific layers did not enhance performance, probably due to HindHold's role in modeling another model, which is simpler than modeling real-world phenomena. Developed using Keras and TensorFlow, HindHold's simplicity is balanced by fewer customization options (Chollet et al., 2015). Training and validation utilized 90% (approximately 45,000 datapoints) of OCTOPUS data, and the remaining 10% (approximately 5,000 datapoints) were reserved for testing.

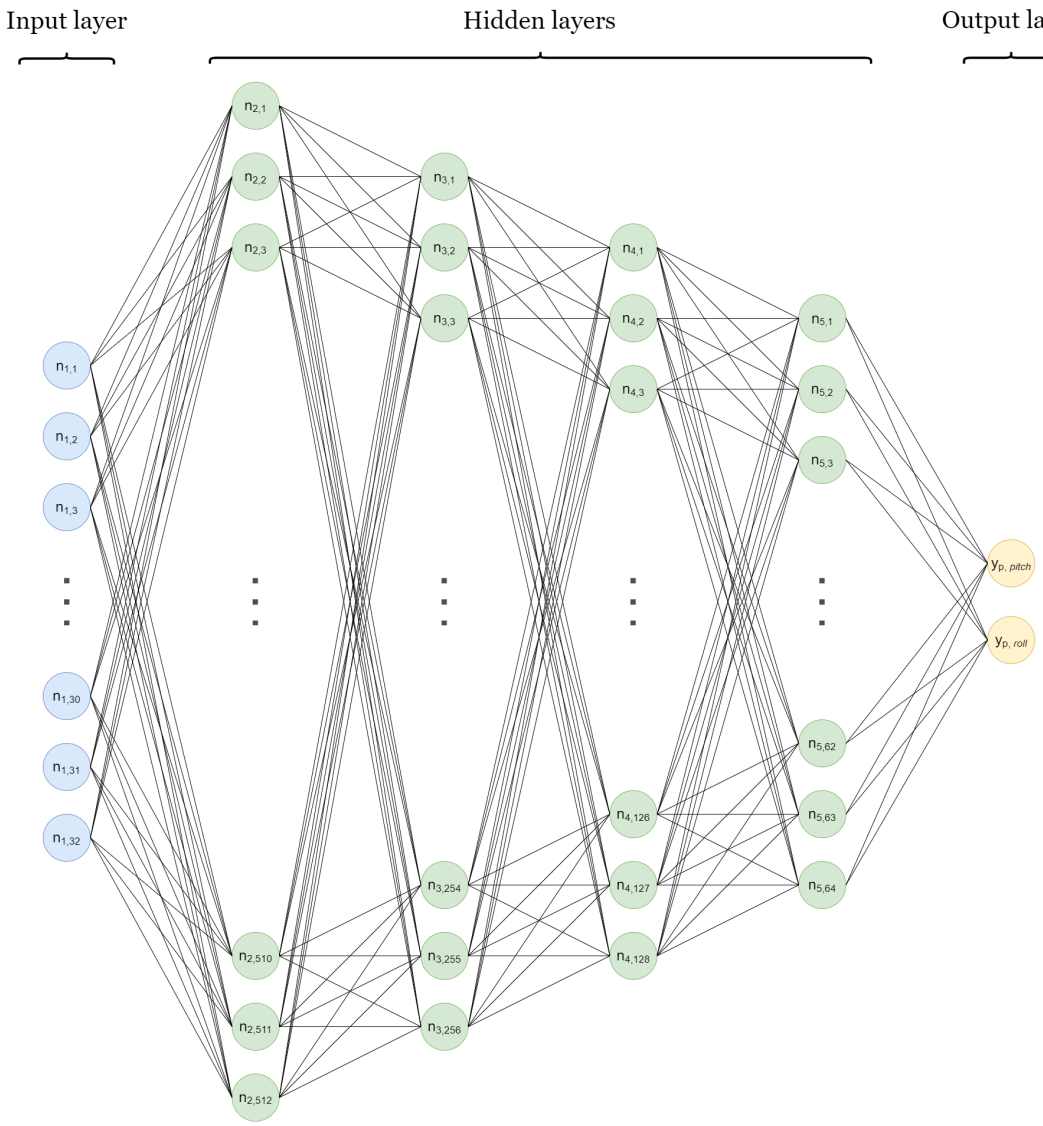


Figure 26: Schematic View of HindHold's Architecture

Hyperparameters

HindHold's hyperparameters, predominantly identified through manual optimizations and grid searches, are outlined below. For detailed descriptions of the Hyperparameters, refer to Section 2.3.

- **Number of epochs:** 700
- **Early stopping patience:** 20 [epochs]
- **Mini batch size:** 256

- **Initial learning rate:** $1.25 \cdot 10^{-4}$
- **Learning rate patience:** 10 [epochs]
- **Learning rate factor:** 0.1
- **Weight decay:** $1.0 \cdot 10^{-3}$

Model Training

The model's development began with a simple neural network, gradually increasing in complexity through additional layers, neurons, and learning techniques. Following significant validation loss reductions, grid searches determined the final architecture. Due to computational limitations, hyperparameters were manually fine-tuned. The mean squared error loss function and common cross-validation techniques, dividing training data into 8 folds, were employed. Figure 27 displays the training process, indicating a well-fitted model through the gradual and smooth decrease in training and validation loss, with early stopping at epoch 331.

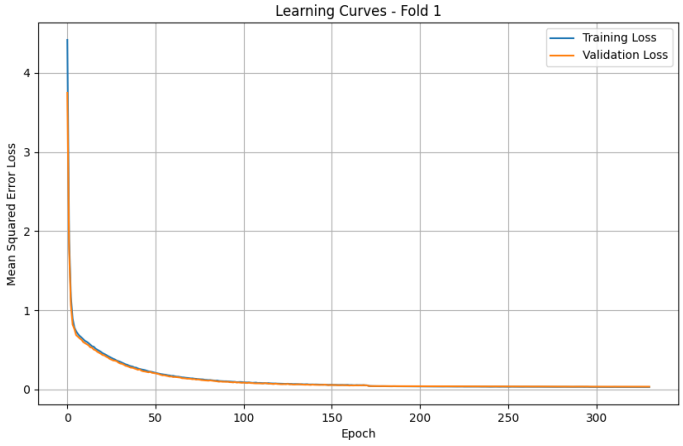
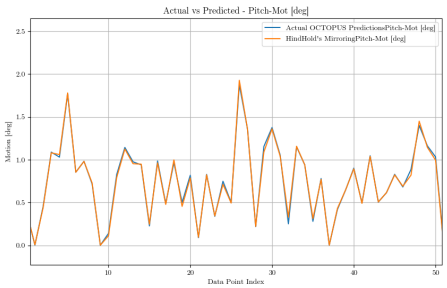


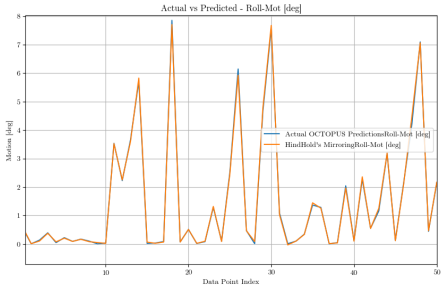
Figure 27: Training of HindHold's Fold 1

Mirroring Capabilities

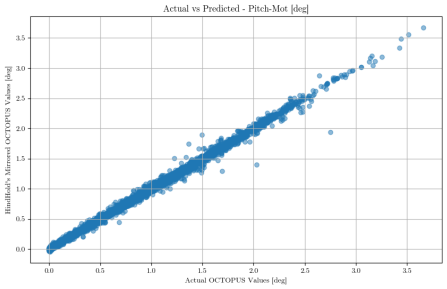
HindHold's predictive performance is validated by comparing its predictions against actual OCTOPUS values. Figure 28 showcases HindHold's mirroring capabilities for pitch and roll motions in various formats. Subplots 28(a) and 28(b) display the first 50 test indices, illustrating HindHold's mimicry of OCTOPUS's predictions for pitch and roll, respectively. Subplots 28(c) and 28(d) plot OCTOPUS versus HindHold values for the entire test set, indicating mirroring proficiency. The mean absolute errors, quantifying the discrepancy between OCTOPUS and HindHold predictions, are 0.02 [deg] for pitch and 0.05 [deg] for roll, evidencing HindHold's effective mirroring of OCTOPUS predictions.



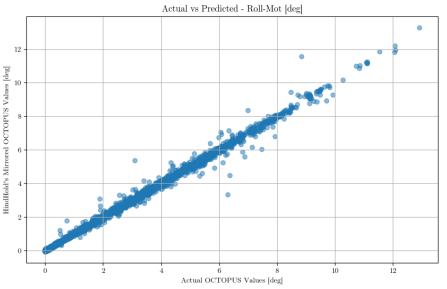
(a) Mirroring OCTOPUS's Pitch Predictions First 50 indices of Test Data



(b) Mirroring OCTOPUS's Roll Predictions First 50 indices of Test Data



(c) HindHold's Mirroring Over All OCTOPUS Pitch Test Data



(d) HindHold's Mirroring Over All OCTOPUS Roll Test Data

Figure 28: Mirroring Capabilities of HindHold Approach

4.3 Validation Framework Design Choices

This section expands on the methodology chapter by detailing the choices made in the validation framework. It focuses on why and how the evaluation metrics and validation strategies were chosen. The metrics used in this study are explained in simple terms, showing their importance in measuring how accurately the different approaches predict vessel motions. Next, the section discusses the two main validation strategies. These strategies are key to testing how well the ML methods work with both familiar and unfamiliar LCs. This part of the research helps to better understand the strengths and weaknesses of these methods. Overall, this section connects the theory from the methodology chapter with the actual steps taken in the research, making it easier to see how the study was carried out in practice.

4.3.1 Elaboration on Evaluation Metrics

A mix of error measures is integrated into the validation process, each offering a distinct display of the comparitive prediction accuracies of the approaches. This subsection is devoted on detailing how the error measures are determined.

Mean Absolute Error Mean Absolute Error (MAE) quantifies the average magnitude of prediction errors, irrespective of their direction. It is computed as shown in Equation 4.3, where n is the total number of predictions, and y_j^t and y_j^p represent the actual and predicted values, respectively.

$$\text{MAE} = \frac{1}{n} \sum_{j=1}^{n} \left| y_j^t - y_j^p \right| \tag{4.3}$$

Root Mean Squared Error Root Mean Squared Error (RMSE) evaluates the standard deviation of the residuals, emphasizing larger errors by squaring the residuals. Equation 4.4 details its computation. Again, n refers to the total number of predictions, while y_j^t and y_j^p represent the actual and predicted values for each j^{th} prediction, respectively. The square root operation ensures the error is expressed in the same units as the response variable.

$$\text{RMSE} = \sqrt{\sum_{j=1}^{n} \frac{(y_j^t - y_j^p)^2}{n}} \tag{4.4}$$

Mean Absolute Percentage Error Mean Absolute Percentage Error (MAPE) offers a relative assessment of prediction errors in percentage terms, providing an intuitive understanding of the error magnitude. Its calculation is outlined in Equation 4.5. Again, n refers to the total number of predictions, while y_j^t and y_j^p represent the actual and predicted values for each j^{th} prediction, respectively. By multiplying the result by 100, the error is expressed as a percentage.

$$\text{MAPE} = \frac{100}{n} \sum_{j=1}^{n} \left| \frac{y_j^t - y_j^p}{y_j^t} \right| \tag{4.5}$$

Over- and Underestimations Separately reported overestimations and underestimations can indicate tendencies towards over- or underprediction, which may have different practical implications. An error exceeding one standard deviation from the actual value is classified as an over- or underprediction. For maximum pitch and roll, these thresholds are 0.56 [deg] and 1.41 [deg], respectively.

4.3.2 Validation Strategy I: Known Loading Conditions

In validation strategy I, a comprehensive assessment of the predictive performance under known LCs is conducted. This strategy involves allocating 90% of the data from each voyage for training the ML-based approaches, including all the LCs encountered during these voyages. The remaining 10% of the data, primarily from the latter part of each voyage, is reserved for testing the predictive accuracy of all four approaches (OCTOPUS, NauticalNet, NauticalNet*, and HindHold).

The rationale behind selecting the final 10% of data for testing lies in the nature of ML training.

By excluding this portion from the training phase, the validity of the test results is ensured, as the approaches are not exposed to these data points during training. The chosen data division yields approximately 5000 test data points, striking a balance between a substantial training dataset and a sufficiently representative test dataset. Figure 29 illustrates this division for a randomly selected sample time series.

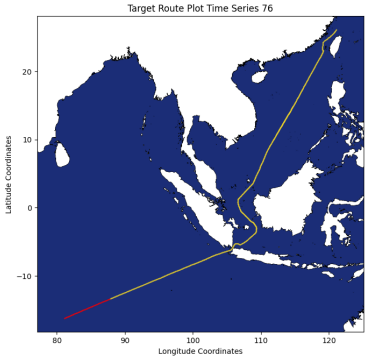


Figure 29: Train (in Yellow) and Test (in Red) Division of Data for Time Series 76

4.3.3 Validation Strategy II: Testing with Unknown Loading Conditions

Validation strategy II aims to evaluate the adaptability of ML-based approaches to unknown LCs, a critical factor in their practical application. This strategy trains the ML-based approaches with data from all voyages except one, specifically excluding the voyage's unique LC from the training dataset. The data from this excluded voyage are then utilized to test the performance of all approaches. The proportion of the different LCs in the data set is shown in Figure 30.

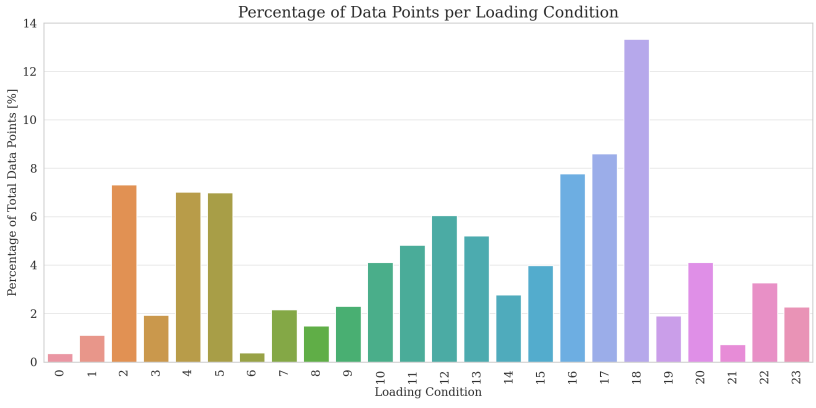


Figure 30: Distribution of Loading Conditions in the Dataset

The selection of which LC to exclude is crucial and is guided by its relevance to future trans-

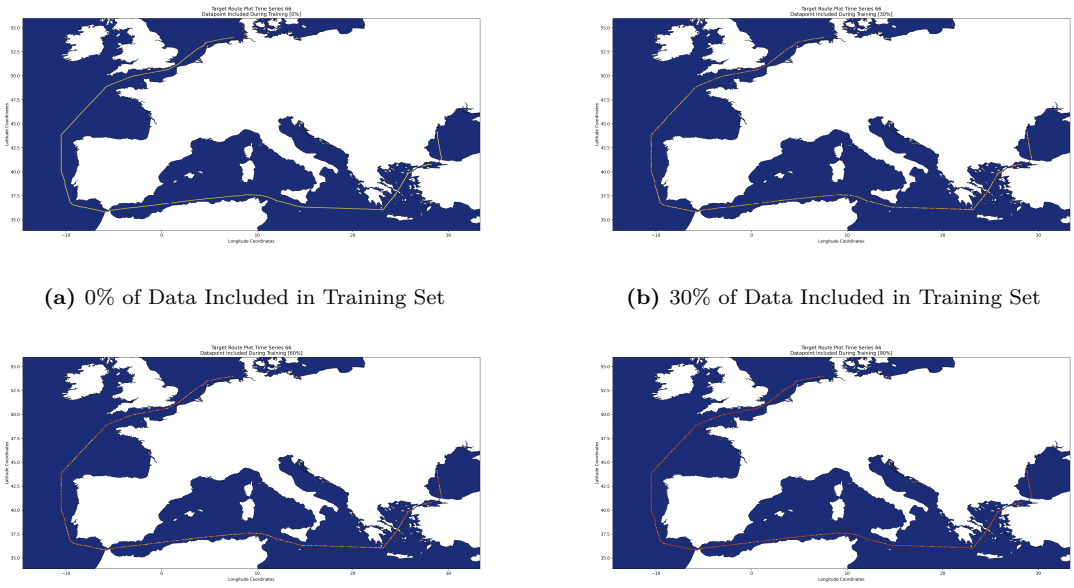
port operations. The chosen LC is one that should closely resemble the LCs expected in upcoming voyages. For this study, LC 7 was selected for exclusion due to its similarities with the anticipated monopile-loaded condition. This decision ensures that the validation effectively simulates a real-world scenario where the approaches encounter a previously unseen LC. Comparative features of LC 7 and the monopile-loaded condition are presented in Table 2.

Loading Condition Feature	**Van Oord Goliath's Loaded Condition**	**Monopile Loaded Conditions**
Displacement [t]	67.225	63.776
GM [m]	12,42	12,35
GM_fluid [m]	8,76	8
xcg [m]	108,42	109
ycg [m]	0	0,02
zcg [m]	9,2	9,6
rxx [m]	13,37	9,68
ryy [m]	52,74	36,04
rzz [m]	53,54	36,41
Draft aft [m]	10,67	10,15
Draft forward [m]	8,15	8,12
Natural Pitch Period [s]	13	13,73
Natural Roll Period [s]	11,54	9,41

Table 2: Loading Condition 7 Feature Values Compared with Monopile Loaded Conditions

4.3.4 Adaptability of Machine Learning-Based Approaches

The adaptability of ML-based approaches is further explored through an iterative training process, reflecting various degrees of exposure to the excluded LC's data. This methodology, ranging from complete exclusion to the inclusion of 90% of the excluded data, links both validation strategies and offers insights into the learning efficiency of ML approaches when faced with new LCs. The process is visualized in Figure 31.



(a) 0% of Data Included in Training Set

(b) 30% of Data Included in Training Set

(c) 60% of Data Included in Training Set

(d) 90% of Data Included in Training Set

Figure 31: Iterative Approach of Randomly Selecting Data Points

The findings from these validation strategies contribute to understanding the capabilities and limitations of ML-based approaches in the context of HTV motion prediction, a central theme in this research.

5. Results

This chapter consolidates results from implemented validation strategies and comparative analyses. In Section 5.1, results from Validation Strategy I are presented, comparing four approaches against observed motion values in the final segments of time series, focusing on maximum roll and pitch motions. Section 5.2 examines the outcomes of excluding LC7 from ML-based approach training. The chapter concludes with the results on the adaptability of ML-based approaches, particularly NauticalNet and NauticalNet*, highlighting their performance against (simulated) conventional approaches, in Section 5.3.

5.1 Validation Strategy I: Known LCs

In Chapter 3, the methodology for validation strategy I was introduced. The focus of this section is on the presentation of results obtained through this strategy, which involves a comparison of the predictions from four comparative approaches with the observed maximum values of motions during the final 10% of each time series. For an effective comparison, a series of steps will be undertaken. Initially, an evaluation of the maximum roll motion will be conducted, followed by an assessment of the maximum pitch motion. Throughout these sections, instances of trend deviations will be examined. This section culminates with an analysis aimed at elucidating the reasons behind the observed results.

5.1.1 Maximum Roll Motion Prediction Results

This subsection will investigate the maximum roll motions by examining the prediction plots to identify general trends. Subsequently, some time series that deviate from the observed trends will be discussed. Finally, overall performance will be evaluated considering overestimations, underestimations and performance metrics. Differences between approaches will be assessed by analyzing these components.

Visual Insights: Maximum Roll Motion Prediction Plots

In the Appendix C, the predicted maximum roll motions from the four approaches are plotted against the date/time values of the time series. These plots, which include eight time series each, show the observed maximum motions in the upper plot, along with the predictions and the corresponding absolute errors in the lower plot. The last data point of one time series overlaps with the next data point of the subsequent time series, with the x-axis label indicating the start of the subsequent

time series. Each minor tick on the x-axis represents a 15-minute interval. The following paragraph examines similar prediction plots that feature only five time series, chosen to represent general trends and compare the performance of the developed approaches with OCTOPUS and the impact of using hindcast versus nowcast data. This analysis seeks to identify overarching patterns and evaluate the general efficacy of the approaches. The subsequent subsection will then delve into a more detailed analysis of individual time series that differ from these general trends.

Contrasting ML-Based and Conventional Prediction Performance The results of the time series in Figure 32 are in line with the general trend observed in all time series in Appendix C. ML-based approaches show higher prediction accuracy than (simulated) conventional approaches. This can, for example, be seen in time series 57 to 60. Time series 56 is an exception, with the performance gap between the two approaches narrowing in the first half. All approaches mainly follow the overall trend in the data, rather than local fluctuations, which could be due to a shared limitation in capturing rapid motion changes. The similarity in performance between NauticalNet and NauticalNet* across these series suggests that the use of different data types (hindcast vs. nowcast) has little effect in this context, a trend that is also seen between OCTOPUS and HindHold. However, the general trend of the impact of using different data types does not match this finding.

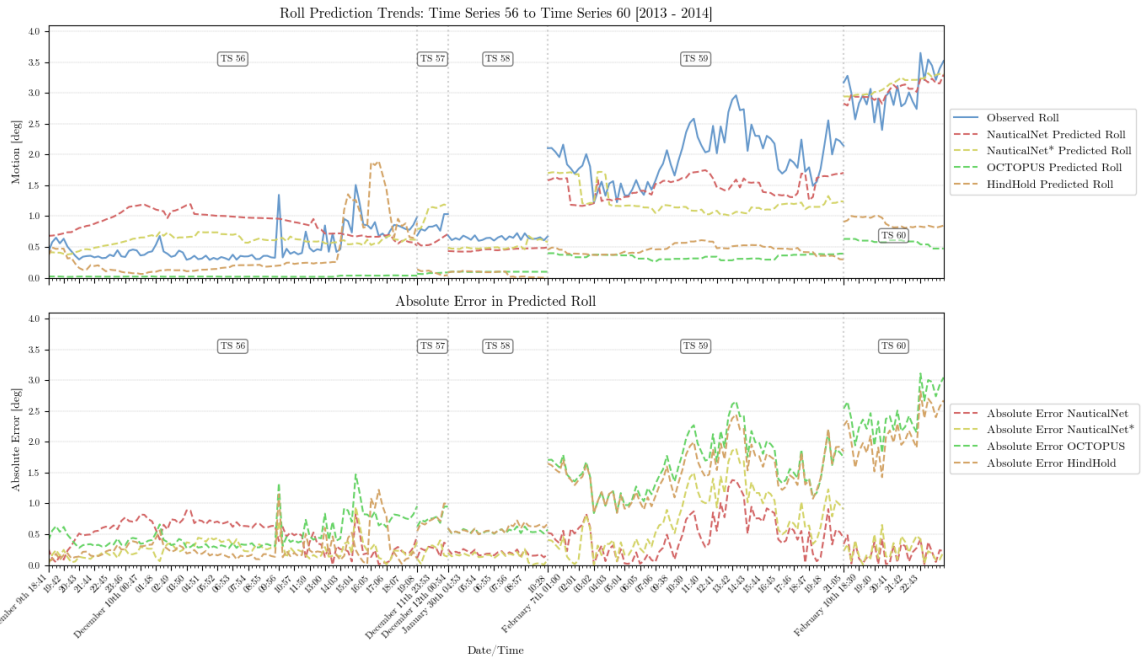


Figure 32: Comparative Analysis of Roll Motion Predictions vs. Actual Data in Time Series 56-60: Evaluating ML-Based Approaches Against Conventional Methodologies.

Hindcast vs. Nowcast Data: Impact on Prediction Accuracy In the second set of time series (Figure 33), a clear pattern is seen in the use of hindcast versus nowcast environmental data. This set

was chosen to demonstrate the improvement in prediction accuracy when using hindcast environmental data, particularly for OCTOPUS and HindHold. Hindcast environmental data leads to a significant decrease in error rates compared to nowcast environmental data. For ML-based approaches, the hindcast environmental data set includes additional features such as wave slope and swell partitions, which are not present in the nowcast environmental dataset, providing a more comprehensive context for prediction. However, it should be noted that this trend is not consistent across all time series in Appendix C. It can also be observed that across all approaches, a common pattern is the tendency to approximate a general trend in the data rather than to capture specific local fluctuations. This suggests that increasing the sensitivity to immediate environmental changes could improve the accuracy of prediction. It could also be a limitation of the environmental data, as this data type is sampled every hour and then linearly interpolated in time every 15 minutes.

Figure 33: Comparative Analysis of Roll Motion Predictions Using Nowcast vs. Hindcast Environmental Data: Time Series 176-180. This figure illustrates the variation in prediction accuracy between (simulated) conventional and ML-based approaches, highlighting the enhanced performance with hindcast data.

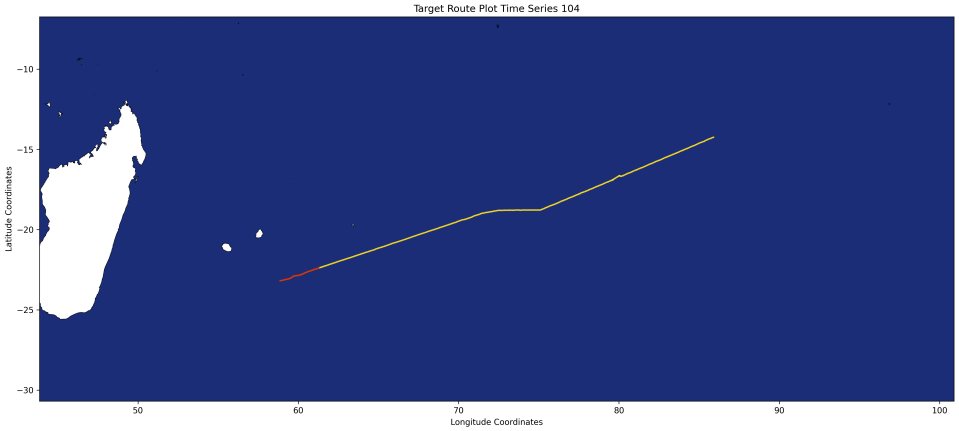
In conclusion, the results of validation strategy I seem to suggest that the developed approaches are more accurate in predicting maximum roll motion than OCTOPUS. Furthermore, the use of hindcast environmental data tends to produce more accurate predictions than nowcast data. However, the type of predictive approach appears to have a greater impact on the accuracy of the predictions than the type of environmental data used. The complexity of the interactions between data quality, environmental factors, and the capabilities of the approach is highlighted by the variation in the predictions across

different time series. Further investigation of these elements is planned in the following subsection.

Analysis by Conditions and Datatypes

In the previous section, the general observations and trends in the comparison of the prediction accuracy of the four approaches were discussed. The trend mentioned in the preceding subsection showed a general better performance of developed ML-based approaches compared to OCTOPUS and that one should prefer the utilization of hindcast- over nowcast environmental data. This subsection is devoted to examining time series that diverge from these trends, thereby describing how different conditions impacted the predictions and their trends, particularly for the maximum roll motions.

Deviation of Trend in Approach Performance - Time Series 104 Out of the 191 time series studied, two stand out from the general pattern of ML-based approaches outperforming OCTOPUS. These two cases, where OCTOPUS provides more accurate predictions, should be further investigated. Time series 104 is particularly noteworthy. The route for this time series is shown in Figure 34(a), with the test data points marked in red. The sea depth at the designated red test points is greater than 1000 [m] (the threshold for this feature is not greater than 1000 [m]). Figure 34(b) reveals that for Time Series 104, OCTOPUS, is far more accurate than ML-based approaches. NauticalNet and NauticalNet* are at some points 4 degrees off. Both approaches tend to underestimate the observed maximum roll motion. Additionally, for OCTOPUS and HindHold the use of nowcast data appears to be slightly better, while for the ML-based approaches the use of hindcast data appears to be slightly better.



(a) Geographical Route of Time Series 104 with Test Data Points Marked in Red)

(b) Comparative Predictions by ML-based and (simulated) conventional approaches for Time Series 104

Figure 34: Detailed Visualization of Time Series 104: Route and Predictive Analysis

An in-depth examination of the environmental conditions during Time Series 104 reveals why OCTOPUS probably outperformed ML-based approaches in this case. The magnitude of the maximum roll motion observed in this series was exceptionally high, reaching a peak of 10.71 degrees, as illustrated in Figure 16(b). This is the highest maximum roll motion recorded in the entire data set (see also

Figure 16 in Chapter 3). ML-based approaches, which are trained on a broader range of data where such extreme response values are uncommon, tend to underestimate in this situation. This conservative prediction pattern is likely due to the models' inherent design to minimize the loss function, as discussed in Chapter 2, Section 2.3, which leads them to avoid predicting rare extreme response values.

In contrast to ML-based approaches, OCTOPUS relies more heavily on theoretical models of vessel dynamics. This appears to be advantageous in this particular situation. It appears that the key factor is the alignment of the peak wave period with the vessel's natural roll frequency. The natural roll period of Time Series 104 was calculated to be 15.14 seconds, and the peak wave period, which was mainly composed of swell waves in these deep-sea conditions, was found to be around 15.02 seconds, as seen in Figure 35. This close alignment suggests a resonance effect, where the wave energy spectrum and the vessel's RAO peak coincide, resulting in increased roll motions. OCTOPUS, which may be better able to estimate the vessel's response based on theoretical RAO, particularly in resonant conditions, thus possibly therefore provides a more accurate prediction for this time series. The density distribution of the peak wave period, which indicates the frequency of occurrence of these wave conditions, is in agreement with the energy density spectrum of the waves, further confirming the resonance hypothesis.

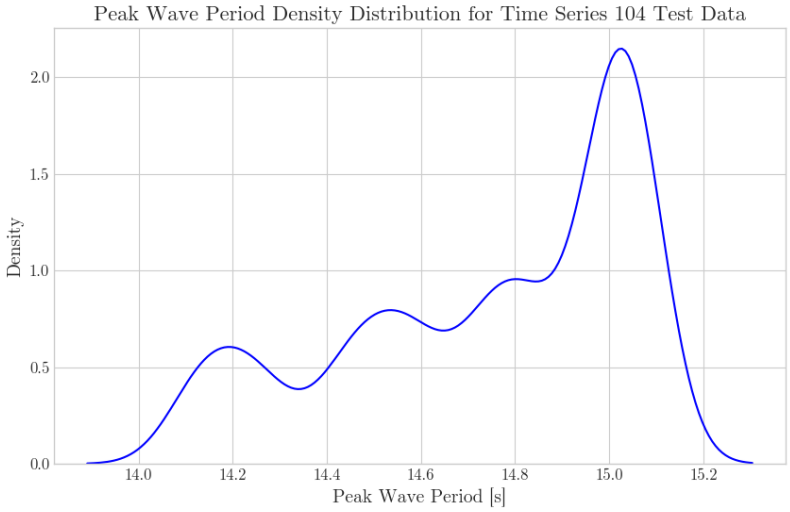
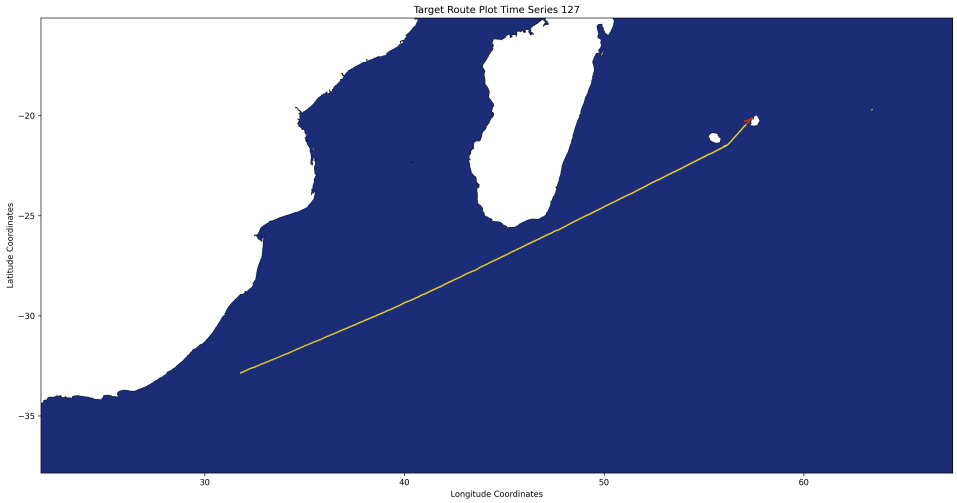


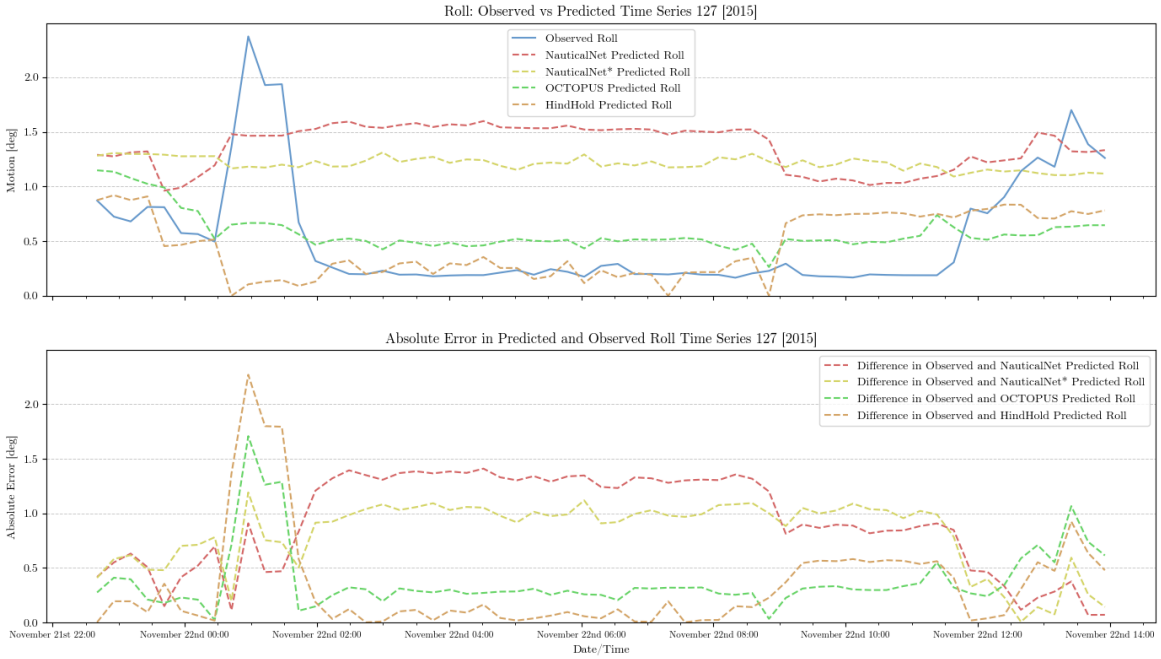
Figure 35: Density Distribution of Peak Wave Periods in Time Series 104: Illustrating Resonance with Vessel's Natural Roll Frequency

Deviation of Trend in Approach Performance - Time Series 127 The second notable divergence from the general trend is observed in Time Series 127, which is geographically close to Time Series 104. This series documents HTV Target's journey from South Africa, passing Madagascar, and approaching Mauritius, with the test data set capturing its maneuvers near the Mauritius shoreline. Figure 36(a) delineates this route, highlighting the test data points in red. A key observation here is

the vessel's abrupt change in direction during the test set, an anomaly compared to the rest of the voyage, predominantly in deep-sea conditions, as indicated in the training data (colored yellow).



(a) Geographical Route of Time Series 127 with Test Data Points Marked in Red)



(b) Comparative Predictions by ML-based and (simulated) conventional approaches for Time Series 127

Figure 36: Detailed Visualization of Time Series 127: Route and Predictive Analysis

The prediction results for time series 127, as shown in Figure 36(b), clearly demonstrate the higher accuracy of OCTOPUS and HindHold compared to NauticalNet and NauticalNet* in this particular

scenario. Contrary to Time Series 104, where purely ML-based approaches tended to underpredict, here they are prone to overpredicting the observed maximum roll motions. The choice between hindcast and nowcast environmental data seems to have a negligible impact in this context. Notably, none of the approaches successfully predicted the local peak observed between November 22nd 00:00 and November 22nd 02:00, suggesting a nonenvironmental factor at play, potentially linked to abrupt navigational changes.

The different conditions of the test data in Time Series 127 pose a unique challenge. Here, HTV Target navigates at significantly lower speeds in shallower waters, in stark contrast to the predominantly high-speed deep-sea conditions of the training data. This discrepancy likely introduces a bias in ML-based approaches, which have been conditioned to associate certain LCs with higher motion responses. This assumption, drawn from training data, leads to inaccuracies when faced with atypical conditions during the test phase. This theory gains support from the observed roll and pitch motion data for the entire series, shown in Figure 37, where the last 10% of the series (the test set) differs markedly from the initial 90% (the training set).

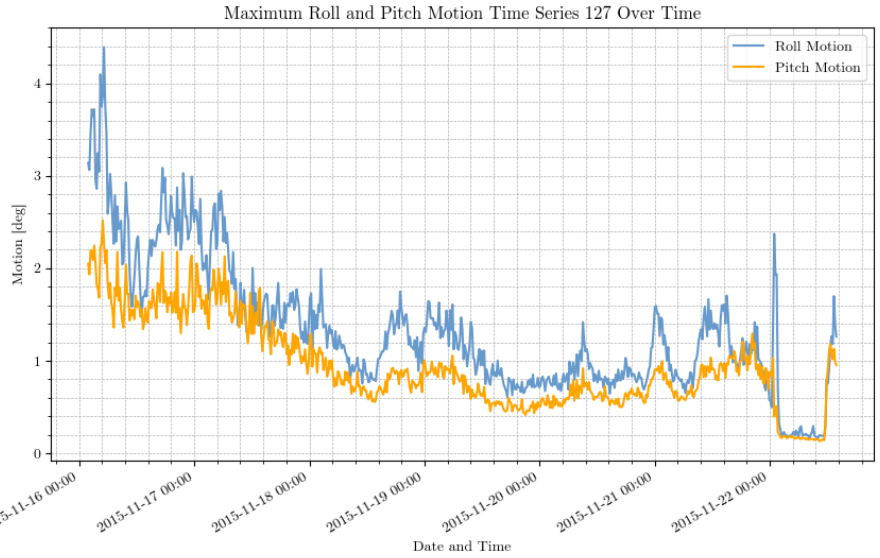


Figure 37: Observed Maximum Roll and Pitch Motion Time Series 127

In contrast to ML-based approaches, OCTOPUS does not require learning from past data and is, in theory, independent of the training data. In the case of Time Series 127, OCTPUS and HindHold have shown higher accuracy than is usually seen. This improved performance could be due to the particular conditions in the test data set, where the vessel is navigating close to shore with relatively high wave frequencies.

The speed of the vessel may have an effect on the accuracy of predictions made by (simulated) conventional approaches. This is evidenced by a correlation analysis, as shown in Figure 38, which reveals a positive correlation between vessel speed and prediction errors for (simulated) conventional approaches. Pearson's coefficients of 0.27 and 0.23 for OCTOPUS and HindHold, respectively, suggest that errors in (simulated) conventional, predictions tend to increase with higher vessel speeds. Interestingly, this trend is not as evident in ML-based approaches, indicating that vessel speed has a different impact on the accuracy of predictions between the two types of approaches. This correlation implies that (simulated) conventional approaches may be more accurate when the vessel is travelling at a lower speed than the average speed of 11.45 knots in the entire data set, which was 3.61 knots in the Time Series 127 test set.

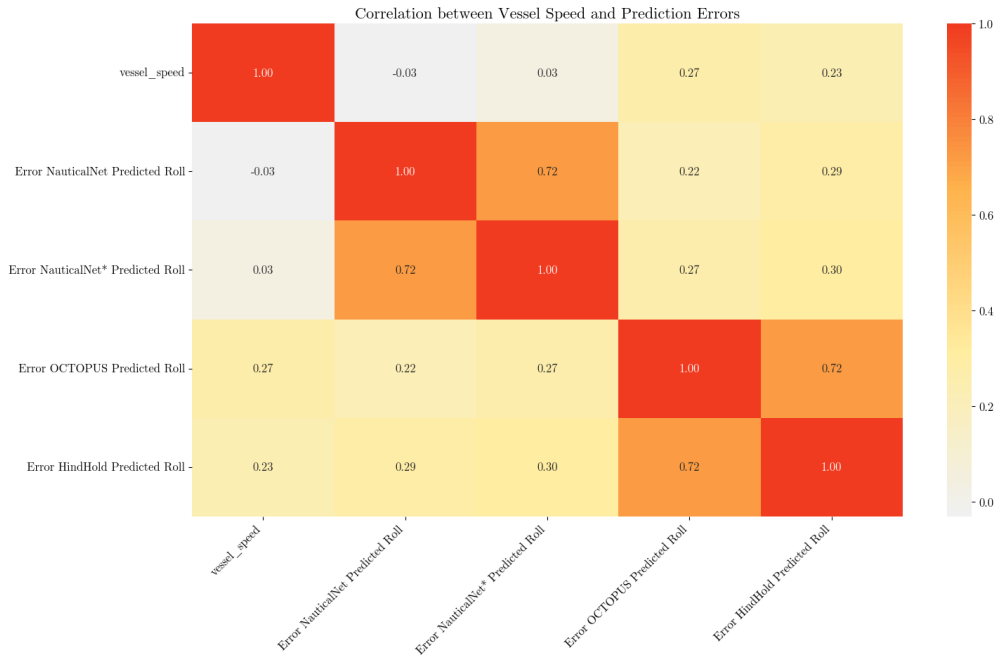
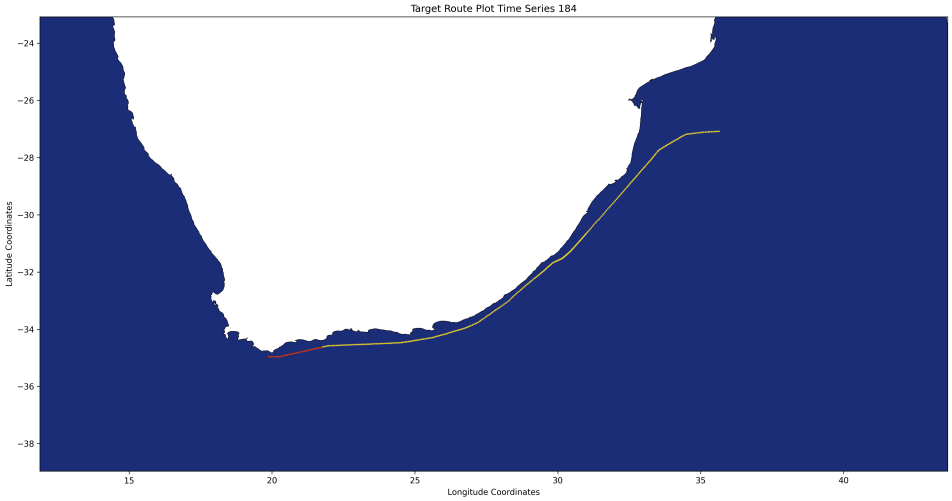


Figure 38: Correlation between Vessel Speed and Prediction Errors in Maximum Roll Motion

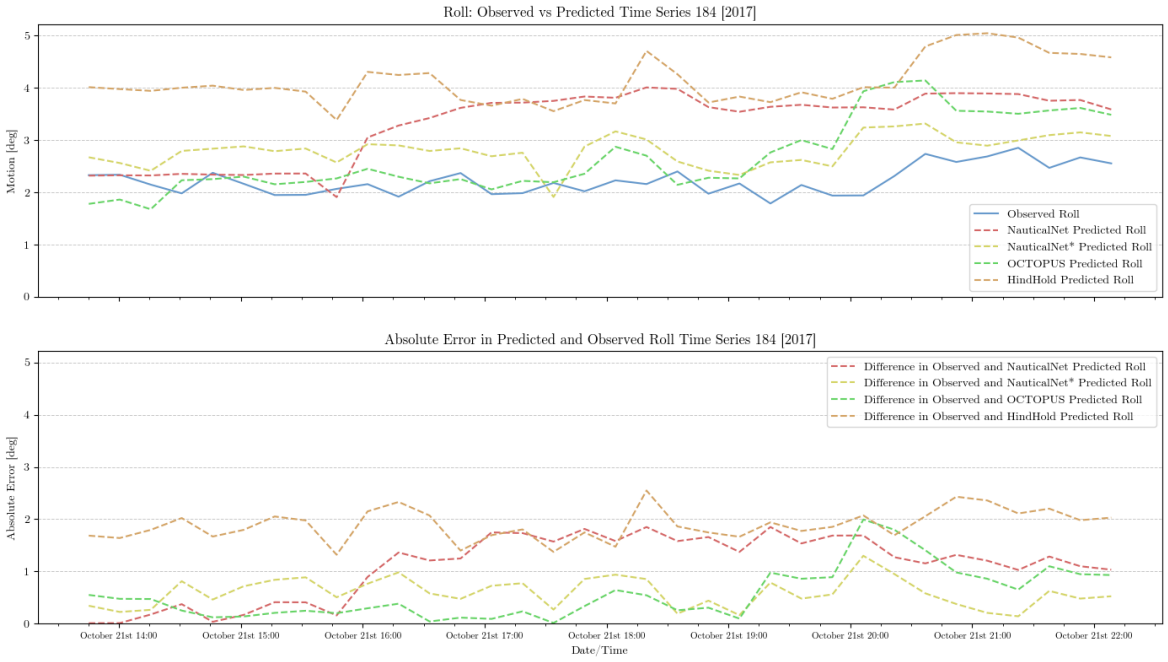
It is essential to recognize that, although these arguments offer an understanding of the effectiveness of prediction techniques in certain circumstances, they do not completely explain the observed divergence in trend. These efforts to explain the divergence are not definitive and imply that other elements may also be involved and associated with prediction errors. Further exploration into these elements is necessary, as will be discussed in the Discussion section.

Deviation of Trend in Datatype Employed - Time Series 184 Time Series 184, which follows HTV Target's course along the African shoreline, shows a remarkable divergence from the general trend towards hindcast environmental data. In contrast to the overall results, this particular time series reveals that the predictions based on nowcast environmental data (by OCTOPUS and NauticalNet*)

are more precise than those using hindcast data (HindHold and NauticalNet). This distinction is clearly demonstrated in Figure 39(b), where the overestimation of maximum roll motion using hindcast data implies a potential inconsistency in the quality or applicability of this type of data for this particular time series.



(a) Geographical Route of Time Series 184 with Test Data Points Marked in Red)



(b) Comparative Predictions by ML-based and (simulated) conventional approaches for Time Series 184

Figure 39: Detailed Visualization of Time Series 184: Route and Predictive Analysis

It may be insightful to investigate why the hindcast data may not be performing as expected in this case. Examining the interpolation methods used is key. As described in Section 4.2.1, the hindcast data, which is sourced from a 0.5-degree grid, utilize nearest interpolation, which could lead to inaccuracies. This is evident around the 10th data point in Time Series 184, as seen in Figure 39(b), where a sudden decrease in prediction accuracy is observed at the same time as a change in the chosen environmental data point. This shift, depicted in Figure 40, raises questions about the reliability of the hindcast data at certain navigational points. Although this does not definitively demonstrate the superiority of nowcast data, the trends in Figure 39 appear to favor it.

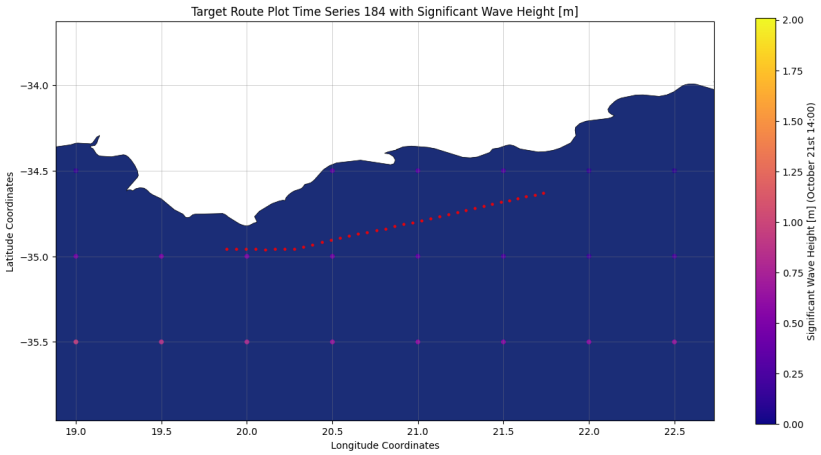


Figure 40: Spatial Distribution of Hindcast Significant Wave Height (in meters) Across the Environmental Data Grid During Time Series 184.

Investigating further, a comparison of the significant wave height and peak wave period between hindcast and nowcast data – two characteristics that have a major impact on the wave spectrum – reveals considerable differences (Figure 41). Hindcast data usually predict higher significant wave heights than nowcast data for time series 191, which is in line with the overestimation of maximum roll motion when using hindcast data. This implies that the forecast data probably overestimate the wave height predictions. Further distinctions between the two data types and their implications will be discussed in the following chapter.

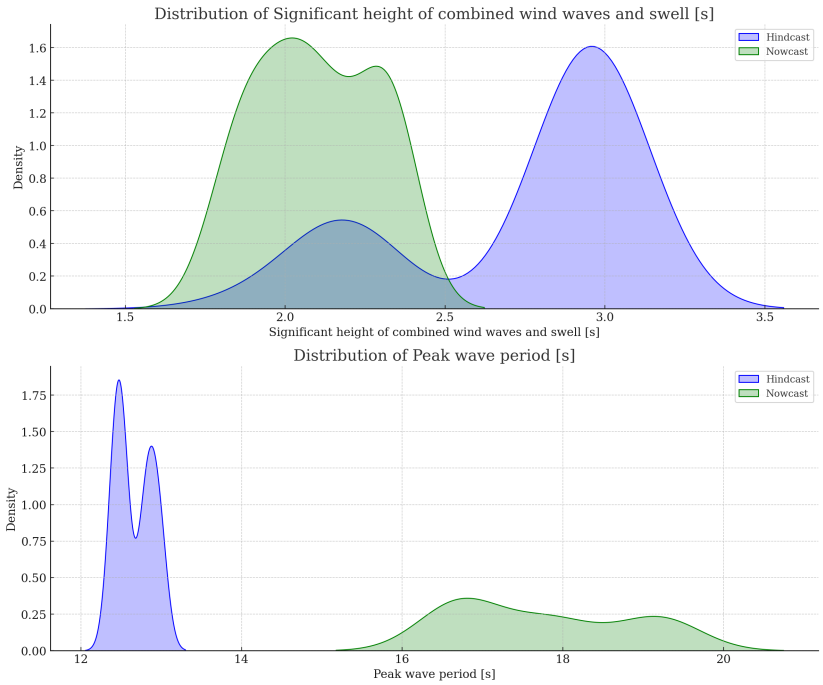


Figure 41: Comparison of Nowcast Environmental Data and Hindcast Environmental Data for Time Series 184

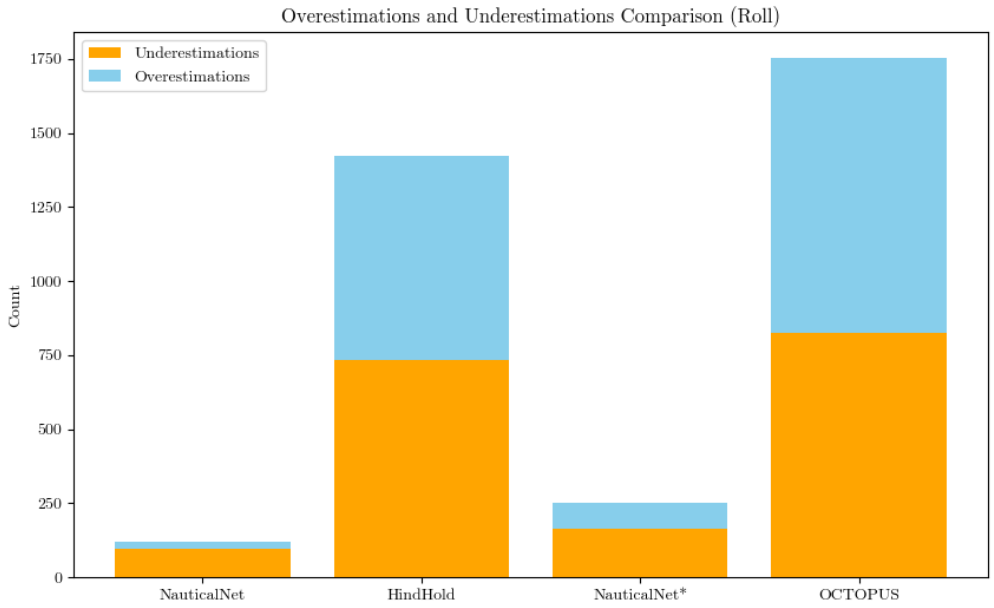
Analysis of Overestimations and Underestimations

As defined in Chapter 2, overestimations and underestimations occur when predictions deviate more than one standard deviation from the mean of a specific time series. The aggregation of overestimations and underestimations over all time series yields a comprehensive view of the predictive accuracy of each of the four approaches. Figure 42(a) presents a bar graph describing these deviations, categorizing the approaches according to the type of environmental data used – hindcast on the left and nowcast on the right. The chart colorcodes underestimations in orange and overestimations in blue, facilitating a direct comparison across methodologies.

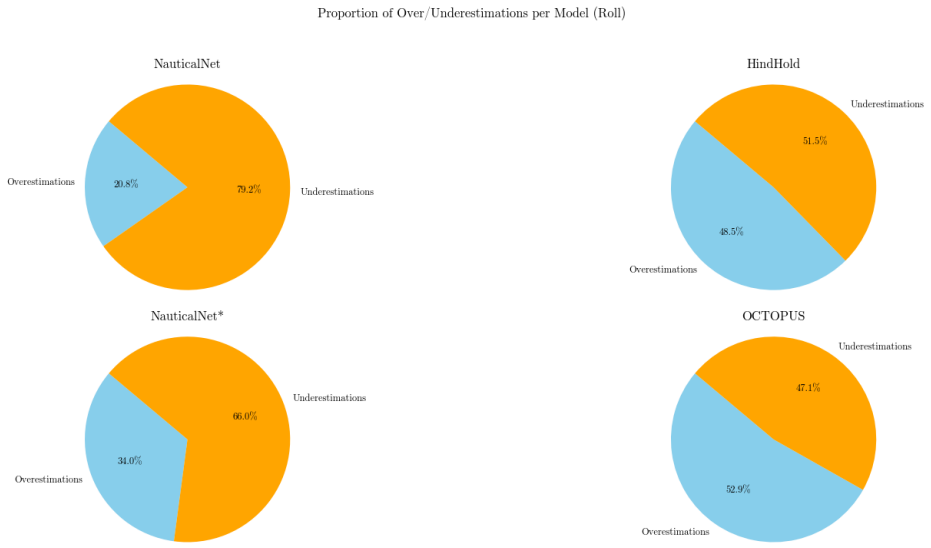
An interesting observation from this analysis is the difference in the total number of overestimations and underestimations between ML-based approaches and (simulated) conventional approaches. ML-based approaches have significantly less overestimations and underestimations than (simulated) conventional approaches. This probably suggests a higher prediction accuracy for the ML-based approaches. There is also a notable difference between using nowcast environmental data and hindcast environmental data. From the barplots it can be observed that using nowcast environmental data tends to increase the total number of overestimations and underestimations for both type of prediction approaches.

In terms of nuances of overestimations versus underestimations, Figure 42(b) shows the proportional

differences between the four approaches. This pie chart reveals that the choice of environmental data, hindcast or nowcast, does affect the balance of over-/underestimation in ML-based methods, while it does not affect the balance over-/underestimations in (simulated) conventional approaches much. In this case, the use of the hindcast environmental data seems to result in a higher relative proportion of underestimations for ML-based approaches. While this does not lead to a definitive conclusion about the role of data type, a potential description could be that by increasing the environmental data quality (by choosing hindcast over nowcast), the total number of overestimations and underestimations lowers, but mainly overestimations decrease w.r.t. underestimations. This could potentially be due to the custom loss function, which gives more weight to higher observed values. Also interesting to note is that (simulated) conventional approaches are more balanced in their predictions. Looking back at the discussion on time series 104, this could maybe indicate why ML-based approaches are tended toward underestimations. However, this remains an inconclusive argument and therefore should be approached with care.



(a)



(b)

Figure 42: Maximum Roll: Over- and Underestimation Characteristics of the Four Prediction Approaches

Performance Metrics

Earlier discussions in this section have indicated a preference for ML-based approaches over (simulated) conventional approaches, and for hindcast over nowcast environmental data in predicting maximum

roll motion. These trends are further supported by analyzing the performance metrics introduced in Chapter 3, Section 3.2. The metrics, summarized in Figure 43, reinforce the previous observations, with ML-based approaches showing superior performance compared to (simulated) conventional approaches. Additionally, the metrics suggest marginally better performance for approaches using hindcast environmental data than those using nowcast data.

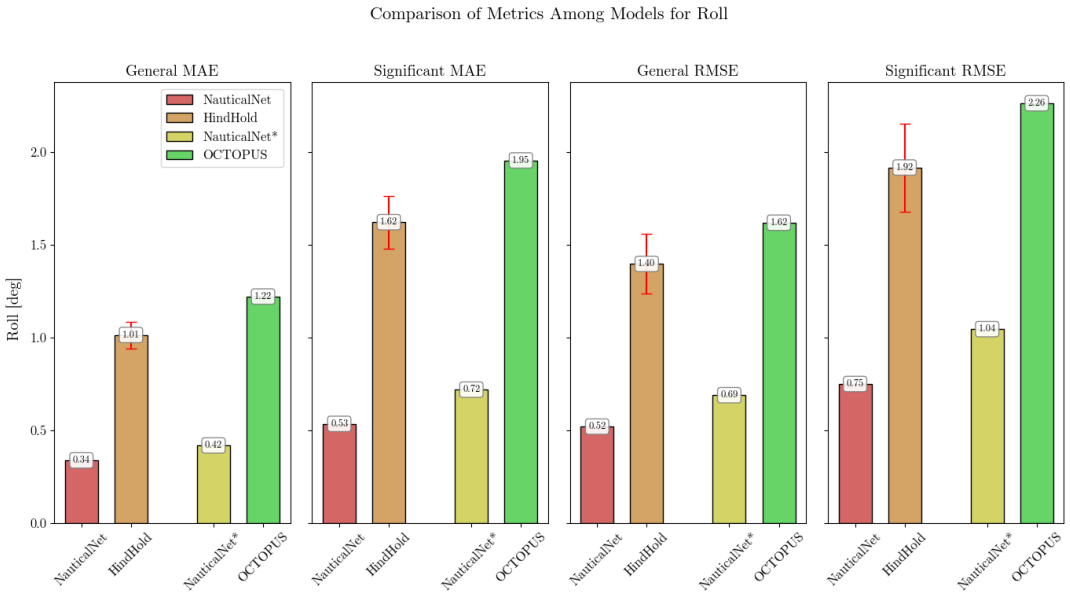


Figure 43: Comparison of Prediction Approaches on RMSE and MAE in Predicting the Maximum Roll Motion

In detail, NauticalNet, which uses hindcast environmental data, achieves the lowest MAE at 0.34 degrees, closely followed by its nowcast variant, NauticalNet*, with 0.42 degrees. The (simulated) conventional approaches, HindHold and OCTOPUS, exhibit higher errors of 1.01 degrees and 1.22 degrees, respectively. Notably, HindHold's performance is slightly better than that of OCTOPUS, and the accompanying error bar in its orange bars (indicating the errors when mirroring OCTOPUS as discussed in Section 4.2.3) adds a layer of uncertainty to its results. The RMSE metric, which gives more weight to higher individual errors, is consistent with these findings, despite the presence of higher magnitudes of the errors. Furthermore, in the significant zone (the top one-third of the data points), the performance gap between the ML-based and (simulated) conventional approaches widens for the approaches utilizing hindcast data but narrows for the approaches using nowcast data.

Further analysis using the Mean Absolute Percentage Error (MAPE), as shown in Figure 44, provides information on the relative errors. This metric evaluates errors in relation to observed values, offering a perspective on the proportionality of errors. In the general case, ML-based approaches exhibit a MAPE of 40% - 45%, significantly lower than the MAPE of 85% - 100% observed in conventional

methods. When focusing on significant data points, the performance gap remains, with ML-based approaches decreasing to around 20% MAPE, while (simulated) conventional approaches drop to about 60%. This confirms the preference for ML-based approaches over (simulated) conventional approaches in this dataset. Regarding the choice of environmental data, while hindcast data slightly edges out nowcast data in terms of error minimization, this difference is less pronounced compared to the choice between ML-based and (simulated) conventional approaches.

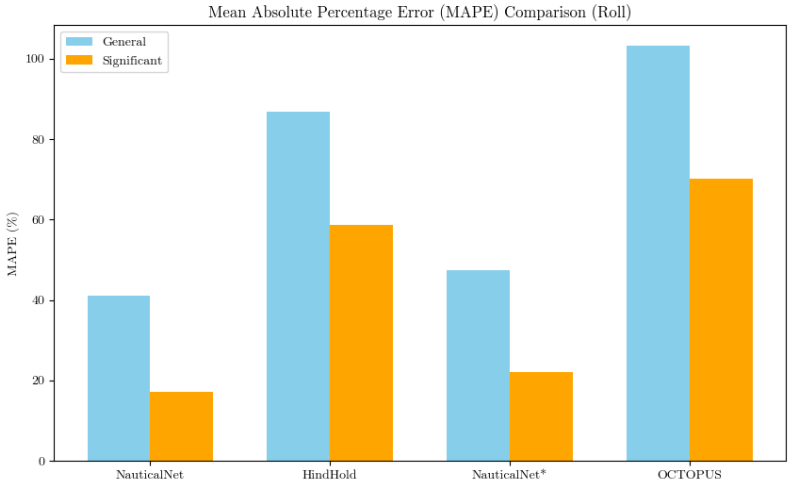


Figure 44: Comparison of Prediction Approaches on MAPE in Predicting the Maximum Roll Motion

5.1.2 Maximum Pitch Motion Prediction Results

This subsection examines the predictions of the maximum pitch motion. The same general trends that were observed in the previous subsection regarding the maximum roll motion are also present here. To avoid repetition, this subsection will be more concise in terms of analysis and discussion. Appendix D contains all maximum pitch predictions from the four approaches in the last 10% of each time series. This subsection will emphasize the prediction performances on maximum pitch motion by confirming the general trend, as also discussed in the previous subsection, in a prediction plot of multiple time series. It will then focus on any deviations from the general trends. This section will conclude with the analysis of overestimations and underestimations and the performance metrics.

Visual Insights: Maximum Pitch Motion Prediction Plots

In the analysis of maximum roll motion, as discussed in the previous subsection, two general trends were identified:

- Lower prediction errors are typically exhibited by ML-based prediction approaches compared to

(simulated) conventional approaches.

- The use of hindcast environmental data is found to be more accurate than nowcast data.

These trends were also observed in the analysis of the maximum pitch motion. Figure 45 shows the predictions for the maximum pitch motion for time series 121 to 125. It is noted that ML-based approaches, specifically NauticalNet and NauticalNet*, outperform (simulated) conventional approaches such as OCTOPUS and HindHold. This is especially true when hindcast environmental data are used, as evidenced in time series 121 and 122.

A comparison of the prediction plots in Appendix C and Appendix D reveals that the distinction between ML-based and (simulated) conventional approaches in predicting pitch motion is not as pronounced as it is for roll motion. All four approaches demonstrate an improved capacity to capture pitch motion, and the influence of different kinds of environmental data appears to be less important for pitch motion than for roll motion.

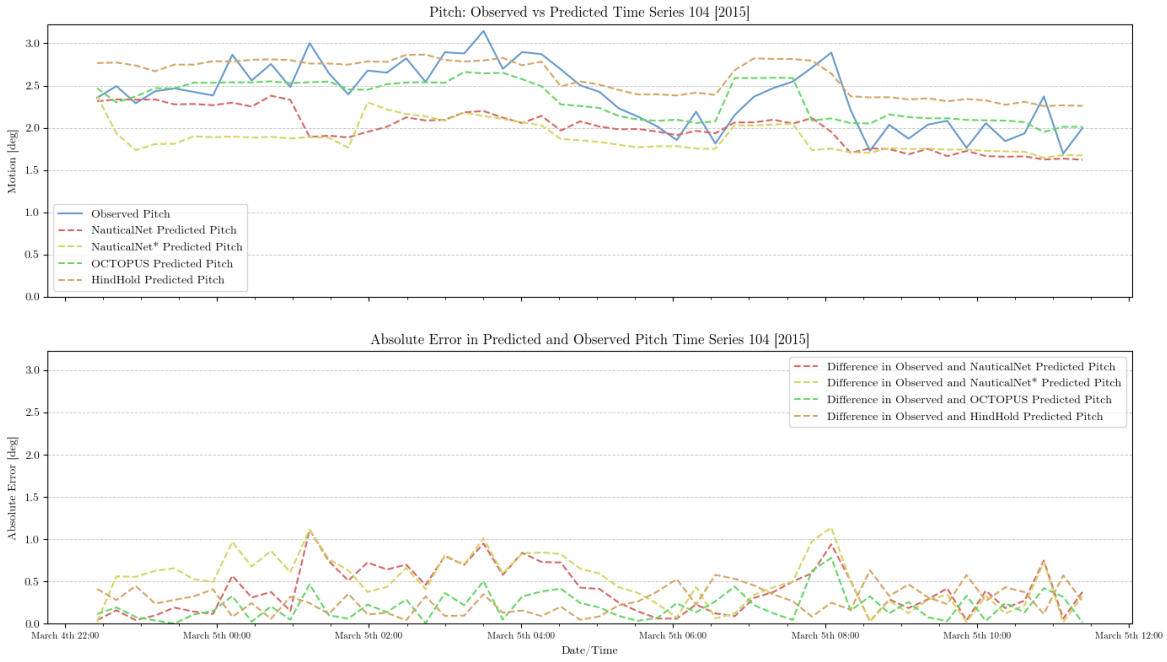
Figure 45: Comparing Pitch Predictions with Observed Motions: Time Series Analysis (121-125). This graph illustrates the effectiveness of ML-based approaches against (simulated) conventional approaches, highlighting the variations in nowcast and hindcast environmental data.

Analysis by Conditions and Datatypes

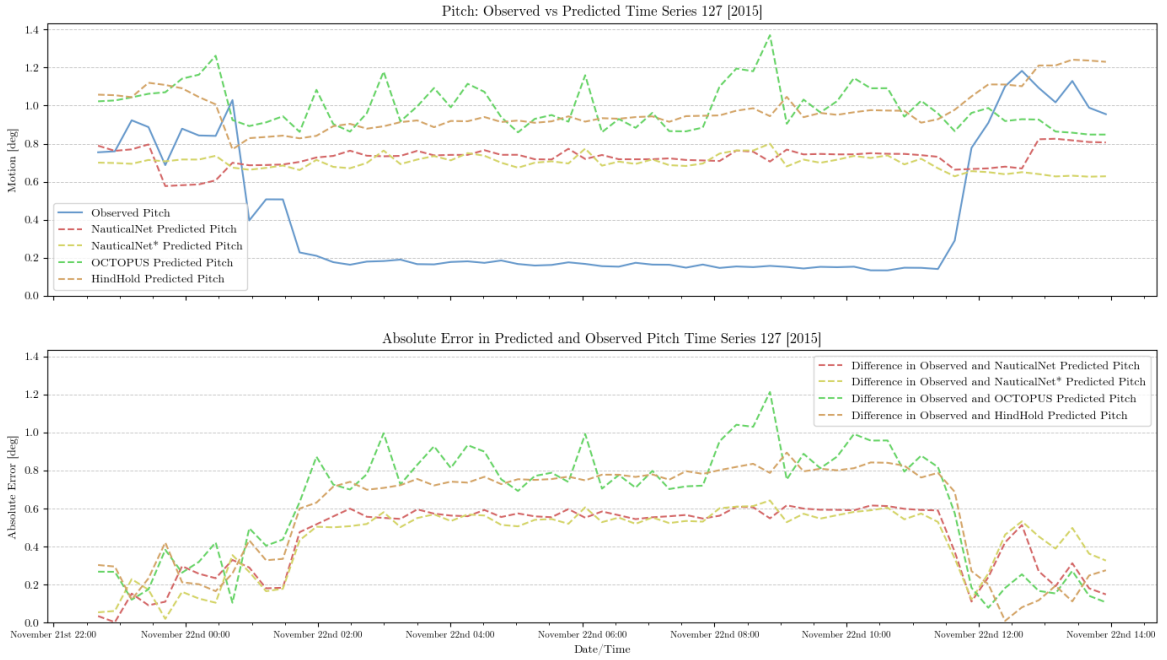
In the predictions for the maximum pitch motion, certain time series are identified as trend-breaking observations. For example, Time Series 104, as shown in Figure 46(a), is an example where (simulated) conventional approaches are found to be more successful than those based on ML. This deviation can

be attributed to two possible factors:

- An extreme value for maximum pitch motion is observed. As shown in Figure 16, the distribution of pitch motion within the data set is presented. Values exceeding 2.5 [deg] for maximum pitch motion are extremely rare. It should be noted that the current configuration of ML-based approaches, due to their training, is limited in predicting such extreme response values.
- The occurrence of an extreme value for maximum pitch motion is likely due to the overlap of pitch Response Amplitude Operator (RAO) and the wave spectrum. The mean peak wave period was previously calculated to be approximately 15 [s], while the natural pitch period is 13 [s]. This proximity suggests a possible resonance effect, where the wave energy spectrum and the vessel's RAO peak coincide, leading to heightened maximum pitch motions. In such resonant conditions, (simulated) conventional approaches, which are potentially more adept at estimating the vessel's response based on theoretical RAO, are observed to potentially provide more accurate predictions for this particular time series.



(a)



(b)

Figure 46: Maximum Pitch Motion Predictions for Time Series 104 and Time Series 127.

In the preceding section on the maximum roll motion, a deviation from the expected trends was discussed by analyzing time series 127, as referenced in 5.1.1. The pitch motion predictions for this

series, depicted in Figure 46(b), show a similar deviation. It is important to note that the reasoning for the inaccuracies in ML-based predictions for roll motion applies to pitch motion as well. Specifically, Figure 37 reveals a pronounced drop in observed motions in the final 10% of the data. This change is likely to be attributed to reduced vessel speeds, altered headings, and the transition to shallower waters. Such conditions were not represented in the ML-based approach's training dataset, which was skewed toward scenarios involving higher speeds and deeper waters. Consequently, as Figure 46(b) shows, ML-based approaches exhibit limited adaptability to these unforeseen conditions, underscoring a critical limitation that warrants further exploration in the following chapter.

In contrast, when it comes to predicting pitch motion, (simulated) conventional approaches yield less accurate results compared to ML-based approaches. This appears to be in contrast to the trends observed for the maximum roll motion, where conventional methods generally outperformed ML-based approaches for this time series. A potential explanation for this disparity can be found in the correlation between vessel speed and prediction errors, as illustrated in Figure 47. Interestingly, the correlation for pitch motion predictions is markedly lower for conventional methods than it was for roll motion, even tending towards a negative correlation. This suggests that lower vessel speeds may lead to higher errors in pitch motion predictions, a trend that is not observed in roll motion predictions. However, this interpretation is not definitive and should be approached with caution. It mainly indicates the distinct nature of pitch and roll motions and their respective influencing factors.

Figure 47: Correlation between Vessel Speed and Prediction Errors in Maximum Pitch Motion

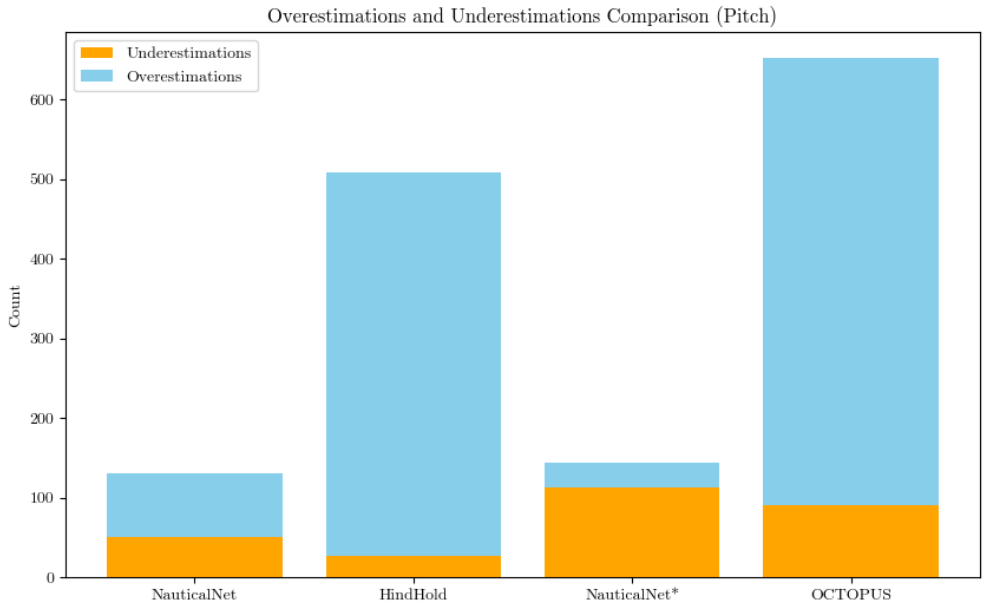
In Section 5.1.1, time series 184 demonstrated different responses to hindcast and nowcast environmental data for roll motion predictions, a pattern not as pronounced in pitch motion (see Figure

114 in Appendix D). This observation suggests a possible difference in the sensitivity of these motions to the quality of environmental data. Roll motion seems to be more affected by the accuracy of these data compared to pitch motion.

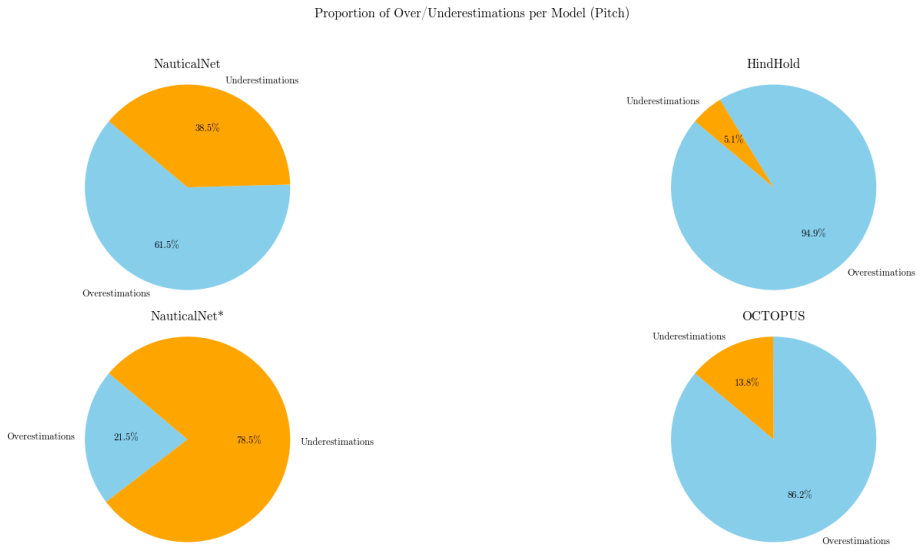
The nature of roll and pitch motions, as discussed in Chapter 2, can help explain the difference between them. Pitch motion is more heavily damped, resulting in lower amplitudes and less extreme responses. This higher damping level may reduce the impact of inaccuracies in the wave spectrum on pitch motion predictions. In other words, the damping in pitch motion could lessen the effect of errors in environmental data. On the other hand, roll motion has lower damping and is more prone to resonance effects. Therefore, inaccuracies in environmental data, such as the wave spectrum, are not as easily mitigated in roll motion. This means that the quality of the wave-spectrum data is more critical for accurately predicting roll motion. Errors in these data are not reduced as in pitch motion, but could be amplified due to resonance. This discussion suggests that, while both types of motion benefit from accurate environmental data, the need for precision is more pressing for roll motion due to its lower damping and susceptibility to resonance. Pitch motion, with its higher damping, may be less affected by minor inaccuracies in the wave spectrum data. However, it is important to note that this is a tentative conclusion based on the observed data and the known characteristics of these motions.

Analysis of Overestimations and Underestimations

The results of the four prediction approaches for maximum pitch motions are shown in Figure 48. Compared to the overestimation and underestimation of the maximum roll motion, the total number of overestimations and underestimations of the maximum pitch motion is lower for NauticalNet * and both (simulated) conventional approaches (see Figure 48(a)). NauticalNet* has approximately 150 overestimations and underestimations, while for roll it had around 250. This implies that the precision in predicting the maximum pitch is higher and that the predictions are more stable. NauticalNet has a similar total number of overestimations and underestimations. However, since 3 of the four approaches appear to have a lower total number of overestimations and underestimations, this could indicate a higher prediction accuracy of all four approaches for the maximum pitch motion. Figure 48(b) reveals interesting behavior. While for the maximum roll motion the (simulated) conventional approaches were evenly balanced, for the maximum pitch motion, they tend to overestimate it significantly. For NauticalNet, instead of having more underestimations for the maximum roll motions, it displays a higher number of overestimations for the maximum pitch motions. For NauticalNet*, the proportions with respect to the maximum roll motion remain relatively similar.



(a)



(b)

Figure 48: Over- and Underestimation Characteristics of the Four Prediction Approaches

Performance Metrics

The performance metrics summarizing the errors of the four approaches confirm the general trends observed in the data. Figure 49 shows the MAE and RMSE for both the general case and the significant

case for all four approaches. The significant thresholds are for the maximum roll ≥ 1.92 [deg], and for maximum pitch ≥ 0.81 [deg]. The ML-based approaches both outperform their conventional relative, based on environmental data type. NauticalNet is the best-performing approach having a 0.11 [deg] MAE. NauticalNet* performs almost as well, having an MAE of 0.13 [deg]. HindHold is arguably the third best performing approach, but with the error bar indicated in red, it makes almost no difference when compared with OCTOPUS. This result is similar for the RMSE of all four approaches. The difference between ML-based approaches and (simulated) conventional approaches is still quite significant. However, the difference between the use of hindcast environmental data and nowcast environmental data is significantly lower than it was for the maximum roll motion. This aligns with the earlier findings when analyzing time series 184, highlighting that the prediction of maximum pitch motion is probably less governed by errors in the environmental data. It also seems like that for the significant case the difference between the four approaches becomes relatively narrower.

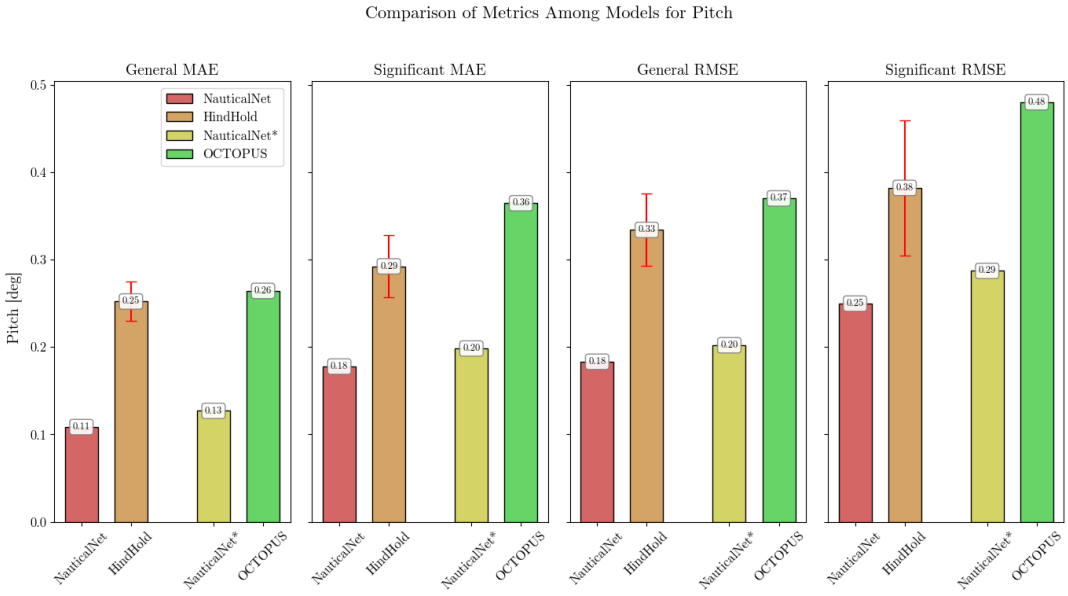


Figure 49: Comparison of Prediction Approaches on RMSE and MAE in Predicting the Maximum Pitch Motion

The MAPE, which describes the relative error with respect to the maximum pitch motion observed, is shown in Figure 50. This bar indicates that the ML-based approaches are also relatively better at predicting the maximum pitch motion than the (simulated) conventional approaches. It appears that all four approaches have a lower MAPE when predicting the maximum pitch motion than when predicting the maximum roll motion, suggesting that it is "simpler" to predict the maximum pitch motion than the maximum roll motion. Both types of approaches have almost half of the MAPE in the general case. For the (simulated) conventional approaches, this is also true for the significant

case, where the MAPE drops from 60%-70% to 25%-30% (with respect to the maximum roll motion). However, for the ML based approaches, this drop is less significant. For the maximum roll prediction the MAPE was around 20% for the significant case, and for the maximum pitch this is about 13%-14%. This could indicate that ML-based approaches outperform (simulated) conventional approaches relatively more in the roll predictions than in the pitch predictions in the significant domain. However, still ML-based approaches should be preferred over (simulated) conventional approaches for maximum pitch predictions.

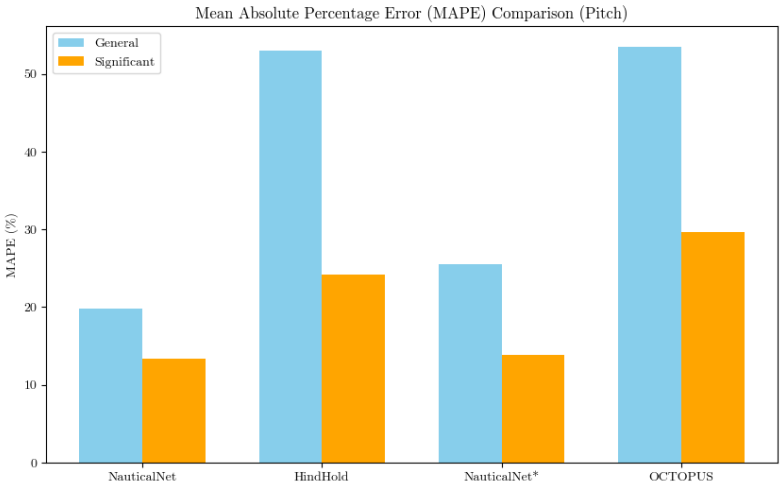


Figure 50: Comparison of Prediction Approaches on MAPE in Predicting the Maximum Pitch Motion

In general, ML-based approaches demonstrate lower prediction errors than (simulated) conventional approaches, as seen in Figure 51. This figure shows that for the majority of time series, both the maximum pitch and maximum roll, the ML-based approaches outperform the (simulated) conventional approaches, with the exception of time series 104 and time series 127. The ML-based approaches outperform the (simulated) conventional approaches for the maximum pitch motion, however, the difference is more significant for the maximum roll motion. This is likely due to the fact that pitch motion is a relatively highly damped motion, as the occurrence of pitch motion generates waves which take energy from the system, resulting in lower responses in general, and so also during resonance. On the other hand, roll motion is not a governed damped motion, but more a governed resonance motion, which is a high nonlinear phenomenon, making it harder to capture with (simulated) conventional approaches. The ML-based approach appears to capture these nonlinearities better. Limitations of ML-based approaches have also been identified in this section, mainly regarding the prediction of extreme maximum response values. This discussion will continue in Chapter 6 'Discussion'.

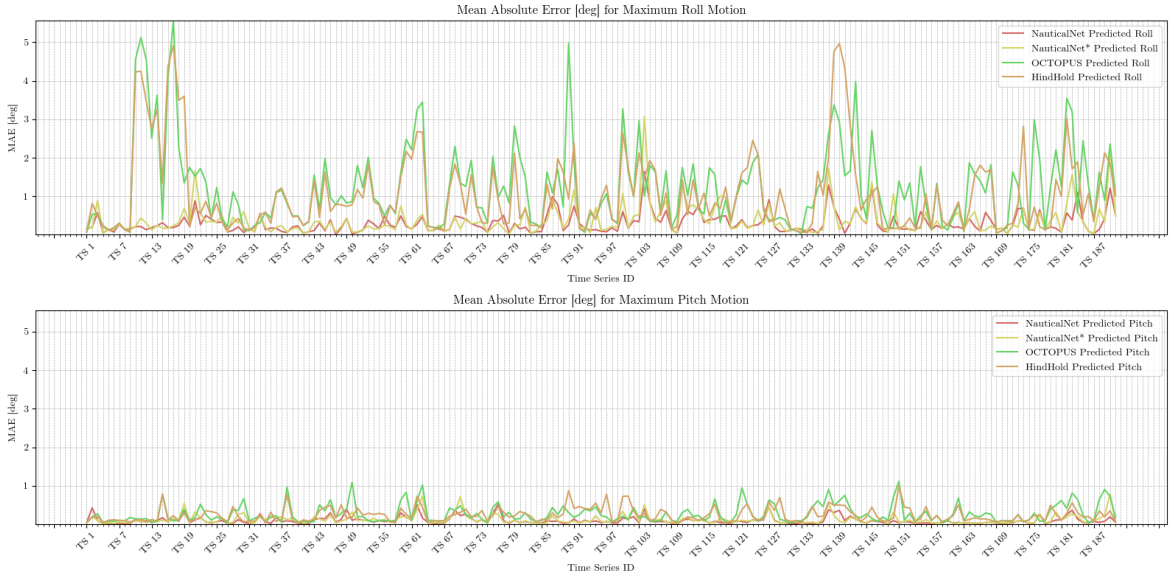


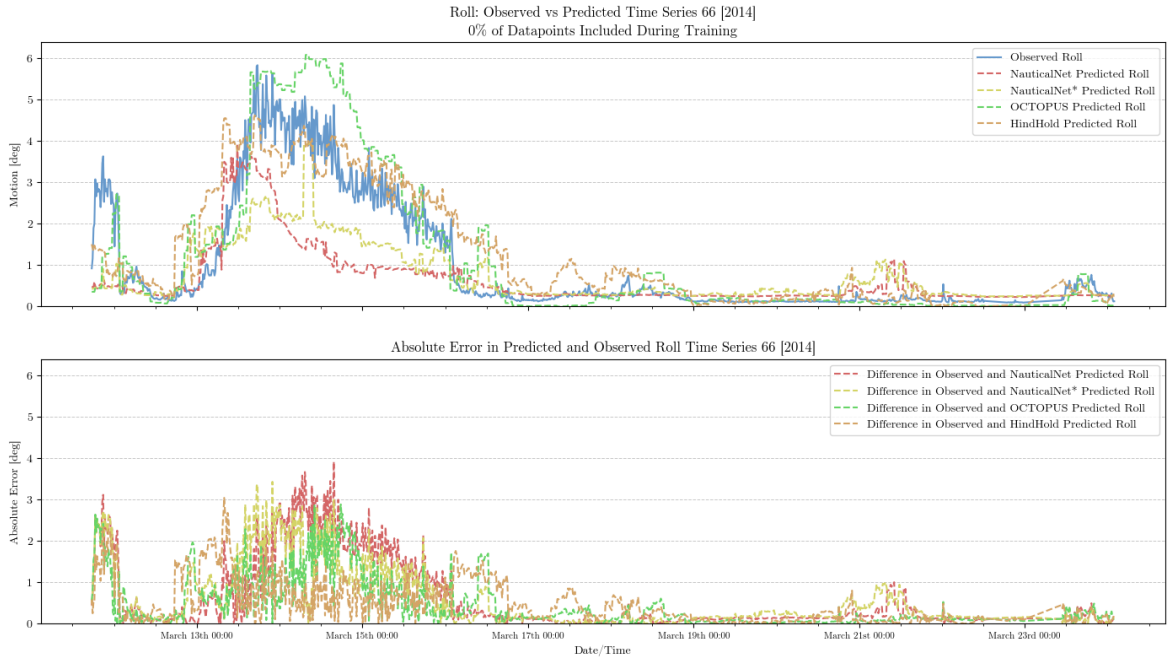
Figure 51: Comparison of Prediction Approaches on MAE over all Time Series

5.2 Validation Strategy II: Testing with Unknown LCs

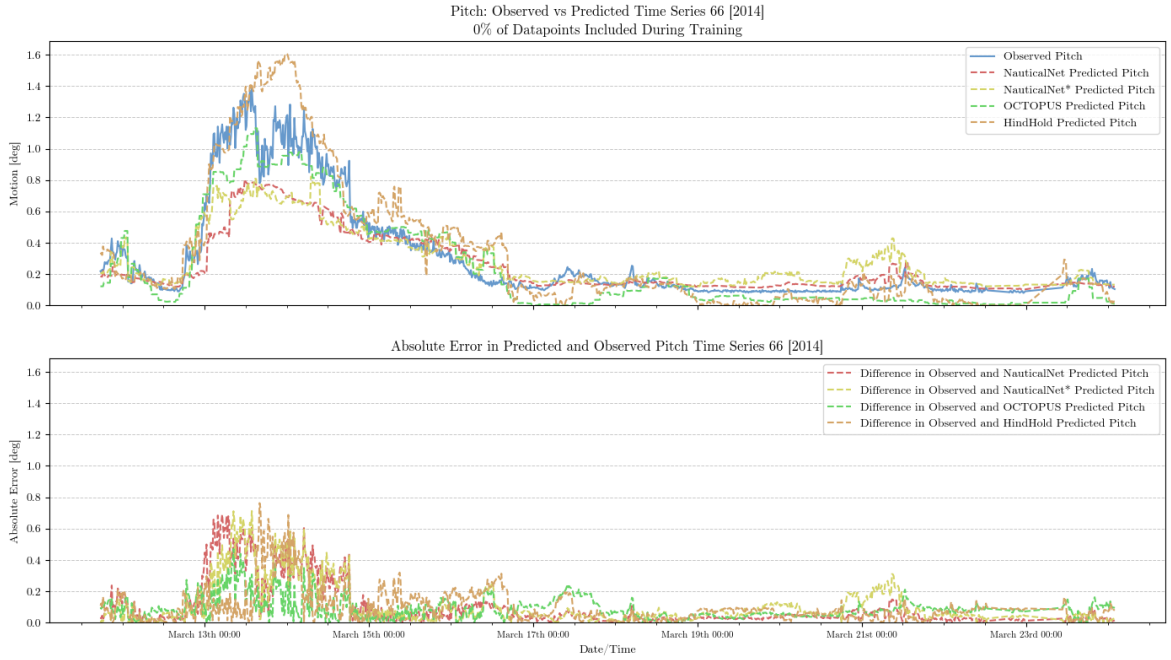
This validation strategy suggests that one LC is excluded in its entirety from training and validating the ML-based approaches. In Chapter 4, Section 4.3.3, LC7 was appointed to be excluded from the learning dataset. LC7 only contains time series 66. In this section, the results of this validation strategy will be analyzed.

5.2.1 Prediction Plots Maximum Roll and Pitch Motions

Using this validation strategy, (simulated) conventional approaches are found to outperform ML-based approaches, as evidenced in Figure 52. Notably, in the higher maximum motion zone (before March 15th 00:00), a significant underestimation by ML-based approaches of the maximum observed values is seen. On the contrary, (simulated) conventional approaches generally exhibit better performance. In the case of the maximum roll motion, as shown in Figure 52(a), the HindHold approach is seen to predict the maximum roll motion quite accurately. For maximum pitch motion, OCTOPUS appears slightly more effective. Overall, a tendency for (simulated) conventional approaches to overestimate the maximum observed motions in this time series is observed. In the latter part of the plot, to the right of the peak, all the approaches demonstrate relatively good performance. For maximum pitch motion, the ML-based approach might slightly outperform (simulated) conventional approaches in this part of the plot.



(a)

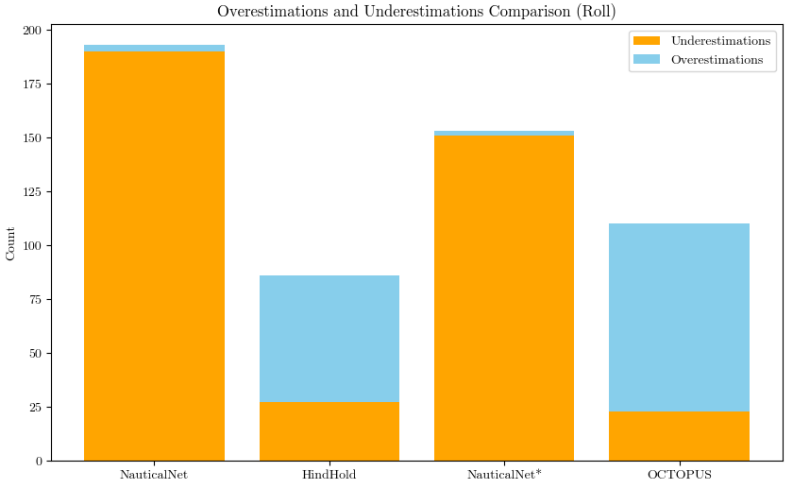


(b)

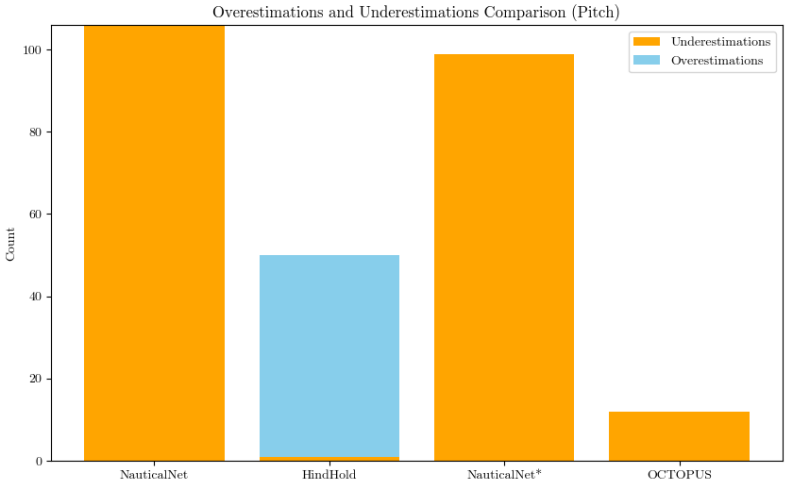
Figure 52: This figure demonstrates the difference between conventional and ML-based approaches when Time Series 66 is not included in the training set. It is evident that ML-based approaches tend to underestimate the maximum roll and pitch motions, while conventional methods, particularly HindHold for maximum roll motion and OCTOPUS for maximum pitch motion, are more accurate in their predictions.

5.2.2 Over- and Underestimations

Significant underestimations of the maximum motions observed are made by both ML-based approaches, as is evident in Figure 53. Overestimations are almost nonexistent for these ML-based approaches. Furthermore, it is observed that the total number of overestimations and underestimations for both ML-based approaches is higher, ranging between 150 and 190 out of 1100 total cases. This indicates a lower prediction accuracy for these approaches. In contrast, (simulated) conventional approaches are seen to have a tendency towards overpredictions, particularly in the case of maximum roll motion. In the context of maximum pitch motion, the total number of overestimations and underestimations is found to be lower, suggesting a better prediction accuracy for this parameter. The proportions of overestimations and underestimations are detailed in more detail in pie charts, presented in Appendix E.1.



(a)



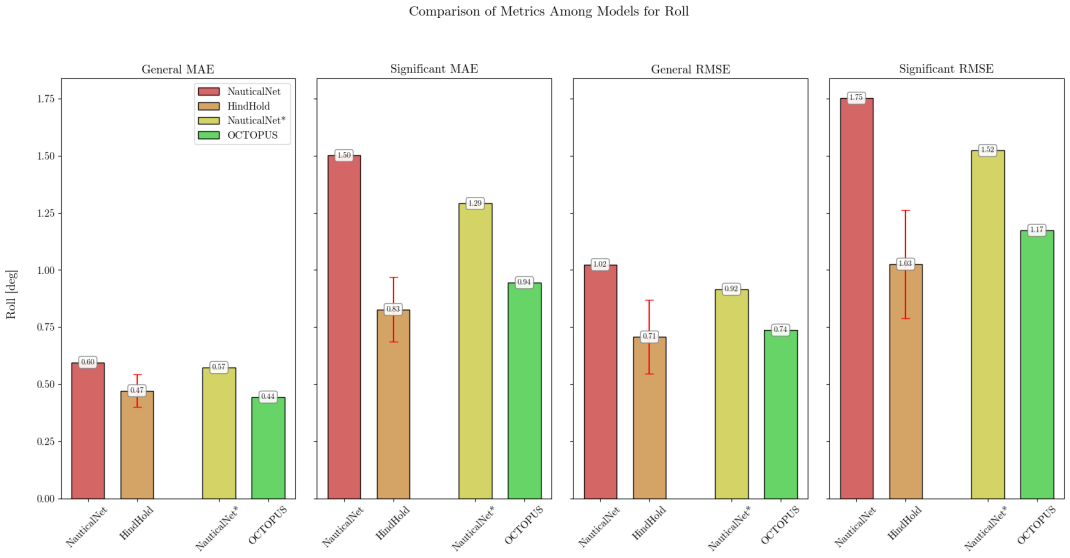
(b)

Figure 53: Over- and Underestimation Characteristics Maximum Roll and Pitch Motion [0% of LC7 Included]

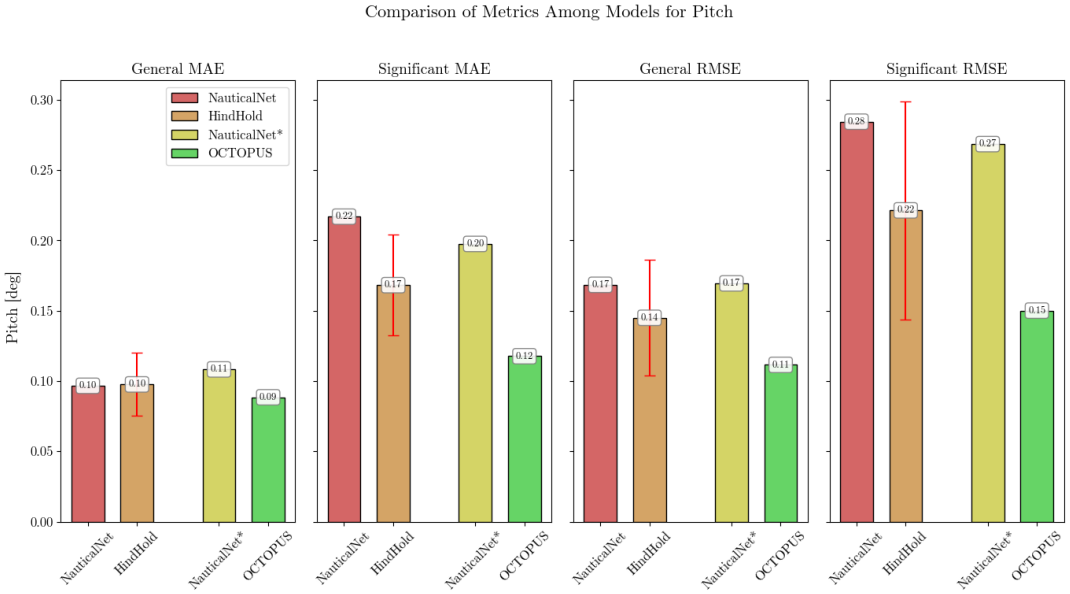
5.2.3 Evaluation Metrics

To validate the findings of the preceding subsections, the same evaluation metrics are used to compare the various prediction approaches. In this case, the significant thresholds are lower than in the first validation strategy. The threshold for maximum roll ≥ 0.55 [deg], and for maximum pitch ≥ 0.23 [deg]. Figure 54 displays the comparison of metrics between the four prediction approaches for the maximum roll and pitch motions. For maximum roll motion, both (simulated) conventional approaches

outperform ML-based approaches. It is noteworthy that NauticalNet, which uses hindcast environmental data, appears to have higher errors than NauticalNet*, which uses nowcast environmental data. This trend is not seen in (simulated) conventional approaches. A potential explanation could be that NauticalNet* has OCTOPUS predictions as an input feature, which is influential enough to make the predictions better than those from NauticalNet, which does not include OCTOPUS predictions. When comparing the general case with the significant case, all prediction approaches have higher absolute errors.



(a) Maximum Roll Motion

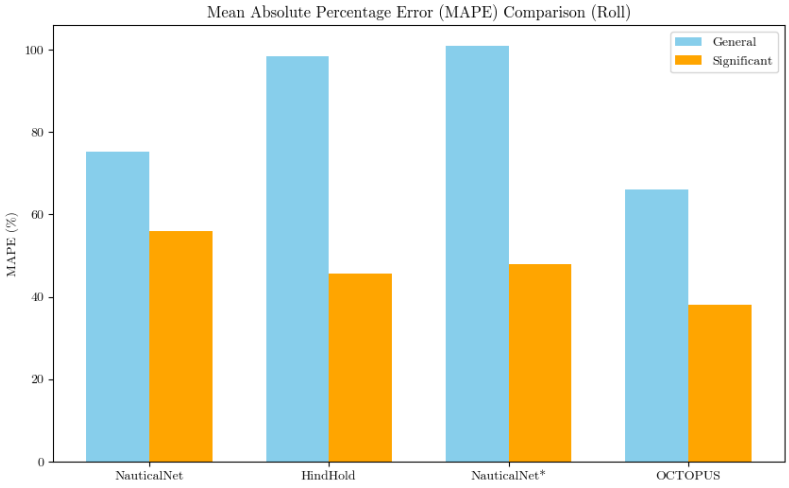


(b) Maximum Pitch Motion

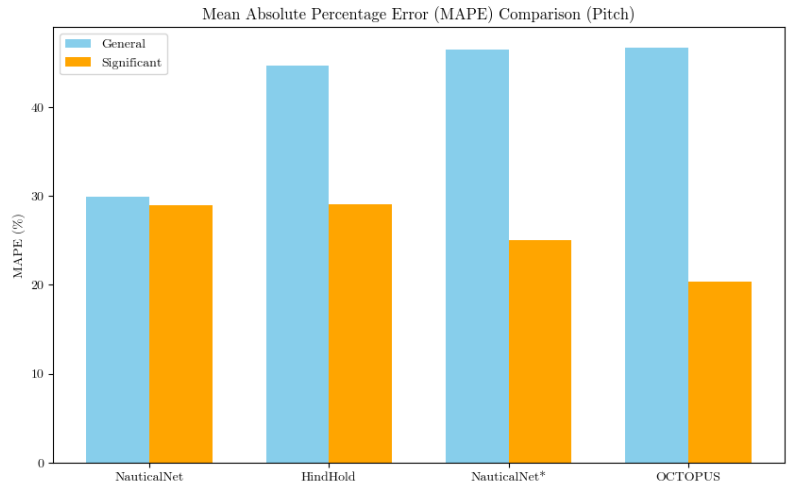
Figure 54: Comparison of Prediction Approaches on RMSE and MAE in Predicting the Maximum Roll and Pitch Motions

Figure 55 shows the MAPE of the two maximum motions. Comparing these results to those from earlier in the chapter, it is evident that both NauticalNet and NauticalNet performed worse than before. The (simulated) conventional approaches yielded similar results for the maximum pitch motion. However, the OCTOPUS predictions appear to be significantly more accurate in predicting

the maximum roll motion in this evaluation case. Before, OCTOPUS had a general MAPE of more than 100%. In the case of time series 66, it achieved an accuracy of 62%.



(a)



(b)

Figure 55: Comparison of Prediction Approaches on MAPE in Predicting the Maximum Roll and Pitch Motions [0% of LC7 Included]

5.3 Adaptability of Machine Learning-Based Approaches

Figure 56 displays the MAE changes with the inclusion of LC7 data to train and validate NauticalNet and NauticalNet* for both roll and pitch motions. The x-axis of these graphs indicates the percentage of

LC7 data included, with 0% being the scenario from Subsection 5.2. It is observed that after including 15-20% of the data, the ML-based approaches begin to outperform (simulated) conventional approaches for both maximum roll and pitch motions. However, more iterations are needed for significant cases before ML-based approaches show better predictions. Specifically, for maximum pitch motion, about 30% of data inclusion (approximately 300 data points) is required for ML-based methods to surpass conventional ones.

A general trend noted is the reduction in MAE for ML-based approaches with increased data inclusion from time series. More data lead to more accurate estimations of how LCs affect the vessel's roll and pitch motions. Initially, a notable reduction in error is observed with a small addition of data, suggesting a significant improvement. As more data are included, this rate of error reduction slows, indicating that ML-based approaches have assimilated enough data to understand the relationship between LCs, environmental factors, and vessel response. Similar observations are shown in Appendix E.2, Figure 117.

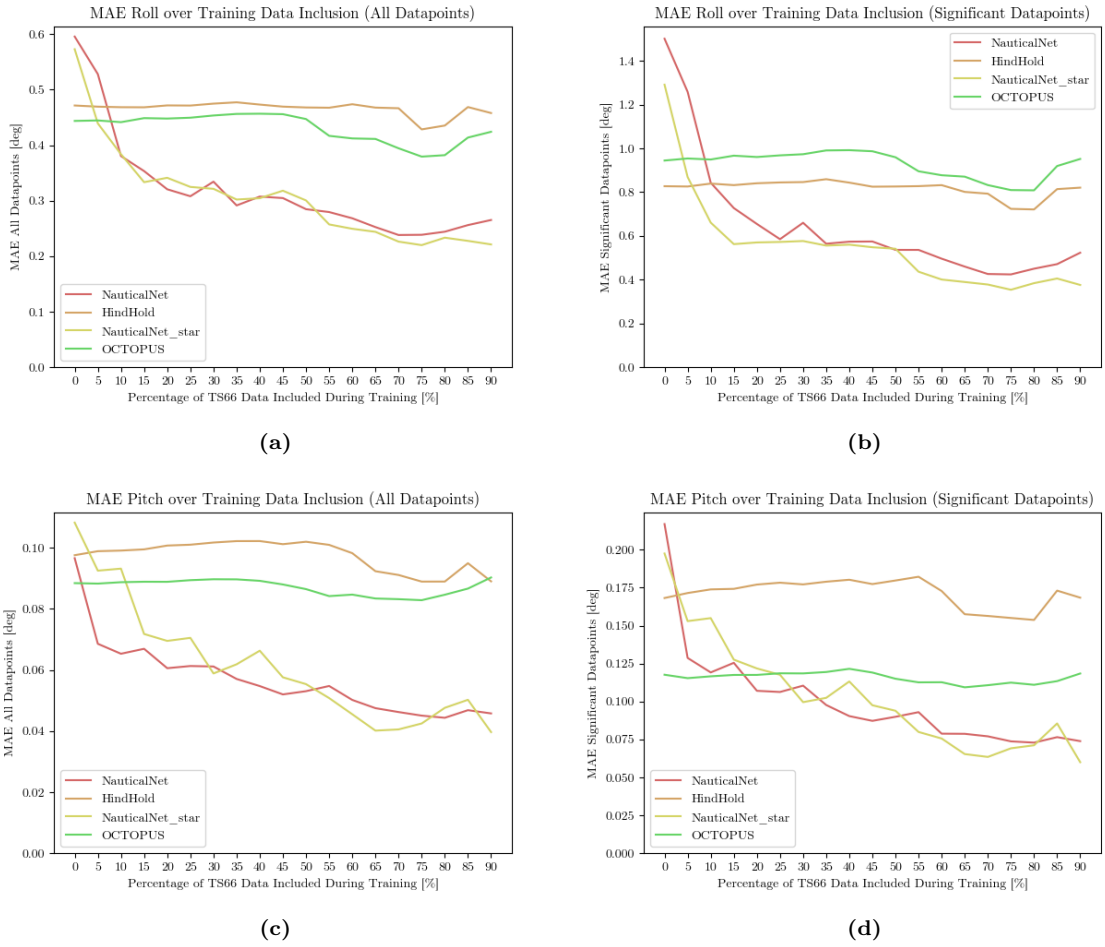
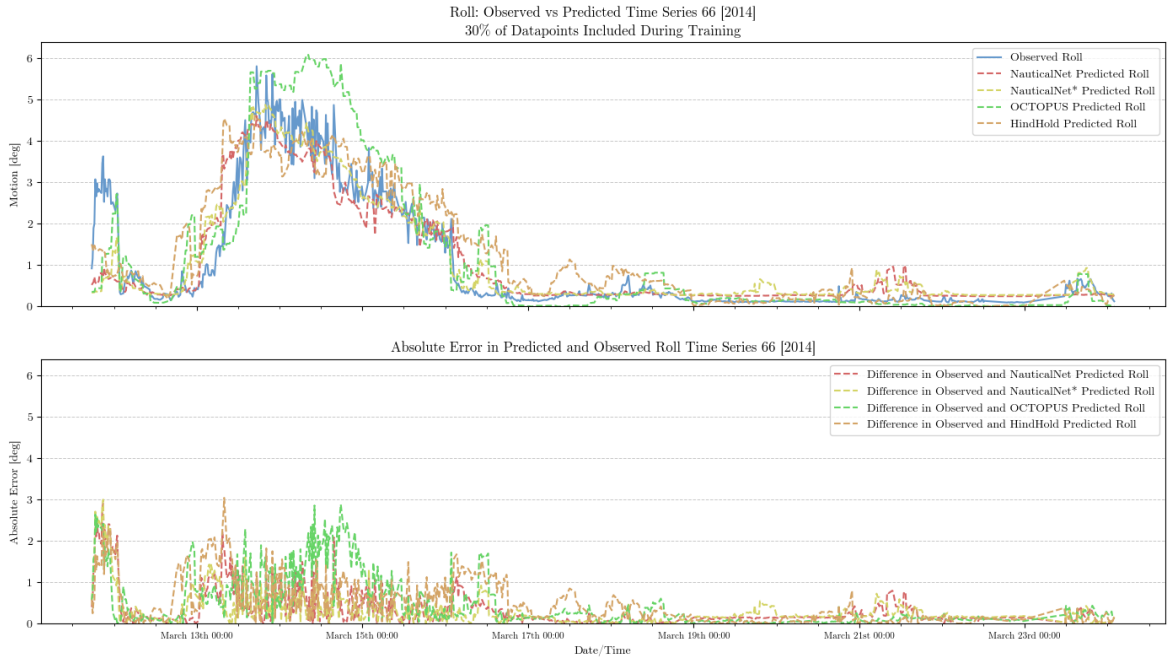
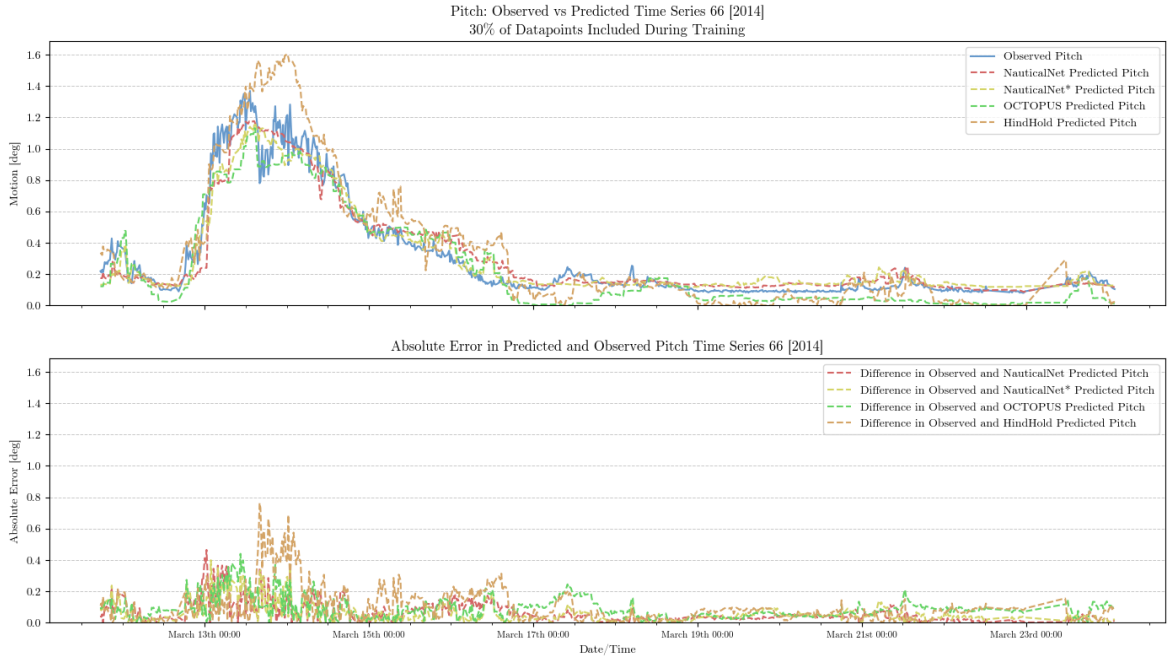


Figure 56: MAE over % of data included during training and validation of NauticalNet and NauticalNet*

The prediction plots corresponding to 30% of data inclusion, which should be the one where both ML approaches outperform (simulated) conventional approaches, is shown in Figure 57. It shows a fairly similar data distribution as before in the 0% inclusion case, indicating that it is still an appropriate test set and the random selection was proper. In both plots, the ML-based approaches have better tracked the higher values present in the data. For both maximum roll and maximum pitch, ML-based approaches are arguably more accurate than (simulated) conventional approaches, in general and in the significant domain, as confirmed in Figure 56. All prediction graphs related to the inclusion of data can be found in the Appendix E.3.



(a)



(b)

Figure 57: Comparison of Prediction Approaches on MAE in Predicting the Maximum Roll and Pitch Motions over Data inclusion of Time Series 66

6. Discussion

This chapter is designed to thoroughly evaluate the methodology employed in this research, delving into aspects such as data quality and feature importance, and seeking explanations for the observed results. These elements collectively lay the groundwork for the study's conclusions and recommendations.

The chapter begins with an analysis of the research methodology and its outcomes. It starts by assessing the quality of input data and features, as detailed in Section 6.1. The development process of the model, including the rationale behind its design and the validation strategies implemented, is discussed in Section 6.2. The chapter concludes with Section 6.3, which is dedicated to uncovering factors that could elucidate the findings of this research.

6.1 Analysis of Input Data: Quality and Feature Exploration

In the context of this study, a direct emphasis has been placed on investigating data quality. This section intends to contribute to this investigation, gradually progressing towards practical recommendations for Boskalis, which will be presented in the subsequent chapter. Initially, the focus will be on examining the quality of hindcast environmental data. Following this, a comparison between hindcast and nowcast environmental data sources will be conducted. Lastly, the section will explore the feature importance in both NauticalNet and NauticalNet*.

6.1.1 Quality Assessment of Hindcast Environmental Data

The accuracy of hindcast environmental data is challenging to determine due to the lack of real-time weather measurements at each location. Uncertainty estimates for each environmental feature in the datasets used by NauticalNet and HindHold are provided by the Copernicus Climate Change Service (C3S), determined using an ensemble approach. This approach involves ten different simulation runs, or 'ensemble members,' with a three-hour temporal resolution. Each member of the ensemble uses slightly different initial conditions or model configurations, resulting in a range of outcomes (Hersbach et al., 2020). The variability in these simulations is used to measure the uncertainty for each environmental feature.

The mean value of each environmental feature is normalized to 1.0 and the variability, or spread, is normalized against this mean, leading to a relative spread. This normalization helps to describe the variability and dependability of the data. Figure 58 shows the weighted relative spread for each

hindcast environmental feature in the data sets of NauticalNet and HindHold. This figure emphasizes the variability in each feature and visually illustrates the uncertainty in the hindcast data. They correspond to the entire data set, so training data, validation data, and testing data are combined. The '10-meter V wind component' has the highest spread percentage at 34%, indicating considerable variability and uncertainty. The '10-meter U wind component' and the 'wave spectral directional width for wind waves' also have high spreads of 24% and 23%, respectively, suggesting significant fluctuations in these parameters. On the other hand, features such as Model Bathymetry have a spread percentage of 0%, which implies consistent data. The mean wave direction and the mean period of total swell also have relatively low spread percentages of 0.96% and 0.98%, respectively, indicating more reliable and less variable data. Features such as 'Mean wave direction of third swell partition' and 'Mean wave period of third swell partition' have intermediate spread percentages of 9% and 17%, respectively, which are not as high as the wind components, but still indicate a moderate level of uncertainty.

These findings collectively provide a clear understanding of the different levels of confidence that can be placed on various aspects of the hindcast environmental data. Features with lower spread percentages are arguably more reliable, whereas those with higher percentages require a more cautious approach and potentially additional verification or corroboration. The importance of uncertainty of particular features is also dependent on how important the feature itself is for ML-based approaches. Some features will have a heavier influence on the maximum roll or pitch motion than others, and therefore the uncertainty in these features is arguably more important than in other features. Therefore, it is difficult to relate the uncertainty of the features directly with the outcome of this research without first investigating the importance of the features, as will be done in Section 6.1.3. It is evident that the features that are exclusive to the hindcast-based prediction approaches (discussed in Chapter 4) have a relatively larger spread, particularly the third-swell wave components. This implies that the uncertainty in the hindcast data is not particularly significant for the features that are shared between the four approaches, apart from the wind features.

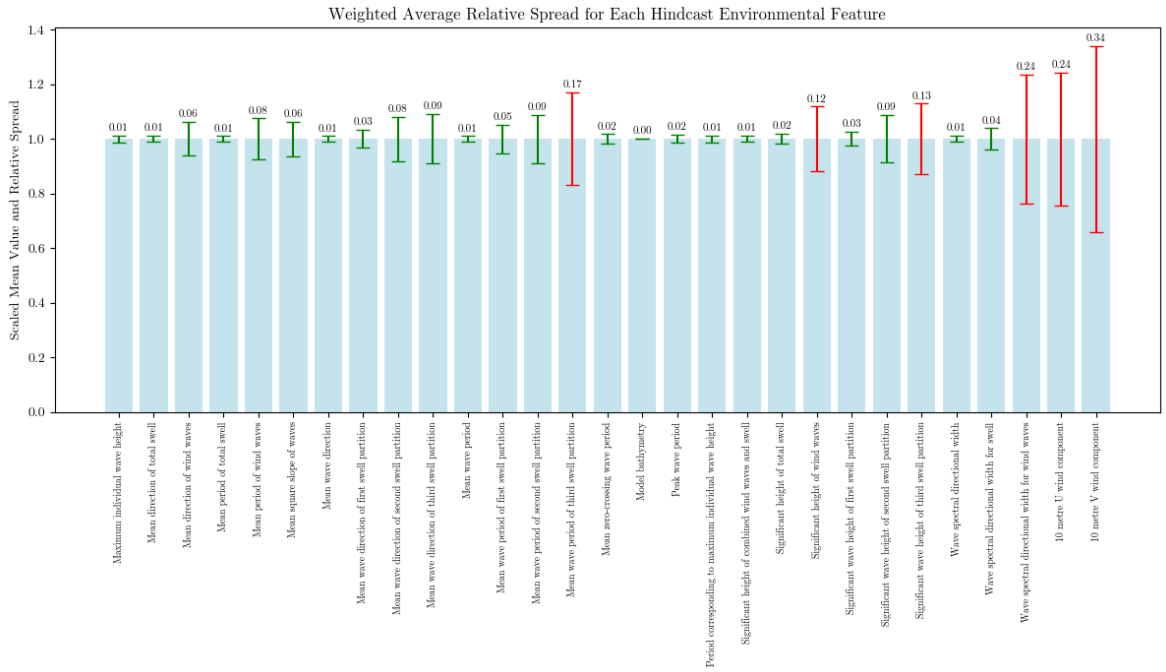


Figure 58: Weighted Relative Spread for Each Hindcast Environmental Feature

6.1.2 Hindcast versus Nowcast Environmental Data

In Chapter 5, Section 5.1, using the first validation approach, it was observed that hindcast data are generally more accurate than nowcast data. To further explore the accuracy of environmental data, an analysis of the shared features between hindcast and nowcast data is performed, providing information on the differences between the two data sources. Furthermore, this assessment will give an indication of the quality of Boskalis' nowcast environmental data in comparison to hindcast environmental data which, in theory, should be closer to reality.

Wave Height Features

Figure 59 shows a clear difference between the wave height features obtained from the hindcast and nowcast data sets. This is especially noticeable in the combined significant wave height of the wind waves and swells (Figure 59(a)) and the maximum individual wave height (Figure 59(b)), where considerable discrepancies can be seen. Although the distributions of the individual swell and wind components are more similar, particularly in terms of the range of the data, there are still notable differences in the weight of the distribution, particularly in the number of occurrences at a certain height.

The effects of these discrepancies on ML-based approaches are difficult to measure in terms of their effect on overall accuracy. In comparison, as discussed in Chapter 2, Section 2.1, OCTOPUS is based

on the wave spectrum in spectral analysis, which is partially dependent on the combined significant wave height. Given the observed variation in the combined significant wave height and assuming hindcast data are usually more in line with reality, it could be argued that by using nowcast data for the determinations of the wave spectrum used in the OCTOPUS approach, it potentially leads to decreased prediction accuracy.

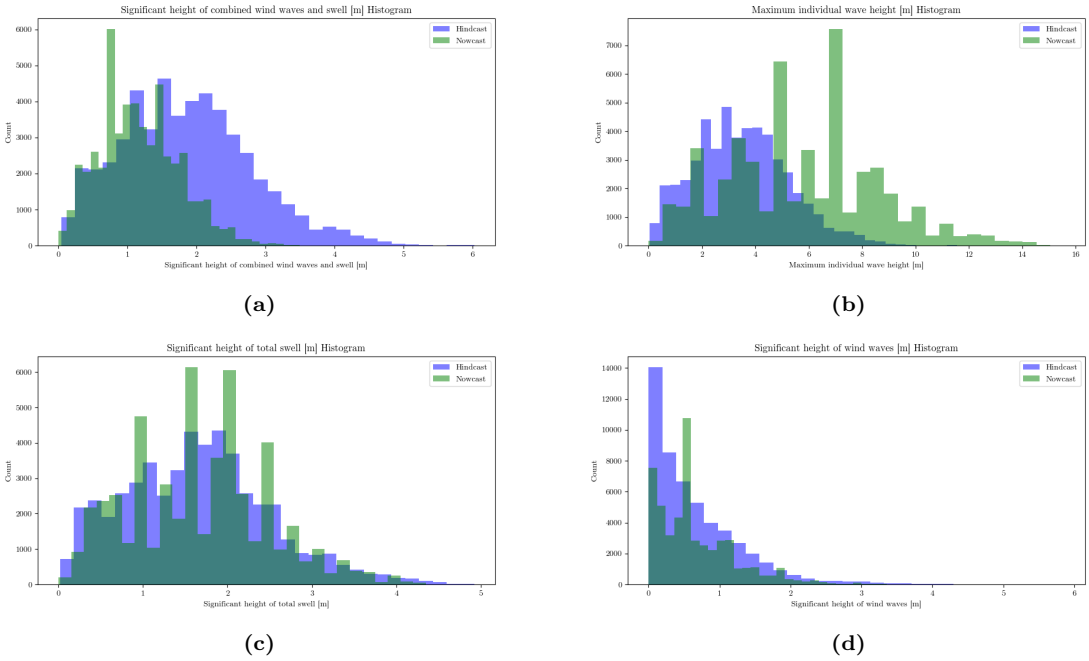


Figure 59: Histogram of Hindcast and Nowcast Data for Wave Height Features

Wave Period Features

Differences in wave period features are illustrated in Figure 60. It is observed that the distribution of the mean wave period, as depicted in Figure 60(a), exhibits a relative similarity between the hindcast and nowcast data sets. However, distinct differences are apparent in the other three shared wave period features (Figures 60(b), 60(c), and 60(d)). The nowcast wave period features are found to be rounded to the nearest whole number, in contrast to the hindcast features which display a broader range of values. This rounding in the nowcast data set is identified as a disadvantage, as it can lead to an inaccurate representation of the sea state surrounding the vessel. This is particularly critical when wave periods are close to the natural periods of the vessel, where these inaccuracies can result in problematic outcomes. The RAO is characterized by a pronounced dense peak around its natural period. Should the actual wave period fall between two whole numbers close to this natural period, the impact of the rounding becomes significant. A slight shift to either side of the peak will arguably result in considerably reduced energy transfer. Consequently, methods that use spectral analysis are more significantly affected by this data rounding, and OCTOPUS is identified as the most affected.

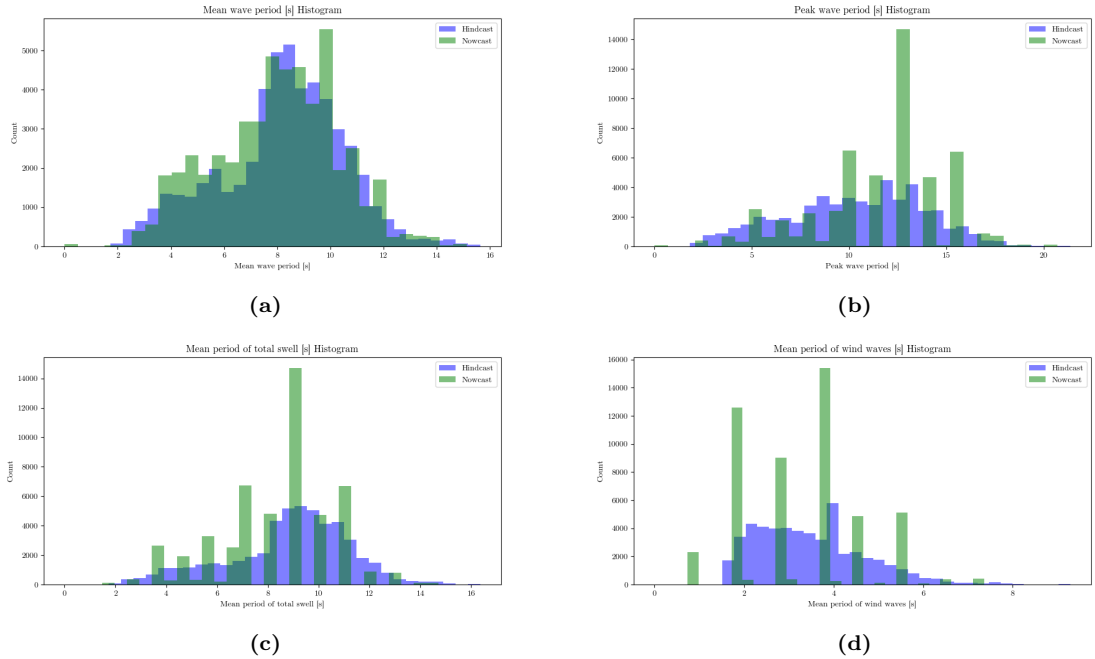


Figure 60: Histogram of Hindcast and Nowcast Data for Wave Period Features

Directional Features Analysis

The histograms in Figure 61 demonstrate similarity in the distributions of the directional features, with the exception of the relative direction of the wind, which has some variations. This implies that these features have a limited effect on the difference in prediction accuracy between NauticalNet and NauticalNet*, and between OCTOPUS and HindHold. Nevertheless, this should not be interpreted as a sign of its insignificance, as these features are likely still essential for all four approaches.

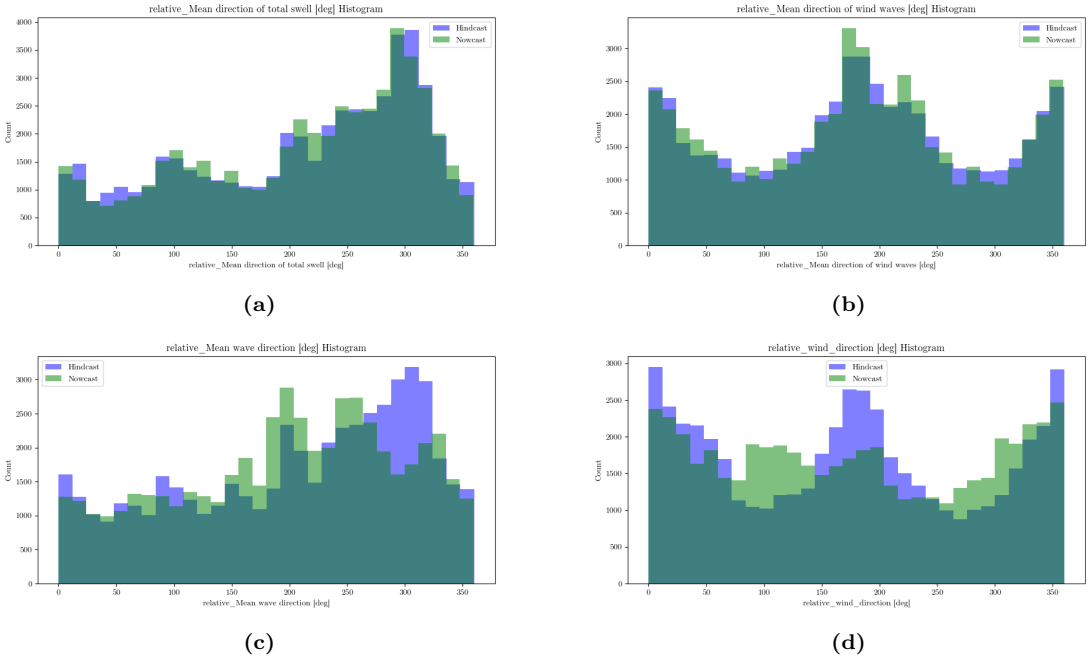


Figure 61: Histogram of Hindcast and Nowcast Data for Directional Features

Potential Influences on Prediction Accuracy

The analysis has revealed the marked distinctions between hindcast and nowcast environmental data, which have an effect on the accuracy of both ML-based methods such as NauticalNet* and traditional approaches such as HindHold. It is recommended to be careful when comparing these approaches, particularly since HindHold is an approach designed with ML that mimics the OCTOPUS approach. The OCTOPUS approach, which is dependent on precise wave-spectrum data, is more vulnerable to inaccuracies in nowcast data than its ML-based counterpart NauticalNet*. This is mainly due to the rounding of wave-period features in nowcast data, which can cause considerable prediction errors for conventional approaches as previously discussed, particularly when these periods are close to the vessel's natural frequencies. NauticalNet* has demonstrated an impressive capacity to adjust and learn from uncertainties in nowcast data, as demonstrated by its prediction accuracy. This adaptability highlights the power of ML-based approaches to deal with data variability and uncertainties.

Conventional approaches such as OCTOPUS thus probably rely more on high-quality data than the ML-based prediction approaches, as they are based on wave spectrum analysis, a critical component of their methodology. This contrast underlines the importance of data quality in conventional prediction approaches and highlights the trade-offs between different approaches. Moving forward, the next critical aspect to explore in ML-based methods is the feature selection process, a key factor in understanding how these models process data and how this influences their prediction accuracy.

6.1.3 Evaluating Feature Selection

The selection of features in ML approaches, as used in NauticalNet and its variant NauticalNet*, plays a crucial role in the results of the research. NauticalNet's initial dataset of 89 features was reduced to 47 to improve model performance. This reduction, guided by Kotsiantis et al. (2006), involved removing features considered irrelevant or redundant, to increase computational efficiency and minimize noise. The selection of features for both ML-based approaches was designed to balance accuracy and manageability. Differences in NauticalNet*, arising from the use of nowcast data and OCTOPUS integration, highlight the complex interplay between data source selection, feature mapping, and model outcomes. These nuances should be carefully considered in interpreting NauticalNet*'s predictions, as they directly impact the model's learning environment and predictive capabilities. The selection of features highlights a fundamental principle in ML research: every decision in data preparation has significant implications.

The following section delves into the specifics of feature selection in NauticalNet and NauticalNet*, examining the significance of chosen features through feature permutation importance analysis. This analysis assesses the impact of individual features on the models' performance and offers insights into their reliability and robustness in handling environmental data variability.

Feature Permutation Importance NauticalNet

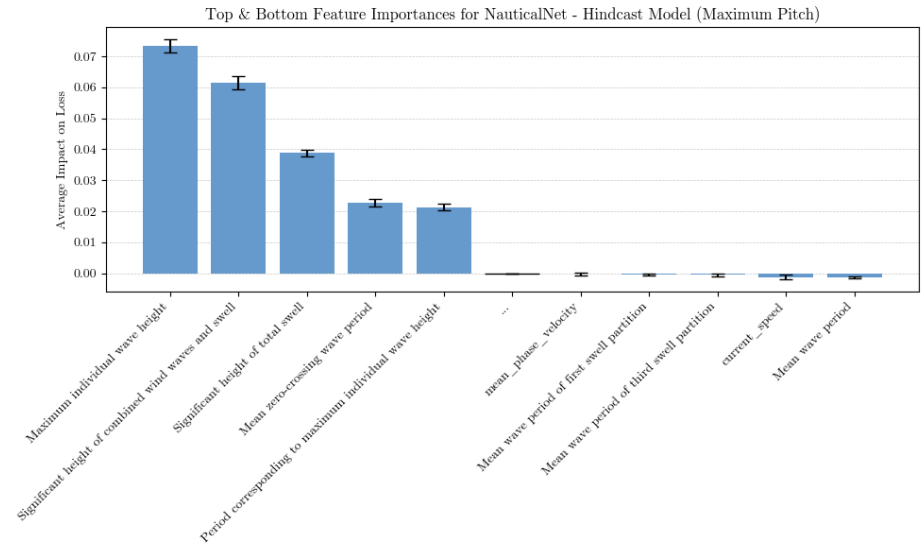
The feature permutation importance analysis for NauticalNet, depicted in Figure 62, highlights the importance of various features in predicting vessel motions. The analysis employs a custom loss function, the same as was used during training (see Chapter 4 'Execution'), to quantify the impact of randomly shuffling individual features while keeping others constant. The higher the score on the y-axis, the greater the impact on the loss function, indicating the importance of the feature. The standard deviation, represented by black minibars within the blue bars, is derived from ten iterations of shuffling to mitigate the influence of random variations.

For the prediction of the maximum pitch motion, NauticalNet depends primarily on environmental features that characterize the wave spectrum, such as significant wave height and wave periods (see Figure 62(a)). These features are crucial in defining the wave environment surrounding the vessel, as outlined in Chapter 2, Section 2.2. In contrast, features like mean wave period and current speed exhibit negative importance scores. The lesser importance of the mean wave period could be attributed to its redundancy, as the information it provides might already be captured by other wave period features. Meanwhile, the minimal impact of current speed aligns with theoretical expectations; although current can affect viscosity, its influence is relatively insignificant for pitch motion, which is predominantly governed by wave-induced damping.

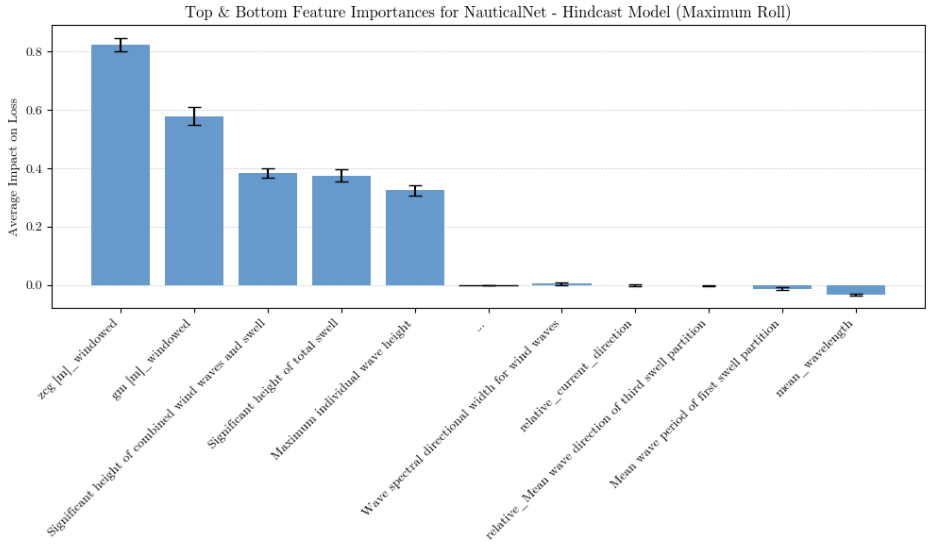
On the contrary, the roll motion analysis (Figure 62(b)) demonstrates the dominance of the stability

characteristics of the internal vessel, such as the Center of Gravity (CoG) in the z direction and the metacentric height (GM). These findings support the hypothesis that a vessel's roll motion is profoundly influenced by its internal dynamics and geometric configuration. The substantial reliance of the approach on these stability parameters implies a potential vulnerability. When faced with new LCs that alter the CoG and GM, the accuracy of the model prediction may be significantly compromised. This aspect potentially contributed to the lower prediction accuracy observed in the second validation approach (see Chapter 5, Section 5.2), where the approach encountered unknown LCs. Interestingly, features such as the mean wave length exhibit minimal importance in the prediction of maximum roll motion. On the contrary, for maximum pitch motion, the wave length might play a more significant role (see Appendix F). This could be due to the vessel's ability to 'ride' along the wave when the wave length is sufficiently long, allowing the vessel to align more with the wave's longitudinal profile. This alignment is less relevant for roll motion, which is primarily influenced by lateral stability factors.

In summary, NauticalNet's FPI analysis offers valuable information on its predictive capabilities for roll and pitch motions and highlights potential vulnerabilities under unfamiliar operational conditions. These insights are aligned with the findings in Chapter 5, Section 5.2.



(a)



(b)

Figure 62: Feature Permutation Importance Results NauticalNet for both Maximum Pitch and Roll

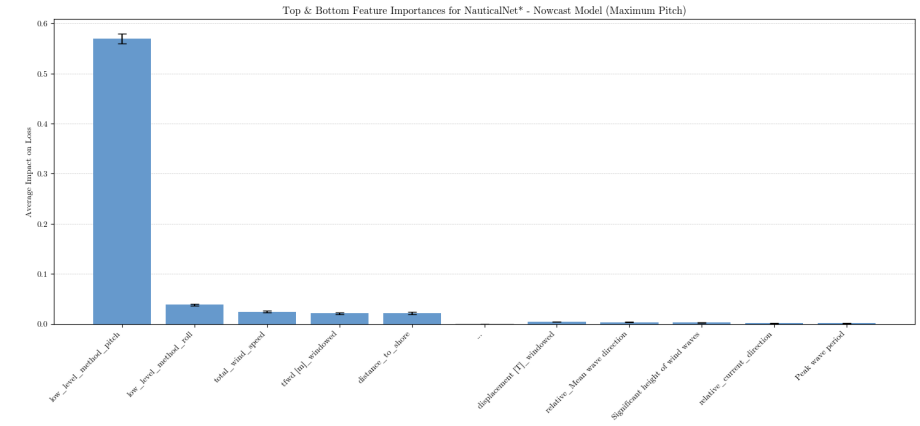
Feature Permutation Importance for NauticalNet*

NauticalNet* uniquely incorporates OCTOPUS predictions (denoted as ‘low_level_method_pitch‘ and ‘low_level_method_roll‘) as features, demonstrating a significant influence of conventional prediction approaches. This integration is reflected in the impacts on the loss function (0.57 for pitch and 2.16 for roll, see Figure 63), suggesting a strategic blend of conventional knowledge and modern ML techniques. Despite this dependence, NauticalNet* outperforms OCTOPUS substantially, indicating that it may have learned to discern the accuracy of OCTOPUS predictions in various scenarios.

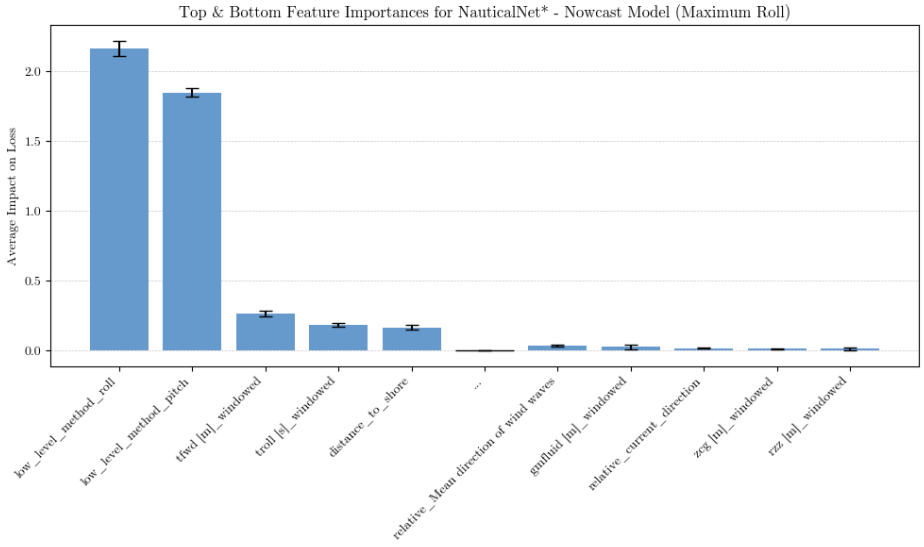
The similarity of NauticalNet* to NauticalNet, rather than to OCTOPUS, despite its reliance on OCTOPUS, is an intriguing aspect of this analysis.

The analysis indicating NauticalNet*'s reliance on the OCTOPUS maximum pitch prediction for predicting maximum roll motion, while demonstrating less dependence on OCTOPUS predictions for pitch motion, suggests a nuanced interplay between pitch and roll dynamics. This observed pattern may imply a potential correlation between pitch and roll motions, particularly in the way in which pitch dynamics could influence roll behavior. However, caution must be exercised in drawing definitive conclusions, as this correlation, while suggested, is not explicitly established. The significant influence of the OCTOPUS prediction on pitch in determining roll motion, as opposed to the relatively lesser impact of the OCTOPUS roll prediction on pitch motion, warrants further investigation to fully understand the underlying dynamics.

Additionally, the finding that the Center of Gravity (CoG) in the z-direction and the metacentric height corrected for the fluid hold less importance in NauticalNet* compared to NauticalNet is notable. It could be presumed that the OCTOPUS predictions integrated into NauticalNet* may already encapsulate aspects of the vessel's LCs. Consequently, this integration might reduce the direct impact of specific LC features, such as the CoG and metacentric height, on the model's predictions. This hypothesis suggests that in NauticalNet*, the LC features are indirectly represented through the OCTOPUS predictions, possibly in a combined or transformed form. This indirect representation could account for the observed diminished importance of these specific features in the model's decision-making process.



(a)



(b)

Figure 63: Feature Permutation Importance Results NauticalNet* for both Maximum Pitch and Roll

6.2 Model Development and Validation Approaches Evaluation

This section revisits parts of the research approaches outlined in Chapter 3 'Methodology' and Chapter 4 'Execution,' which together detailed the process of model development and validation to explore the potential of ML-based approaches to predict vessel motion. Evaluating these approaches is crucial to understand their effectiveness and identify areas for improvement. This section addresses the model choices made during development before discussing the advantages and disadvantages of each validation strategy, providing a thorough discussion of the methodology applied.

6.2.1 Model Development

The development of NauticalNet and NauticalNet* involved important decisions that significantly influenced the results of the research. Both approaches employ a multitask deep learning neural network architecture, but this architecture was optimized for NauticalNet, not NauticalNet*. NauticalNet's architecture was designed to incorporate a large number of input features, necessitating a complex structure with shared and task-specific hidden layers. This complexity was justified by the diversity of integrated hindcast environmental data into the model. However, the same architecture may not be suitable for NauticalNet*, which utilizes a reduced feature set of 32 inputs, primarily due to its reliance on nowcast data. This disparity in input features could have implications for the effectiveness of the shared architecture, and a more streamlined architecture might have been more suitable for NauticalNet*, potentially enhancing its learning efficiency and predictive accuracy. The hyperparameter settings for both models, which involved decisions on the number of epochs, batch size, learning rate, and other factors, were critical in shaping their training processes. These settings were determined through a combination of manual optimization and grid searches, reflecting a balance between computational feasibility and model performance. The differences in training duration between NauticalNet and NauticalNet* (approximately 310 versus 440 epochs, respectively) suggest varying degrees of complexity and challenges in model convergence. Particularly for NauticalNet*, the longer training and observed differences between training and validation losses might indicate a tendency toward overfitting, caused by the complex architecture with respect to the reduced feature set. Incorporating a custom loss function into both approaches, designed to penalize errors more heavily for higher target values, was a strategic choice. This function was intended to address inaccuracies in high-value predictions, a crucial aspect given the operational significance of vessel motion predictions. However, this emphasis on high-value errors could mean that the approaches are less sensitive to smaller, but still operationally relevant, variations in vessel motions.

HindHold is a method designed to bridge the gap between conventional and ML-based approaches by trying to replicate what OCTOPUS would have predicted using hindcast data. This concept raises questions about the reliability of its predictive capabilities. The development process, which included manual optimizations and grid searches for hyperparameters, was essential in determining HindHold's predictive performance. HindHold's mirroring capabilities, as demonstrated by the mean absolute errors in its predictions compared to OCTOPUS, show its effectiveness in replicating OCTOPUS outcomes with nowcast data. However, this mirroring raises questions about how much HindHold truly reflects what OCTOPUS would predict if given hindcast data. The mean absolute errors of 0.02 [deg] for pitch and 0.05 [deg] for roll suggest a high degree of accuracy, yet it is uncertain if OCTOPUS would return these values under hindcast conditions. The reliance on a simulated approach, while innovative, carries the limitations and assumptions.

6.2.2 Limitations of the Custom Cross Validation Strategy

The custom training validation strategy, as outlined in Chapter 4, Section 4.2.1, ensured the equal representation of all 191 time series in the training and validation datasets. This approach was designed to expose the ML-based approaches to a wide range of different LCs, thus enhancing their learning potential. However, a significant limitation arises from this methodology. The ML-based approaches become inherently biased towards the 191 time series and, by extension, the LCs these series represent. This bias manifests itself during the validation phase, where the approaches are assessed against conditions on which they have already been trained. Consequently, this could lead to an underestimation of validation losses, potentially giving a skewed perspective of the models' performance in handling new, unseen LCs. This limitation is a common factor in both validation strategies, as they share the same approach to divide the training and validation data. The following subsections explore the strengths and weaknesses of each strategy, considering their individual characteristics.

6.2.3 Strengths and Weaknesses of Validation Strategy I: Training and Validating on All Time Series

The first validation strategy in this study, which involved training and validation of ML-based approaches using all 191 time series, was important in evaluating the approaches under known LCs.

Strengths: This strategy offers a comprehensive overview of the performance of ML-based approaches under previously encountered LCs. By incorporating all available time series for training and validation, it ensures an exhaustive exploration of the capabilities of these approaches in familiar scenarios. This is vital for applications where similar conditions are frequently encountered or for post-processing purposes. The evaluation of all 191 time series in the testing phase provides a diverse range of data, crucial for a robust assessment of the performance across various situations.

An additional strength of this approach is its facilitation of a broad evaluation of hindcast versus nowcast environmental data. This aspect is critical as it is less dependent on whether specific LCs are included but is pivotal in assessing the approaches on a wide range of data. This strategy played a key role in concluding that nowcast data is generally of lesser quality than hindcast data, a finding relevant for both ML-based and (simulated) conventional approaches. The broad dataset used in this strategy enhances the understanding of how different environmental data types affect prediction accuracy.

Moreover, this validation strategy enables extensive learning for ML-based approaches, exposing them to a spectrum of data. This exposure is imperative for these approaches to discern complex patterns and relationships within the data, potentially elevating their accuracy and reliability in familiar conditions.

Weaknesses: This strategy has some notable weaknesses. Primarily, the inclusion of all time series in the training phase restricts the ML-based approaches' ability to generalize to unseen LCs. This is a major issue, as it could reduce the predictive robustness of ML-based approaches in real-world scenarios. Furthermore, this approach may lead to an overly optimistic assessment of the accuracy of ML-based approaches. Since the ML-based approaches are validated on conditions they have already been trained on, there is an inherent bias towards these familiar conditions. This bias can hide the true predictive performance of the approaches, as their ability to handle new unencountered scenarios remains untested. As a result, while the ML-based approaches may appear highly accurate within the validation framework, their performance in real-world applications, where novel conditions are likely to be encountered, may not be as reliable.

This analysis of validation strategy I has revealed the trade-offs associated with it. It is excellent in providing thorough understanding of the approach's performance in known LCs and offers a wide range of testing scenarios. However, it has drawbacks in terms of its ability to generalize and the possibility of overestimating the approach's accuracy.

6.2.4 Strengths and Weaknesses of Validation Strategy II: Excluding a Specific Loading Condition

The purpose of the second validation strategy was to assess the generalizability of the ML-based approaches by omitting a particular LC from the training and validation datasets.

Strengths: This strategy has a major advantage in that it evaluates the ML-based approaches' capacity to generalize to new conditions, which is essential for practical applications. By omitting a certain LC, this strategy provides a unique chance to assess how well the ML-based approaches can adjust to situations they have not experienced during training. This is especially relevant in real-world settings where conditions can vary significantly and unpredictably. The knowledge gained from this approach is invaluable in determining the practical applicability and robustness of the ML-based approaches in different environments. Furthermore, this strategy establishes a strong basis for future research. By emphasizing the performance of the ML-based approaches (or lack thereof) in unexplored conditions, it gives a clear direction for further development and refinement of the approaches. This aspect is essential to progress the field, as it pinpoints particular areas where more research and improvement are needed.

Weaknesses: This strategy's reliance on a single excluded LC may lead to a limited and potentially non-representative test set, which could affect the conclusiveness of the results. Additionally, it may lack the depth and rigor of a primary strategy. For instance, the custom cross-validation strategy, discussed in the previous subsection, was designed for the primary validation strategy, which may not be suitable for the objectives of this secondary validation strategy. A random approach to training and

validation might have been beneficial, giving the ML-based approaches the opportunity to encounter unseen LCs and potentially leading to a more generic approach. These factors should be taken into consideration when interpreting the results, as they may not fully capture the complexity of the problem. Moreover, by excluding certain conditions, there is a risk that the approaches become less versatile compared to those developed under the first strategy, potentially impacting their effectiveness in a broader range of real-world scenarios.

In conclusion, this second validation approach provides essential information on the applicability of ML-based approaches. However, its restrictions must be taken into account. The possibility of the test set not being representative and the particular design decisions made in the strategy emphasize the need for a thorough comprehension of its results and their implications for model creation.

6.2.5 Reflection on the Influence of Focused Methodological Choices

Reflecting on the methodology employed in this study, it becomes evident that a significant portion of the research effort was dedicated to optimizing ML-based approaches. This focus, while central to advancing the field of ML in maritime motion prediction, came with its own set of trade-offs and influenced the study's outcomes in specific ways.

Impact of Integrating OCTOPUS Predictions into NauticalNet*: Incorporating OCTOPUS predictions as an input feature in NauticalNet* adds a layer of complexity to the comparative analysis between (simulated) conventional and ML-based approaches. This means that NauticalNet* is indirectly taking advantage of the predictive capabilities of OCTOPUS in its own predictions. This setup may create bias in the study results. The primary goal of contrasting the effectiveness of (simulated) conventional versus ML-based approaches becomes somewhat muddled, as NauticalNet* no longer operates independently of the (simulated) conventional approach to which it is compared. Although this approach is beneficial for exploring the synergies between different prediction approaches, it may not be in line with the study's original aim of a direct comparison.

Validation Strategy and Its Limitations: The chosen validation strategies, although robust in their design, mainly tested ML-based approaches within a range of known LCs (validation strategy I) or a limited scope of new LCs (Validation Strategy II). This approach limits the exploration of the ability of ML-based approaches to generalize in a wider range of real-world scenarios. A custom cross-validation strategy, tailored to the first validation strategy, but influencing both validation strategies, further reinforces this focus. It is evident that the methodological choices facilitated a detailed exploration of ML in a specific maritime context, but also shaped the study results in certain directions. To fully comprehend the scope and implications of the findings, it is essential to recognize these influences. Future research efforts in this area should include a broader approach, encompassing a wider variety of

vessels and operational conditions, and a more balanced focus between (simulated) conventional and ML-based approaches. Recommendations for this will be discussed in the next chapter.

6.3 Exploring Factors Influencing Research Outcomes

To provide valuable recommendations and draw insightful conclusions, it is crucial to examine the factors influencing the outcomes of this research. The exploration of these aspects not only enriches the current study but also serves as a potential guide for future research. Key findings have been outlined in Chapter 5 'Results'. This section aims to offer potential explanations for these findings, thereby establishing a solid foundation for the discussions in the subsequent chapter.

6.3.1 Environmental Data

The findings regarding the differences between environmental data types indicated that hindcast data is of better quality than the nowcast data (see section 6.1.2). This is expected, as hindcast data is essentially nowcast data corrected with actual observations and weather models. Having investigated the quality of the nowcast data provider in the thesis in more detail revealed significant discrepancies for significant wave height and wave period features compared to hindcast data. Specifically, for wave period features, the nowcast data provider rounds values to whole integers, likely impacting the predictions of the current prediction approach OCTOPUS. Interestingly, NauticalNet*, using the same nowcast environmental database, appeared better equipped to account for data inaccuracies. The ML-based approach learned not to rely too much on these features, and maybe even considered the error in the data. This underscores an advantage of ML over (simulated) conventional approaches: the ability of ML-based approach to adjust to data imperfections more effectively.

6.3.2 ML-based Approaches vs (Simulated) Conventional Approaches under Known Loading Conditions

The core finding of this research highlights the superior performance of ML-based approaches under known LCs. Although exceptions exist, the overarching trend favors ML-based approaches. The limitations of OCTOPUS, particularly its reliance on linear theory, shed light on this outcome. As discussed in the introduction, OCTOPUS faces problems in incorporating nonlinear phenomena, such as vertical vessel motions and nonlinear damping in roll motion. It also does not inherently include forward vessel speed. Although Ikeda et al. (1978) offered methods to address some of these issues, the linearization process inherently leads to a loss in predictive accuracy. This is evident in OCTOPUS's MAPE on a broad range of validation data, exceeding 100%, signaling a clear need for advancements in vessel motion prediction systems.

Deep neural networks, used in this research, excel in identifying and learning nonlinear relationships

within the data. This capability can be important in offshore environments, where each observed data point likely embodies inherent nonlinearities. ML-based approaches adopt a retrospective learning method, effectively uncovering hidden nonlinearities in historical data. This contrasts with conventional approaches like OCTOPUS, which employ a forward-looking methodology and are constrained by the impracticality of integrating nonlinear dynamics computationally and mathematically. This fundamental difference potentially explains the improved performance of ML-based approaches.

The benefits of ML are particularly noticeable in predicting roll motion compared to pitch. Roll motion, with its complex nonlinear damping, is more susceptible to dynamic amplification effects and is not well represented by linear models. For instance, vessel speed influences the extent of roll damping, with higher speeds inducing more skin friction on the hull and lift effect damping, which are known to be nonlinear. Figure 64 illustrates that ML-based approaches are better able to capture these complexities by correlating the observed roll motion with the speed of the vessel. The color of the data points shifts from green to red, where red represents larger prediction errors. It is important to note that many data points overlap in these plots. The purpose of these plots is to show the distribution of errors in both approaches. Data points with higher errors are superimposed over those with lower errors. For higher vessel speeds, OCTOPUS shows increased errors, indicating it finds it hard to include these induced nonlinearities. Pitch motion, in contrast, is predominantly linear and heavily influenced by radiation damping, which is effectively captured by linear theory. Therefore, the advantage of ML in roll motion prediction is probably more pronounced, as it effectively addresses the nonlinear aspects. This theory is further supported by the correlation matrices shown in Chapter 5 'Results', Section 5.1.1. The higher errors seen in the left plot are related to a different issue, regarding the rare extreme motion response values.

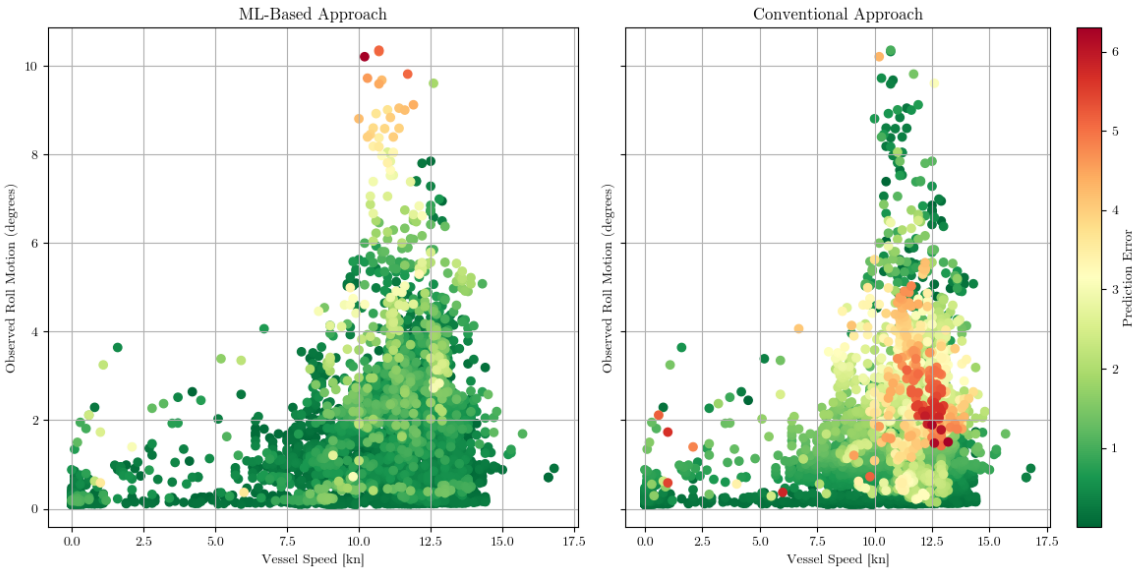
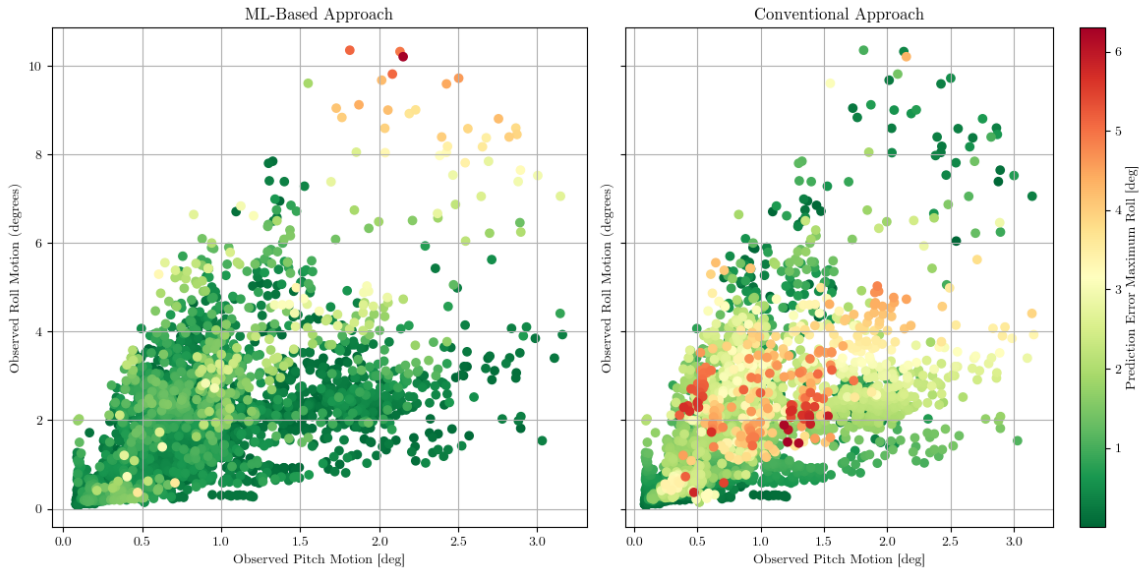


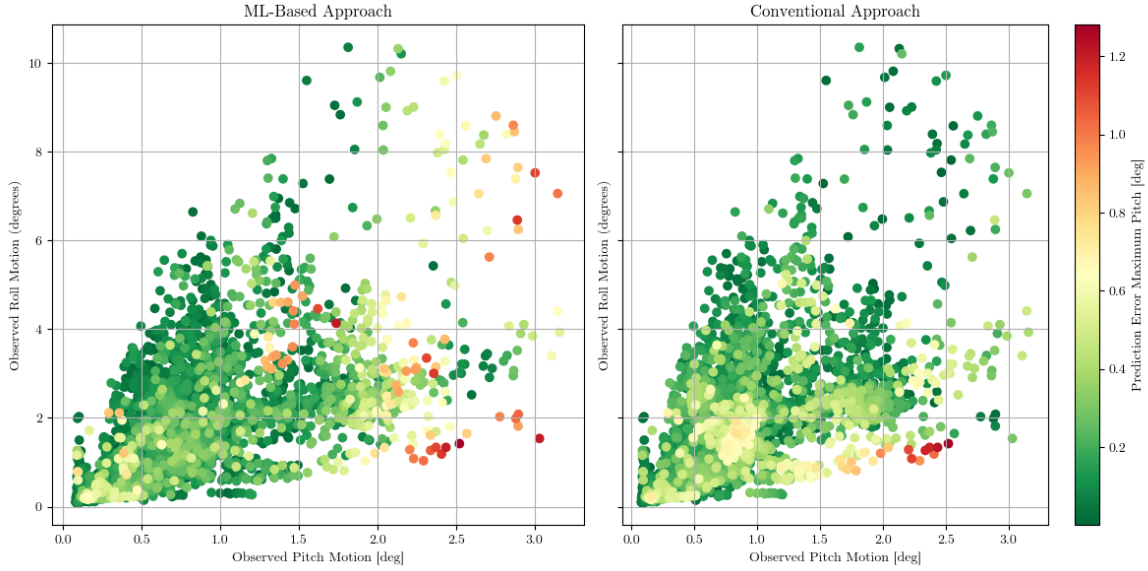
Figure 64: Comparative spread graph depicting the prediction differences between ML-based (NauticalNet*) and conventional approach (OCTPUS), suggesting the potential of ML-based approaches in handling nonlinear phenomena

6.3.3 ML-based Approaches vs (Simulated) Conventional Approaches in Extreme Motion Response Conditions

Extreme motion response conditions are characterized by rare high-magnitude responses of the HTV Target. Figure 65 includes scatter plots that compare the accuracy of ML-based approaches with conventional prediction approaches, focusing on errors in roll (Figure 65(a)) and pitch motion predictions (Figure (65(b)). Each plot displays observed roll motions against observed pitch motions. It is observed that ML-based approaches demonstrate increased errors for the most extreme roll and pitch motions. In areas where extreme motions are present, particularly at the higher end of the scale, a concentration of red data points is visible, indicating significant deviations from expected predictions. These findings correspond to the results found in Chapter 5 'Results', Section 5.1.1, for time series 104. The reason for this trend is found in the training data used for ML-based approaches. Extreme motion events are rare occurrences and, as such, are underrepresented in training datasets. Consequently, the ML-based approaches are skewed toward accurately predicting more frequent, moderate scenarios. This skew results in a conservative bias, especially evident when the approaches encounter rare, high-magnitude events, leading to underestimation. This behavior is indicative of a limitation inherent to ML-based approaches: they excel at generalizing across common scenarios, but falter when predicting outliers, which could be vital for operational safety and efficiency.



(a)



(b)

Figure 65: Comparative spread graph depicting the prediction differences between ML-based (NauticalNet*) and a conventional approach (OCTOPUS), showing the inability of ML-based approaches predicting extreme responses

6.3.4 ML-based Approaches vs (Simulated) Conventional Approaches under Unknown Loading Conditions

The limited capacity of ML-based approaches to generalize between LCs is evident in their decreased performance when predicting maximum roll and pitch motions under unknown LCs. This can be

attributed to the models' limited exposure to a variety of LCs. The dataset comprised only 24 different LCs, which are not evenly distributed. This limitation is analogous to training the model with 23 scenarios to predict an outcome for the 24th, a sample size that is generally inadequate in ML.

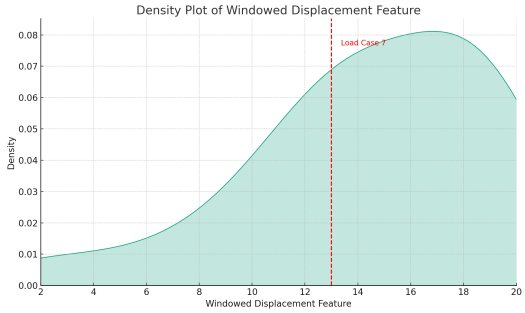
In the utilized dataset, there are 13 distinct features that describe the LCs on the vessel. Evaluating how common the excluded LC is within the dataset assists in contextualizing the results of this research. Figure 66 aids in this assessment. For select LC features, density plots are presented, with annotations for the values corresponding to the excluded LC. For instance, Subfigure 66(a) shows the distribution of the discretized displacement feature, highlighting the project's condition with a dashed red line. These plots reveal that while certain features, such as displacement and center of gravity in the x direction, are relatively common in the dataset, others, like corrected metacentric height and center of gravity in the z direction, are less so. These less common features are crucial for predicting maximum roll motion and may have contributed to the decreased performance of ML-based approaches, as they had relatively less data to learn from for these specific LCs.

The commonly used '10 times rule' suggests needing at least ten times more data points than features to adequately learn the relationship between features and target values. According to Alwosheel et al. (2018), the ideal sample size should even be at least 50 times the number of features, indicating a requirement of 130, using the commonly used '10 times rule', to 650, using Alwosheel et al. (2018) findings, distinct LCs for a generalized ML-based approach. By applying dimensionality reduction to the LC features, it is possible to reduce the 13 initial LC features. However, this process may lead to a loss of variance in the data, potentially impacting the model's ability to accurately generalize across different LCs.

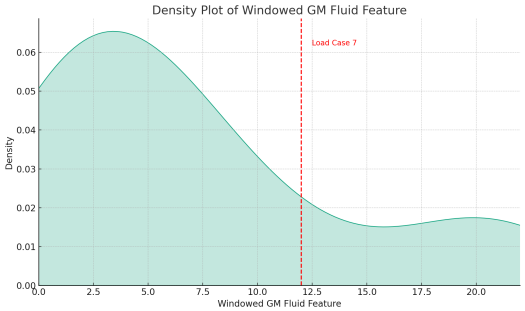
Expanding the dataset with new LCs is expected to narrow the performance gap between known and unknown scenarios. Given the current superior performance of ML-based approaches over OCTOPUS under known LCs, it is plausible to anticipate their effectiveness under unknown LCs, even before reaching the ideal sample size of 130 different LCs. In essence, ML-based approaches may outperform (simulated) conventional approaches in these scenarios before achieving the high accuracy levels seen in known LCs.

The decline in predictive performance for maximum roll motion in unknown scenarios is particularly significant. An importance analysis conducted in this chapter revealed the crucial influence of specific LC features, such as metacentric height and free surface moment, on roll motion. These features are vital in determining the stability and natural roll period of the vessel to environmental forces, making them particularly critical for accurate prediction of the maximum roll motion. A model's failure to precisely learn the intricate relationship between these LC features and roll motion response leads to greater inaccuracies, especially in scenarios not previously encountered. Therefore, reliance on these characteristics for roll motion, rather than pitch, may be the reason why the predictions of roll motions

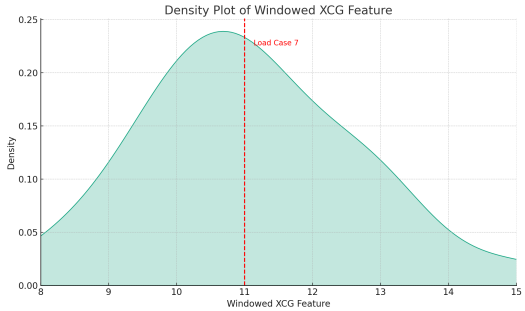
are more affected by the failure to learn the connection of the unfamiliar LC. A dataset that includes a wide range of LCs would allow ML approaches to better understand the intricate interactions between these conditions and roll motion, thereby improving the accuracy of predictions in both familiar and unfamiliar LCs.



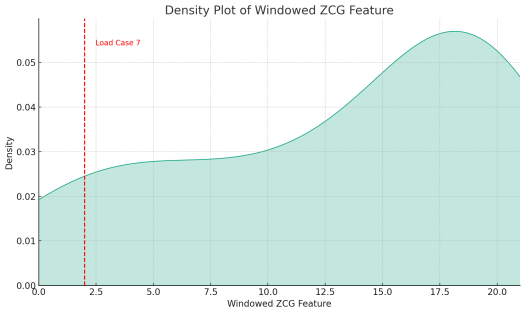
(a) Density Distribution of Windowed Displacement Feature



(b) Density Distribution of Windowed Corrected Metacentric Height Feature



(c) Density Distribution of Windowed xcg Feature



(d) Density Distribution of Windowed zcg Feature

Figure 66: Loading Condition Data Densities with LC7 Annotations

6.3.5 Adaptability of ML-based Prediction Approaches

ML-based approaches approaches, such as NauticalNet and NauticalNet *, demonstrate remarkable adaptability to unknown LCs. This adaptability is primarily due to their exposure to a vast amount of environmental data during training, enabling them to recognize patterns linking environmental factors with the vessel's behavior. With this foundational knowledge, the approaches can extrapolate to new LCs more effectively.

In the context of new LCs, these ML-based approaches take advantage of their existing knowledge base. They do not start from scratch; instead, they apply their understanding of how environmental factors influence vessel behavior, facilitating rapid adaptation to new data. This phenomenon is related to transfer learning, a concept in which a model developed for one task is adapted for a related one with minimal additional data, as explored by Tan et al. (2018) in their survey on deep transfer learning.

This could indicate its effectiveness in contexts such as predicting vessel motions under varying LCs. This ability to adapt quickly is a strong point of ML-based approaches, but still raises questions on how much data is needed for every new LC experienced by the HTV Target. This will be addressed in the recommendations section of the next chapter.

Amount of Data Theoretically Required for Adaptation

Determining the sufficient amount of new data for ML-based approaches to adapt to new LCs involves several factors. The inclusion of more diverse LCs in the training data would probably reduce the amount of new data required for adaption to a new LC. The similarity between new and previously trained LCs also plays a crucial role. ML models will probably adapt more quickly to LCs that are similar to those they have already encountered. For instance, with the current omitted LC, a significant improvement in both roll and pitch was observed after incorporating only 10% of the new data, which is equivalent to about 100 data points out of a total training set of 49,000 (approximately 0.2% additional data). However, this finding, based on a single condition, may not be universally applicable to other LCs, where the required data amount could vary. The amount of new data needed for transfer learning to provide reliable estimates for maximum roll and pitch motions depends on several factors:

1. **Similarity to Known Conditions:** The closer the new LCs are to those previously encountered, the fewer data points may be needed. The model can effectively utilize its learned patterns to these similar conditions.

2. **Complexity of New Conditions:** More complex or different LCs might necessitate more data for the model to understand and adapt.

3. **Model Learning Capacity:** The inherent learning capacity of the ML-based approach, including its architecture and learning algorithms, influences the amount of data needed. Advanced models might generalize from fewer examples compared to simpler models.

4. **Few-Shot Learning Techniques:** Incorporating few-shot learning techniques, as explored by Song et al. (2022) in their comprehensive survey, can reduce the data requirement. These techniques enable models to make accurate predictions with limited training examples, enhancing the adaptability of the approach with minimal data (Song et al., 2022).

5. **Data Quality:** High-quality representative data can significantly enhance the model's learning efficiency.

In practical terms, the required amount of new data for reliable estimates could range from a small fraction of the total dataset to a larger subset, depending on the above factors. The key is to provide enough data to cover the variability and complexity of the new conditions without overwhelming the approach. Although there is no definitive answer to the amount of new data required, a combination

of transfer learning, few-shot learning techniques, and careful consideration of the new conditions' complexity and similarity to known scenarios can significantly enhance the model's adaptability. This approach can effectively reduce data requirements to obtain accurate predictions under new LCs.

7. Conclusion and Recommendations

This thesis conducted a detailed investigation to compare ML-based approaches with (simulated) conventional prediction approaches, particularly in terms of their ability to predict maximum roll and pitch motions of the HTV Target. In doing so, it aims to address the limitations currently faced by OCTOPUS, further helping the industry and academia on the journey towards more accurate prediction approaches. The findings of this study not only demonstrate the potential and restrictions of ML-based approaches, but also open up possibilities for future developments in maritime operations. This chapter consolidates these findings, drawing conclusions from the extensive research conducted, and addresses recommendations for further advancing the field of maritime motion prediction. In Section 7.1, the main research question is addressed and an overarching conclusion is presented. Section 7.2 focuses on offering theoretical recommendations to overcome the challenges identified in the developed ML approaches. Section 7.3 expands the scope, providing practical recommendations for the implementation of these methodologies onboard HTVs and their potential applications in other sectors within Boskalis. Finally, Section 7.4 presents concluding thoughts and reflections on this research.

7.1 Conclusion

This study was initiated to address the challenges faced by Boskalis in predicting maximum roll and pitch motions of HTVs. Accurate predictions are important to prevent exceeding the predefined motion limits, a critical factor in ensuring the integrity of the vessel and cargo. The current reliance on OCTOPUS, a system with limitations in handling nonlinearities and dependent on nowcast environmental data, led to the exploration of alternative ML-based approaches. This research developed three novel prediction approaches: NauticalNet, NauticalNet*, and HindHold. These approaches were compared with OCTOPUS, and their performance was evaluated using different validation strategies to address the main research question:

How do ML-based prediction approaches compare with (simulated) conventional prediction approaches when predicting maximum roll and pitch motions of an HTV?

The investigation led to several key insights, summarized as follows:

- **ML versus (Simulated) Conventional Approaches under Known LCs:** The ML-based approaches NauticalNet and NauticalNet* generally outperformed both OCTOPUS and HindHold when predicting maximum roll and pitch motions under known LCs. They achieved an

MAPE approximately half that of their (simulated) conventional counterparts. The superior performance of ML approaches under known LCs is attributed to their ability to identify and learn nonlinear relationships within the data, as well as to effectively handle errors in environmental data (Section 6.3.2, Chapter 6).

- **Performance in Extreme Response Scenarios:** In scenarios involving extreme response motion, OCTOPUS and HindHold exhibited a relative advantage over NauticalNet and NauticalNet*. This advantage is particularly evident in the predictions of maximum roll motion, where ML-based approaches often underestimated extreme motion values. This limitation in ML approaches for handling extreme response scenarios could be attributed to the underrepresentation of extreme motion events in the training dataset (Section 6.3.3, Chapter 6).

- **ML versus (Simulated) Conventional Approaches under Unknown LCs:** Under unknown LCs, the effectiveness of ML-based approaches decreased, as evidenced by an increase in MAPE by a factor of 1.5 to 2.0. This decrease in performance is due to the limited exposure of ML-based approaches to a variety of LCs, leading to difficulties in generalizing to new LCs (Section 6.3.4, Chapter 6).

- **Adaptability to Data from Unknown LCs:** NauticalNet and NauticalNet* exhibited rapid adaptation to unknown LCs. This adaptability is primarily due to their exposure to a vast amount of environmental data during training, enabling them to recognize patterns linking environmental factors with the vessel's behavior. With this foundational knowledge, the ML-based approaches can extrapolate to new LCs more effectively (related to transfer learning, a concept in which a model developed for one task is adapted for a related one with minimal additional data (Tan et al., 2018), see Section 6.3.5, Chapter 6).

- **Impact of Environmental Data Type:** The application of hindcast environmental data in NauticalNet and HindHold led to improved predictive performance compared to NauticalNet* and OCTOPUS, especially in predicting maximum roll motions. This improvement was more pronounced for the simulated conventional approach HindHold compared to OCTOPUS than for the ML-based approaches. The difference in the impact of environmental data types can be linked to the quality of the hindcast data, which proved to be higher than nowcast data due to its correction with actual observations and weather models (see Sections 6.1 and 6.3.1, Chapter 6).

ML-based prediction approaches have demonstrated their potential for superior performance in predicting the maximum roll and pitch motions of the HTV Target, compared to (simulated) conventional prediction approaches. Despite their limitations, particularly in extreme response motions and under unknown LCs, the potential of ML-based approaches is undeniable. The primary advantage of ML-based approaches comes from their retrospective methodology, learning from observations in a

backward manner. This approach inherently incorporates nonlinear phenomena in motion responses and also learns to adapt to errors in environmental data. In contrast, OCTOPUS does not inherently include nonlinear phenomena, although it attempts to linearize nonlinear elements, and is not able to adapt or consider errors in environmental data.

The impact of the type of environmental data on (simulated) conventional approaches is found to be greater compared to its impact on ML-based approaches. This is particularly evident in maximum roll motion, indicating that precise environmental conditions play a crucial role in this type of motion. This observation is reasonable since the pitch motion is heavily damped and less affected by specific environmental conditions, since it naturally diminishes through wave damping. In contrast, roll motions experience minimal wave damping, leading to resonance dominance and increased vulnerability to dynamic amplifications and heightened roll responses (see Section 2.2.6). Consequently, the accuracy of environmental data becomes more critical in these cases.

Furthermore, the substantial improvement in predicting maximum roll motion by ML-based approaches, as compared to (simulated) conventional approaches, considering the roll motion's inherent nonlinearity, substantiates the earlier assertion that ML-based approaches are more adept at incorporating nonlinear aspects. This effectiveness is further exemplified by the findings that ML-based approaches encountered fewer difficulties compared to (simulated) conventional approaches when dealing with higher nonlinear roll damping, caused by higher vessel speeds. Therefore, ML-based approaches' proficiency in handling increased vessel speed shows their capability to capture these nonlinearities, posing a significant advantage over (simulated) conventional approaches.

In conclusion, this research has provided a more nuanced understanding of the comparative performance of ML-based approaches with (simulated) conventional approaches, and the impact of the type of environmental data. Challenges such as the need for a more diverse range of LCs, the flawed environmental data quality, and the reduced efficacy of ML-based approaches in extreme motion response scenarios are identified, highlighting areas for future exploration and improvement. Consequently, Boskalis can now focus on addressing these challenges, informed by the insights gained from this study, rather than questioning the applicability of ML. The subsequent sections will explore recommendations for approaching these challenges, offering both theoretical and practical perspectives.

7.2 Theoretical Recommendations: Addressing Challenges

The underlying goal of this thesis is to improve vessel motion prediction approaches, thus increasing their effectiveness in providing operational guidance to adhere to operational limits. The challenges faced in this research, particularly the limitations of current ML-based approaches and environmental data quality, are not setbacks. Instead, they are integral steps towards achieving this underlying goal. Consequently, the next phase involves addressing these identified challenges. This section offers theo-

retical recommendations to tackle these issues for future research, leading to practical recommendations for Boskalis, and the wider industry, in the following section.

7.2.1 Environmental Data Quality

The accuracy of prediction approaches that rely on nowcast environmental data may be influenced by the quality of these data. In particular, OCTOPUS, which relies on a nowcast wave spectrum, is affected by this issue. The wave period features are observed to be rounded to whole numbers, potentially causing a shift in the peak of the wave spectrum in cases where the most frequently occurring waves have a period between whole numbers (see Section 6.1.2). This highlights the limitations of rule-based and fixed approaches, such as OCTOPUS, which are sensitive to data quality issues. On the other hand, NauticalNet*, which also utilizes nowcast environmental data, is less likely to be affected by data quality problems due to its non-rule-based nature and ability to learn retrospectively from data, potentially enabling it to handle errors present in the data or rely less on these features. Unlike (simulated) conventional approaches, ML-based methods have the capability to assign importance to different features on their own, without relying on the designer's knowledge. In terms of performance, the simulated conventional approach HindHold outperformed OCTOPUS significantly on the diverse test dataset from validation strategy I, clearly demonstrating the impact of data quality.

However, ML-based approaches could also benefit from improving data quality, as NauticalNet also demonstrated a significant improvement in predicting the maximum roll motion compared to NauticalNet* (see Section 5.1.2 for a discussion of why this improvement was more pronounced for roll than for pitch). Therefore, enhancing the quality of environmental data could benefit both ML-based approaches and conventional approaches such as OCTOPUS. To put it simply, one obvious recommendation would be to explore different nowcast environmental data providers and compare their data with Copernicus' hindcast environmental data. Data that closely resemble this hindcast environmental data should theoretically yield more accurate vessel motion predictions. This solution is particularly relevant for OCTOPUS. However, for ML-based approaches, if the nowcast environmental data provider is changed, a significant portion of the historical environmental data would also need to be updated. This presents a challenge for Boskalis, as training the ML-based approaches would only be possible if the new nowcast environmental data provider can supply the corresponding historical nowcast environmental data for each voyage. If this is not possible, there is little data to learn from, making this an unpractical solution for ML-based approaches.

An alternative solution that does not have this problem is to incorporate error estimates of the existing nowcast environmental data provider into the input features of NauticalNet*. Since NauticalNet* consists of 17 environmental data features, it is recommended to include 17 error or uncertainty estimate features in the input dataset corresponding to all these features. This would likely enhance NauticalNet*'s ability to generate predictions that are more closely aligned with those of NauticalNet.

Two methods for deriving the error or uncertainty estimates are outlined below:

1. **ML Model for Error Prediction:** According to Foldesi and Valdenegro-Toro (2022), uncertainty estimation methods in ML can address limitations such as overconfidence in predictions, which could be a problem when using faulty data. For Boskalis, the implementation of an ML model that predicts errors in environmental data, such as wind and wave conditions, could therefore improve the accuracy of vessel motion predictions. Training this model on historical data, for example, by relating nowcast with hindcast environmental data, can help identify and correct systematic errors in environmental data. Possible input features for these models could include various environmental parameters from the nowcast data. The target value would be the difference between the nowcast data and the corresponding hindcast data for the same target parameter (Foldesi and Valdenegro-Toro, 2022).

2. **Bayesian Uncertainty Estimation:** Héas et al. (2023) discuss a Bayesian approach to quantification of uncertainty in meteorological data. Boskalis can implement Bayesian uncertainty estimation to assess the confidence level of each data point in the nowcast environmental data. This method provides a probabilistic measure of uncertainty, which can be incorporated into the input features of NauticalNet* to improve its predictive accuracy. By understanding the uncertainty in the data, the ML model can weigh the inputs accordingly, potentially leading to more robust and reliable vessel motion predictions. For example, if the model estimates that there is a high level of uncertainty in the wind speed data, this information can be used to adjust the weight given to the wind speed in the vessel motion prediction model (Héas et al., 2023).

7.2.2 Performance in Extreme Response Scenarios

Extreme response motions are in this thesis identified at maximum roll motions above 8.0 degrees, and maximum pitch motions above 2.5 degrees (see Section 6.3.3) .The analysis of prediction errors in extreme response scenarios revealed a critical challenge: ML-based approaches exhibit higher errors when faced with rare, high-magnitude maximum roll and pitch motions. This issue probably originates from the underrepresentation of such extreme motion events in the training dataset, leading to a conservative bias in the predictions. Accurate prediction of these events is crucial as one of the primary reasons for the need to predict maximum vessel motions is to prevent unexpected high-magnitude responses. If such high-magnitude responses occur because they were not predicted, the vessel operator cannot take preventive action, resulting in the joints within the cargo, and between the cargo and the vessel, experiencing undesired high loads and stresses. Addressing this issue could involve refining the custom loss function and deploying strategic data augmentation techniques.

1. **Refining the Custom Loss Function and Ensemble Methods:**

    - Optimize β in the current custom loss function (see Chapter 4 'Execution', Section 4.2.1).

This customized loss function employs a weighting mechanism (β) to address the magnitude of errors in the significant domain. An increase of β should theoretically boost the model's sensitivity to higher-magnitude responses, reducing the conservative bias. However, this probably comes with a trade-off. The higher β becomes, the less accurate the ML-based approaches become for low-magnitude responses, thus influencing the general MAPE.

- Introduce an additional parameter, α, to apply a more nuanced penalty based on the nature of the prediction error. This parameter is particularly crucial for handling underpredictions, which are more prevalent in extreme scenarios. The value of α will vary depending on whether the prediction is an underprediction or overprediction, relative to the true value. This adaptive mechanism, in conjunction with β (which continues to govern the overall weighting), enhances the model's sensitivity to different types of prediction errors. The conditional assignment for α could be mathematically represented as follows (but a more gradual version could also be investigated):

$$\alpha_i = \begin{cases} 2.0, & \text{if } y_{\text{pred},i} \leq y_{\text{true},i}; \\ 0.5, & \text{otherwise.} \end{cases}$$

. Integrating it into the custom loss function:

$$L^*(y_\text{p}, y_\text{t}, \beta = 1.0, \alpha) = \frac{1}{N} \sum_{i=1}^{N} \left(\left(y_{\text{p}_i} - y_{\text{t}_i} \right)^2 \cdot (\log(|y_{\text{t}_i}| + 1) \cdot \beta + 1) * \alpha_i \right) \tag{7.1}$$

- An ensemble approach is suggested, where different ML models are tailored to specific response magnitudes. For example, Model A could target low significance magnitudes with a lower β value, Model B for medium significances, and Model C for extreme response cases, employing higher β and α values.

2. **Implementing Data Augmentation Techniques:** Given the importance of having a diverse and representative dataset, especially in the context of extreme roll and pitch motions, data augmentation techniques emerge as a possible solution. These techniques can artificially expand the dataset with more examples of extreme response conditions, thereby reducing the model's bias towards more frequently occurring moderate scenarios. A relevant approach in this regard can be found in the work by Shorten and Khoshgoftaar (2019) in their study "A Survey on Image Data Augmentation for Deep Learning". While their focus is on image data, the underlying principles can be adapted for time series data such as vessel motions.

   - **Synthetic Data Generation:** Generating synthetic data that simulates extreme response conditions can help in balancing the dataset. This could involve using physical models or

simulations to create realistic scenarios of extreme roll and pitch motions that might not be well-represented in the collected data.

- **Data Smoothing and Resampling:** Smoothing and resampling techniques can create new data points by interpolating between existing ones. This approach can be particularly useful in time series data, helping to create more examples of gradual transitions into extreme response conditions.

Implementing these data augmentation techniques would require careful consideration to ensure that the synthetic data is realistic, and that the augmentation does not introduce unintended biases. Boskalis can use these strategies to create a more robust and diverse dataset, potentially improving the performance of ML-based approaches in predicting extreme vessel motions. This would be a significant step forward in enhancing the predictive accuracy and reliability of these models in critical scenarios.

7.2.3 Performance under Unknown Loading Conditions

In addressing the challenge of prediction in unknown LCs, it is crucial to understand why current ML-based approaches, such as NauticalNet*, may not generalize well between different LCs. The core issue originates from the limited diversity in the dataset: with only 24 distinct LCs included, the models have not been exposed to a wide enough spectrum of scenarios. This limited exposure has hindered the ability to truly learn the relationship between LC values, the environment, and responses, as reflected in higher prediction errors (as addressed in Section 6.3.4). To enhance predictive accuracy under unknown LCs, the following strategic roadmap is proposed:

1. **Expand the Dataset with Diverse Loading Conditions:**

   (a) As per the commonly used 10-times rule and the 50-times rule suggested by Alwosheel et al. (2018), the objective is to compile a dataset comprising between 130 to 650 distinct LCs, considering the presence of only 13 LC features. It is recognized that dimensionality reduction could potentially lower the required range of LCs, though this might result in some loss of information (see Section 6.3.4). Such an approach is definitely worth investigating. For the time being, the range of 130 to 650 LCs is being utilized. Achieving even the minimum of this spectrum should markedly enhance the models' generalization capabilities.

   (b) Utilize data from other t-class HTVs within the Boskalis fleet to augment the current dataset. These vessels, sharing similar dimensions and hydrostatic conditions, offer a vast array of LCs. A rough estimation of the available data is outlined in Table 3. Empty sailing voyages represent voyages in which there was no cargo onboard the HTV.

T-class Transport Vessel	*Recorded Voyages 2009-2021*
Target	82 (24 empty sailing)*
Talisman	88 (35 empty sailing)
Transporter	68 (30 empty sailing)
Trustee	60 (16 empty sailing)
Triumph	107 (43 empty sailing)
Treasure	64 (26 empty sailing)
Total	**469 (174 empty sailing)**

Table 3: Estimation of available data from T-class HTVs.

* In this study, the empty sailing conditions were filtered out. Moreover, only 24 out of the remaining 58 available LCs were used. This choice was driven by the focus on data that was immediately accessible and supported by the corresponding annual reports, limiting the scope of this analysis to the period between 2012 and 2018. It is worth noting that a more extensive search could potentially uncover additional relevant data for the HTV Target.

(c) By incorporating these additional vessels, the 130-condition threshold is significantly exceeded, which should ideally bridge the gap between known and unknown LCs. In addition, it is also recommended to include empty sailing conditions to let ML-based approaches better learn the relationship between the LC and the response of the vessel. Increasing the amount of data available for a specific LC, such as an empty sailing condition, would likely benefit data augmentation techniques for other LCs.

2. **Refinement and Training of ML Models:**

(a) **Data Enrichment and Model Training:** Once the data from the six t-class HTVs has been combined, resulting in a total of 301 distinct LCs (295 transport operations and 6 empty sailing condition), the next phase is to thoroughly train NauticalNet. This training will utilize hindcast environmental data along with the extensive range of LCs and navigational information. The aim here is to create a robust model, NauticalNet, that is well-versed in a wide array of scenarios, enhancing its predictive accuracy across both known and unknown LCs.

(b) **Data Gap Filling and Augmentation:** In order to ensure that the dataset is complete and includes all the required information for each LC, the improved and enhanced NauticalNet is employed to fill in any sensor data that may be missing. This enables a more

comprehensive depiction of each LC. If NauticalNet's outputs are not suitable or applicable in certain scenarios, synthetic data augmentation techniques can be resorted to (as discussed earlier in this section). These techniques are especially beneficial for LCs that are not well-represented, as they enhance the dataset and improve the model's capability to handle a broader range of conditions.

(c) **Adapting NauticalNet* for Enhanced Predictions:** With the enhanced dataset, attention will turn to training and optimizing NauticalNet*. This stage involves integrating the α parameter into the custom loss function, a modification aimed at boosting the model's accuracy in extreme response situations by imposing a higher penalty on underpredictions. Additionally, during this phase, NauticalNet* will be trained using all 301 distinct LCs, and uncertainty estimates of the environmental data features.

3. **Implementation and Further Refinement:**

   - **Cross-Validation Technique Update:** A modification of the cross-validation technique is proposed to incorporate validations with unknown LCs. This modification aims to improve the reliability and predictive accuracy of models in unknown LCs, which is crucial for their practical application. By including unknown LCs in the validation set during model development, a more versatile model can be created.

   - **Continuous Model Enhancement:** Ongoing refinement of NauticalNet and NauticalNet* will be important to keep up with evolving maritime conditions and technological advances. This includes leveraging Boskalis' advanced computational resources for regular updates and adjustments of the custom loss function and cross-validation techniques.

   - **Operational Implementation:** The next step involves the operational implementation of NauticalNet* on t-class HTVs. This phase, detailed in the subsequent implementation section, marks the transition from theoretical modeling to practical application, utilizing and testing the model's effectiveness in real-world scenarios.

4. **Additional Considerations:** Additional refinements of the ML-based approaches are listed in Appendix G.

7.3 Practical Recommendations

With the primary focus of this thesis on t-class HTVs, the immediate goal should be to further develop and implement ML approaches tailored for these vessels. However, looking to the future, it is equally important to expand this scope to other vessel types in Boskalis' fleet, such as heavy lift, cable laying, and support vessels. Each of these vessel types has different operational characteristics and

LCs, requiring the development of tailored ML models with specific feature sets for accurate motion prediction. This section will first describe recommendations for implementation onboard of HTVs. Thereafter the focus will shift towards other types of vessels.

7.3.1 Implementation onboard of HTVs

Given the challenges highlighted in this thesis, particularly in generalizing across various LCs using ML-based approaches, a well-defined strategy for the implementation on t-class HTVs is important. A practical step forward is to enhance NauticalNet* by integrating comprehensive data from all relevant vessels, enriching the model's exposure to a wide array of scenarios. The following steps and considerations are critical for an effective implementation:

1. **Group-Based Implementation Strategy:** Initially, the focus should be on implementing ML-based approaches for the t-class HTVs as a collective group. This approach will leverage the comprehensive dataset obtained from the 301 distinct LCs across these vessels, providing a robust training ground for a future version of NauticalNet*. To effectively manage data from different vessels, each t-class HTV's dataset, such as Target and Triumph, will be prepared individually. This preparation includes preprocessing and data cleaning steps, as outlined in this thesis (see Chapter 4), to ensure consistency and quality. Each vessel will be assigned a unique identifier (e.g., Target as 0, Triumph as 1) to maintain a clear distinction within the dataset. After these preprocessing steps, the datasets will be merged. This strategy ensures that the same features result from the preprocessing and data cleaning stages, thus maintaining consistency across the datasets. The date/time values will be dropped before merging to address any potential discrepancies related to temporal data.

2. **Computation Requirements:** Evaluate and guarantee sufficient computational capability on board. Increased computational power enables quicker dynamic retraining of the models, essential for adapting to new situations and data in real-time, with every new data point recorded every 15 minutes. It would be ideal to be able to retrain the ML-based approach every 15 minutes with a new observed data point. The ML techniques developed in this study were trained using a computing power of 16GB RAM, with 4 physical and 4 virtual cores. The training process for the ML-based approaches took approximately 2 hours. Two feasible solutions to reach the 15 minute retraining are:

   (a) **Upgrade Onboard Hardware**: Enhance the onboard system with a server-grade CPU, like Intel Xeon or AMD EPYC, and at least 64GB of RAM, complemented by a high-performance GPU such as NVIDIA Tesla. This upgrade can significantly accelerate the training process to meet the 15-minute retraining interval.

   (b) **Cloud Computing Integration**: Using a cloud computing service, for example, AWS

EC2 P3 or G4 instances, optimized for ML tasks. This setup can facilitate rapid model retraining off-board, in line with the 15-minute data update frequency.

3. **Adaptability and Continuous Learning:** The adaptability of ML-based approaches to unknown LCs, as was seen by using the iterative training approach, is a key strength. ML-based approaches can quickly adjust to new LCs as a result of their initial extensive training on diverse environmental data. To improve the accuracy of the prediction to unknown LCs, continuous learning and updates during the voyages are recommended. Incorporating methods such as transfer learning and few-shot learning techniques will probably further increase the adaptability of models.

4. **Uncertainty Quantification:** Implement uncertainty quantification to evaluate the reliability of predictions, particularly in unknown LCs, helping operators in their decision making.

5. **Incorporation of OCTOPUS predictions:** In the initial phase, where the dataset might not cover all LCs, include OCTOPUS predictions as input features to NauticalNet*. This integration can provide a safety net and enhance the model's performance (as was seen in this thesis). As the model evolves and becomes capable of generalizing across conditions, the reliance on OCTOPUS predictions can be re-evaluated and potentially phased out.

6. **Phased Implementation Across Vessel Types:** After successful implementation in t-class HTVs, gradually extend the ML-based prediction models to other vessel types. Each vessel type would require an assessment of specific features and conditions unique to its operational environment, necessitating individual model development and training, as will be discussed in the following section. The eventual goal could be to create a vessel independent model by incorporating vessel dimension features such as vessel length, width, and a block coefficient.

7.3.2 Development of ML-based Approaches for Other Business Units

ML-based prediction approaches have the potential to be transformative and relevant for other business units within Boskalis. Although cable-laying and heavy-lift vessels have different operational characteristics compared to transport operations, it is likely that tailored ML-approaches can predict vessel motions for these types of vessel as well. Excessive vessel motion in cable laying vessels can increase cable stress, while for heavy-lift it can lead to large amplitude swings of the lifted cargo resulting in large moments and loads on the equipment and vessels. Therefore, an accurate prediction of the vessel motion is important for both operations. ML-based approaches for cable-laying and heavy-lift vessels would require different architectures and input features compared to those used for transport vessels. Since cable-laying and heavy-lift vessels typically operate over shorter distances, the accuracy of environmental data for their specific operation areas becomes even more critical. Having a reliable environmental data provider that can deliver precise data for these smaller areas would significantly

improve the effectiveness of ML models in these operations. It is also worth noting that the LCs for heavy-lift vessels do not remain constant throughout the voyages, making data gathering potentially more challenging for heavy-lift operations. Additionally, when developing these models, it is important to consider other vessel motions, such as heaving or sway, which may also be important for these types of operations.

Beyond vessel motion prediction systems, there are numerous opportunities for ML-based approaches in other areas of Boskalis' operations. In the "Dredging & Inland Infra" division, ML can optimize dredging activities, predict maintenance needs, and improve project efficiency. In the "Towage & Salvage" division, ML can assist in planning towing paths and salvage strategies. For heavy lift operations, ML models could predict the duration required for tasks like monopile installation, while for cable laying, ML could classify soil types or predict soil density. These examples illustrate the vast potential of ML in various aspects of Boskalis' operations across all its divisions, including "Dredging & Inland Infra" and "Towage & Salvage." It is recommended that Boskalis continues to explore and invest in ML-based solutions across its different business units. This could involve setting up dedicated teams or departments focused on developing and implementing ML solutions, collaborating with academic institutions or technology partners, and investing in data infrastructure to support these initiatives. By continuing the journey started with this project, Boskalis can leverage ML to improve operational efficiency, reduce risks, and maintain its competitive edge in the industry.

7.4 Creating New Horizons

This thesis, centered on the application of ML in vessel motion prediction for Boskalis' HTV Target, serves as an exploratory step in understanding the potential of ML-based approaches in offshore engineering. Although it lays the foundation for further exploration, it is important to recognize that the results primarily highlight the potential of ML-based approaches rather than their immediate readiness for widespread operational implementation.

While the scope of data used in this study provides valuable insights, it also reveals limitations in developing an approach that can predict across a wider array of operational scenarios, particularly under unknown LCs where the limited data scope impacted the model's effectiveness. However, it is noteworthy that this thesis explored just a fraction of the available data, concentrating on 24 out of 57 conditions for the HTV Target alone. The potential for ML models becomes more pronounced when considering the larger dataset available within Boskalis, which encompasses around 301 distinct LCs for other t-class HTVs.

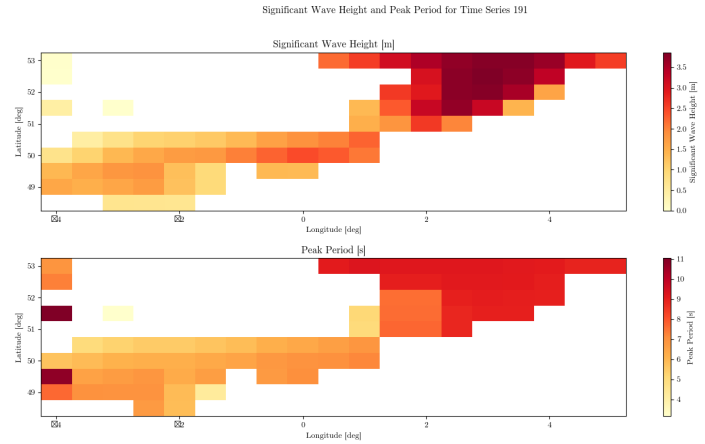
Reflecting on the broader objectives and scope of this research, the potential of ML compared to conventional approaches such as OCTOPUS has been established, yet there is a significant journey ahead to fully harness this potential. Comparative analysis and evaluation of environmental data types

have laid the groundwork for more comprehensive future studies, highlighting the substantial promise that ML-based approaches hold in offshore engineering, particularly for companies like Boskalis. The success of these approaches hinges on their ability to process and learn from large, diverse datasets, underscoring the need to expand these datasets to encompass a broader range of LCs. The insights gained are invaluable, guiding future efforts towards the development of a universal ML model for every degree of freedom (DOF) for each type of vessel. This points to an exciting future for research as more data become available and innovative techniques are developed, aligning practical maritime needs with academic innovation. With continued investment in data collection and model refinement, ML-based approaches could become a powerful tool in vessel motion prediction.

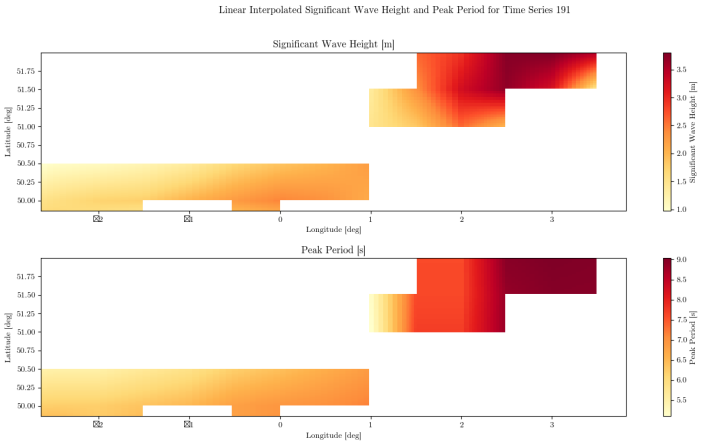
To achieve this ambitious goal, ongoing collaboration between industry and academia is essential. The synergy of Boskalis' extensive operational data and the expertise from academia is envisioned to propel this field to new heights. This journey, which includes steady data collection, model enhancement, and the adoption of the latest computational technologies, is dedicated to bolstering operational safety and efficiency, thereby reinforcing Boskalis' position as a frontrunner in the offshore energy sector. Embarking on this steadfast journey towards innovation and excellence resonates perfectly with Boskalis' ethos of 'Creating New Horizons', encapsulating the spirit of continuous advancement and aspiration that marks the conclusion of this thesis.

Appendix

A. Linear Interpolation Losses



(a)



(b)

Figure 67: Illustration of Data Losses in Environmental Data due to Linear Interpolation for Time Series 191

B. Feature Description

Table 4: Feature Description Group 1: Temporal Data

Name	**Description**
Arrival date	This parameter represents the vessel's arrival date from the voyage, from which the data point was extracted.
data_and_time	This parameter represents the date and time information that corresponds to the time that the geographical location data was recorded.
Date/Time	This parameter represents the time on which the General Hydrostatics (GHS) were logged in the vessel's system.
Departure date	This parameter represents the vessel's departure date from the voyage, from which the data point was extracted.
Time Series ID	This parameter assigns a time series identifier to the data point.

Table 5: Feature Description Group 2: Geospatial Data

Name	**Description**
latitude	This parameter represents the latitude coordinate at which the sensor data, including the vessel speed and course were recorded.
longitude	This parameter represents the longitude coordinate at which the sensor data, including the vessel speed and course were recorded.
Model bathymetry $[m]$	This parameter represents the depth of water from the surface to the bottom of the ocean.

Table 6: Feature Description Group 3A: Environmental Data: Weather Conditions

Name	Description
Air density over the oceans	This parameter is the mass of air per cubic metre over the oceans, derived from the temperature, specific humidity and pressure at the lowest model level in the atmospheric model.
Coefficient of drag with waves	This parameter is the resistance that ocean waves exert on the atmosphere.
10 metre U wind component	This parameter is the eastward component of the 10m wind. It is the horizontal speed of air moving towards the east, at a height of ten metres above the surface of the Earth, in metres per second.
10 metre V wind component	This parameter is the northward component of the 10m wind. It is the vertical speed of air moving towards the east, at a height of ten metres above the surface of the Earth, in metres per second.
10 metre wind direction	This parameter is the direction from which the "neutral wind" blows, in degrees clockwise from true north, at a height of ten metres above the surface of the Earth.
10 metre wind gust since previous post-processing	This parameter is the maximum 3 second wind at 10 m height as defined by the World Meteorological Organization.
10 metre wind speed	This parameter is the horizontal speed of the "neutral wind", at a height of ten metres above the surface of the Earth. The units of this parameter are metres per second.
Free convective velocity over the oceans	The parameter represents the vertical velocity of updraughts resulting from free convection, a fluid motion caused by buoyancy forces and density gradients. It helps estimate the influence of wind gusts on ocean wave development.
Instantaneous 10 metre wind gust	This parameter is the maximum wind gust at the specified time, at a height of ten metres above the surface of the Earth.
Neutral wind at 10 m u-component $[\frac{m}{s}]$	This parameter is the eastward component of the 'neutral wind', at a height of 10 metres above the surface of the Earth.
Neutral wind at 10 m v-component	This parameter is the northward component of the 'neutral wind', at a height of 10 metres above the surface of the Earth.

Continued on next page

Table 6 – *Continued from previous page*

Name	**Description**
Total precipitation	This parameter is the accumulated liquid and frozen water, comprising rain and snow, that falls to the Earth's surface.

Table 7: Feature Description Group 3B: Environmental Data: Wave Conditions

Name	**Description**
current_direction [deg]	This parameter represents the sea current's direction, measured in degrees.
current_speed [kn]	This parameter represents the sea current's speed, measured in knot.
Normalized energy flux into ocean	This parameter is the normalised vertical flux of turbulent kinetic energy from ocean waves into the ocean.
Normalized energy flux into waves	This parameter is the normalised vertical flux of energy from wind into the ocean waves. A positive flux implies a flux into the waves.
Normalized stress into ocean	This parameter is the normalised surface stress, or momentum flux, from the air into the ocean due to turbulence at the air-sea interface and breaking waves.
Maximum individual wave height	This parameter is an estimate of the height of the expected highest individual wave within a 20 minute time window.
Mean direction of total swell	This parameter is the mean direction of waves associated with swell.
Mean direction of wind waves	The mean direction of waves generated by local winds.
Mean period of total swell	This parameter is the average time it takes for two consecutive wave crests, on the surface of the ocean/sea associated with swell, to pass through a fixed point.
Mean period of wind waves	This parameter is the average time it takes for two consecutive wave crests, on the surface of the ocean/sea generated by local winds, to pass through a fixed point.
Mean square slope of waves	This parameter can be related analytically to the average slope of combined wind-sea and swell waves.

Continued on next page

Table 7 – *Continued from previous page*

Name	**Description**
Mean wave direction	This parameter is the mean direction of ocean/sea surface waves.
Mean wave direction of first swell partition	This parameter is the mean direction of waves in the first swell partition.
Mean wave direction of second swell partition	This parameter is the mean direction of waves in the second swell partition.
Mean wave direction of third swell partition	This parameter is the mean direction of waves in the third swell partition.
Mean wave period	This parameter is the average time it takes for two consecutive wave crests, on the surface of the ocean/sea, to pass through a fixed point.
Mean wave period based on first moment	This parameter is the reciprocal of the mean frequency of the wave components that represent the sea state.
Mean wave period based on first moment for swell	This parameter is the reciprocal of the mean frequency of the wave components associated with swell.
Mean wave period based on first moment for wind waves	This parameter is the reciprocal of the mean frequency of the wave components generated by local winds.
Mean wave period based on second moment for swell	This parameter is equivalent to the zero-crossing mean wave period for swell.
Mean wave period based on second moment for wind waves	This parameter is equivalent to the zero-crossing mean wave period for waves generated by local winds.
Mean wave period of first swell partition	This parameter is the mean period of waves in the first swell partition.
Mean wave period of second swell partition	This parameter is the mean period of waves in the second swell partition.
Mean wave period of third swell partition	This parameter is the mean period of waves in the third swell partition.

Continued on next page

Table 7 – *Continued from previous page*

Name	**Description**
Mean zero-crossing wave period	This parameter represents the mean length of time between occasions where the sea/ocean surface crosses mean sea level.
Peak wave period	his parameter represents the period of the most energetic ocean waves generated by local winds and associated with swell.
Period corresponding to maximum individual wave height	This parameter is the period of the expected highest individual wave within a 20-minute time window.
Significant height of combined wind waves and swell	This parameter represents the average height of the highest third of surface ocean/sea waves generated by wind and swell.
Significant height of total swell	This parameter represents the average height of the highest third of surface ocean/sea waves associated with swell.
Significant height of wind waves	This parameter represents the average height of the highest third of surface ocean/sea waves generated by the local wind.
Significant wave height of first swell partition	This parameter represents the average height of the highest third of surface ocean/sea waves associated with the first swell partition.
Significant wave height of second swell partition	This parameter represents the average height of the highest third of surface ocean/sea waves associated with the second swell partition.
Significant wave height of third swell partition	This parameter represents the average height of the highest third of surface ocean/sea waves associated with the second third partition.
Wave spectral directional width	This parameter indicates whether waves (generated by local winds and associated with swell) are coming from similar directions or from a wide range of directions.
Wave spectral directional width for swell	This parameter indicates whether waves associated with swell are coming from similar directions or from a wide range of directions.
Wave spectral directional width for wind waves	This parameter indicates whether waves generated by the local wind are coming from similar directions or from a wide range of directions.
Wave Spectral Skewness	This parameter is a statistical measure used to forecast extreme or freak ocean/sea waves. It is a measure of the asymmetry of the probability density function of the sea surface elevation.

Table 8: Feature Description Group 4: Loading Conditions

Name	Description
Cargo Id	This parameter summarizes information about the cargo and recording date.
vessel_course [deg]	This parameter describes the heading of the vessel w.r.t. the north. It is measured in degrees.
displacement [T]	This parameter represents the total weight of the ship, measured indirectly by calculating the displaced water weight. It is measured in tonnes.
fsm [Tm]	This parameter represents the "free surface moment" (FSM). It is the product of the ship's weight and the elevation of the center of gravity (C.G.). It is determined to create an equivalent righting moment as the shifting of tank contents when the ship heels.
gg [m]	This parameter represents the change in metacentric height due to free surface moments. The change is chosen so the righting moment from a slight change in heel matches the moment generated by the tank content's shift.
gm [m]	This parameter represents the distance from the center of gravity to the metacenter.
gmfluid [m]	This parameter represents the distance from the center of gravity to the metacenter corrected with the free surface correction.
Project ID	This parameter assigns the project ID to which the time series belongs to the data point.
rxx [m]	This parameter represents the radius of gyration about the x axis.
ryy [m]	This parameter represents the radius of gyration about the y axis.
rzz [m]	This parameter represents the radius of gyration about the z axis.
vessel_speed [kn]	This parameter describes the speed of the vessel. It is measured in knots.
taft [m]	This parameter describes the draft at the aft side of the vessel.
tfwd [m]	This parameter describes the draft at the forward side of the vessel.
tpitch [s]	This parameter describes the natural pitch period of the system.

Continued on next page

Table 8 – *Continued from previous page*

Name	**Description**
troll [s]	This parameter describes the natural roll period of the system.
xcg [m]	This parameter describes the location of center of gravity (CoG) in x-direction, measured from the aft perpendicular.
ycg [m]	This parameter describes the location of center of gravity (CoG) in y-direction, measured from the center line of the vessel.
zcg [m]	This parameter describes the location of center of gravity (CoG) in z-direction, measured from the keel of the vessel.

C. Roll Prediction Plots

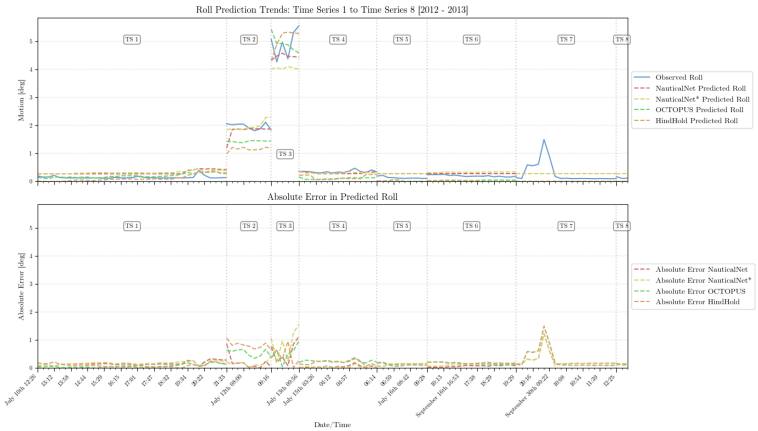


Figure 68: Roll Predictions vs. Observed Roll Motions: Time Series 1 to 8

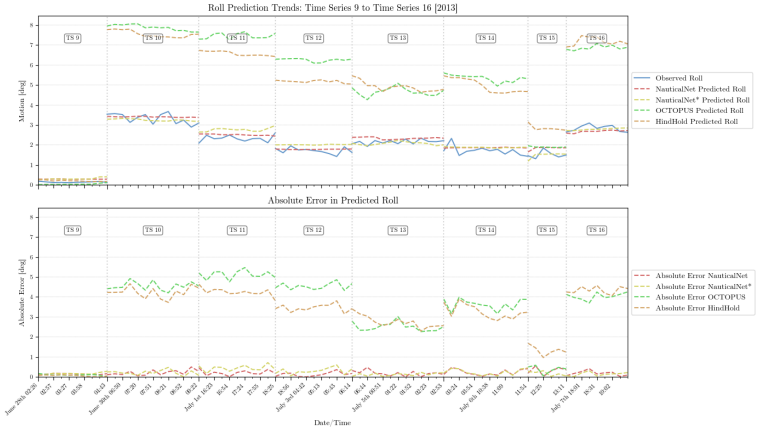


Figure 69: Roll Predictions vs. Observed Roll Motions: Time Series 9 to 16

Figure 70: Roll Predictions vs. Observed Roll Motions: Time Series 17 to 24

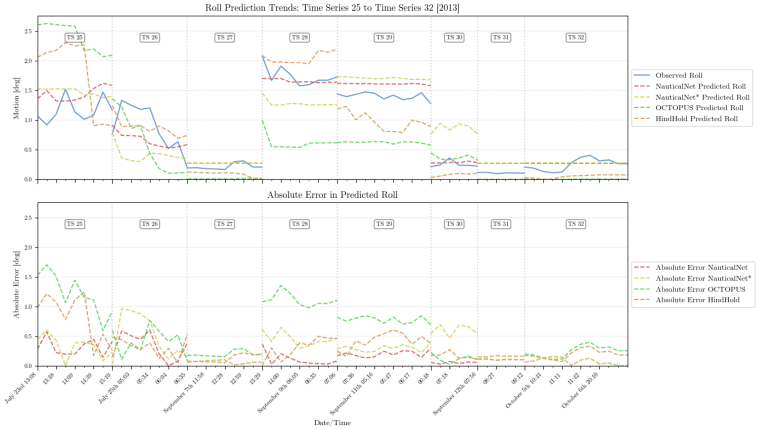


Figure 71: Roll Predictions vs. Observed Roll Motions: Time Series 25 to 32

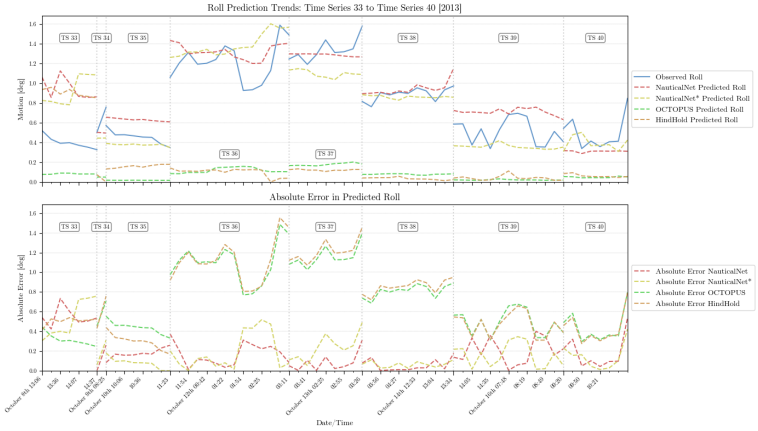


Figure 72: Roll Predictions vs. Observed Roll Motions: Time Series 33 to 40

Figure 73: Roll Predictions vs. Observed Roll Motions: Time Series 41 to 48

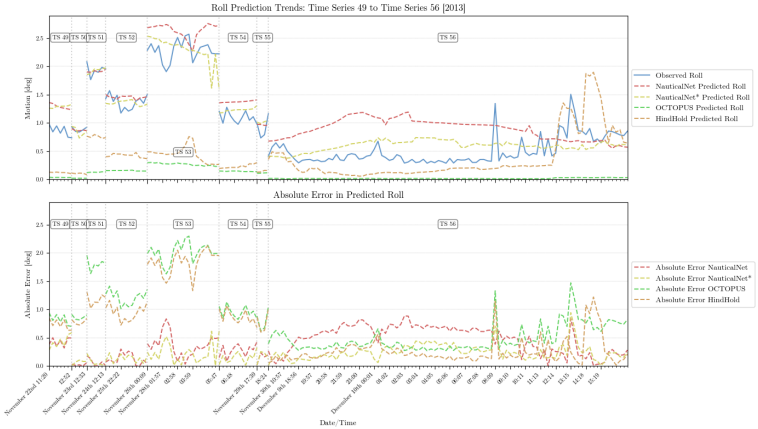


Figure 74: Roll Predictions vs. Observed Roll Motions: Time Series 49 to 56

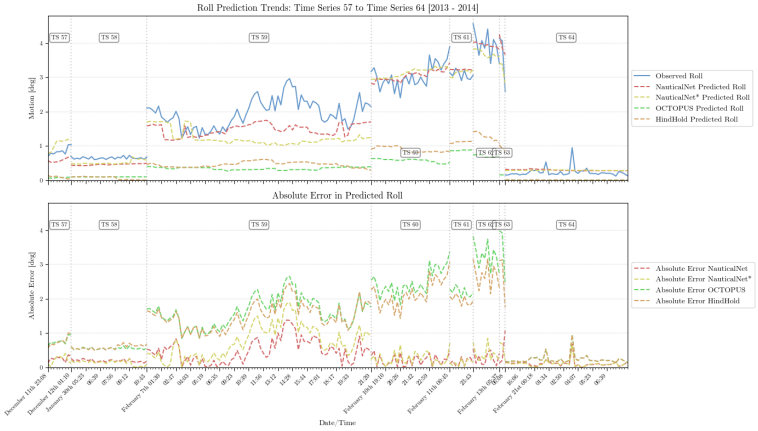


Figure 75: Roll Predictions vs. Observed Roll Motions: Time Series 57 to 64

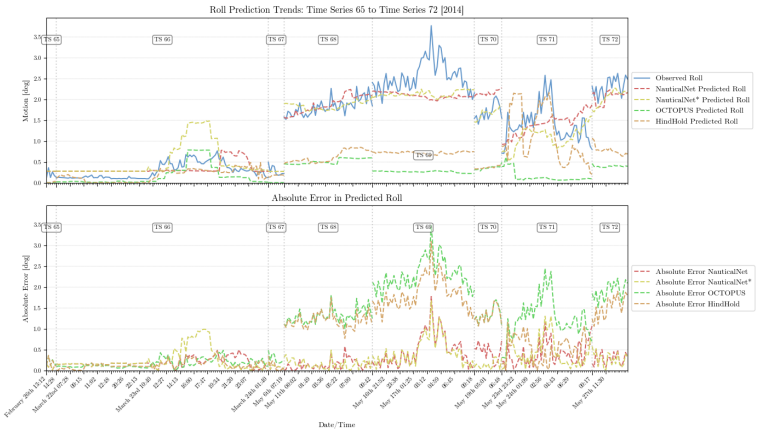


Figure 76: Roll Predictions vs. Observed Roll Motions: Time Series 65 to 72

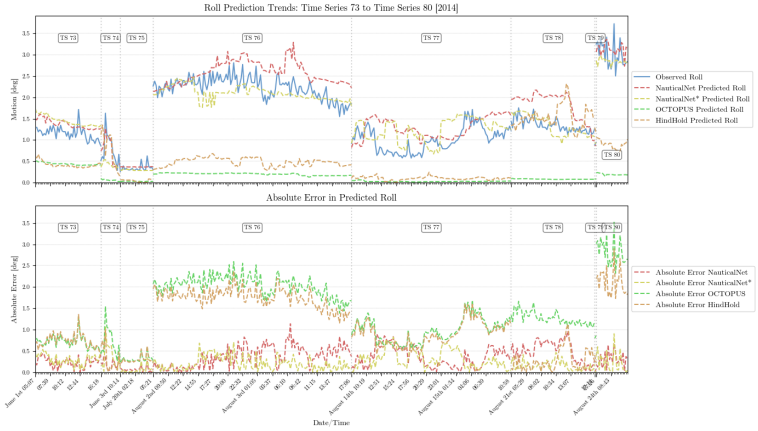


Figure 77: Roll Predictions vs. Observed Roll Motions: Time Series 73 to 80

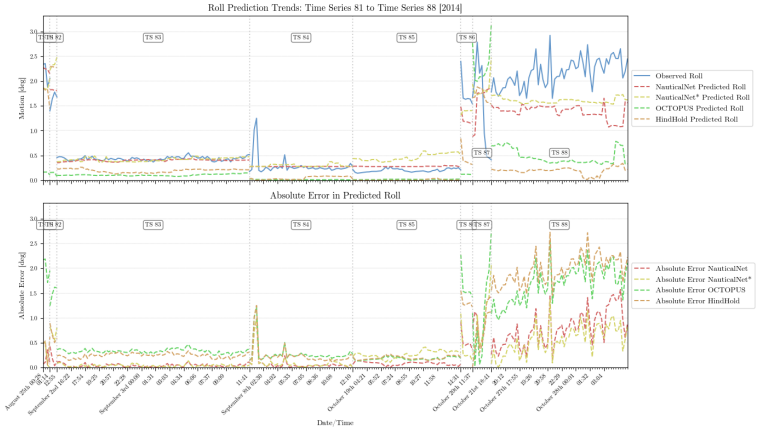


Figure 78: Roll Predictions vs. Observed Roll Motions: Time Series 81 to 88

Figure 79: Roll Predictions vs. Observed Roll Motions: Time Series 89 to 96

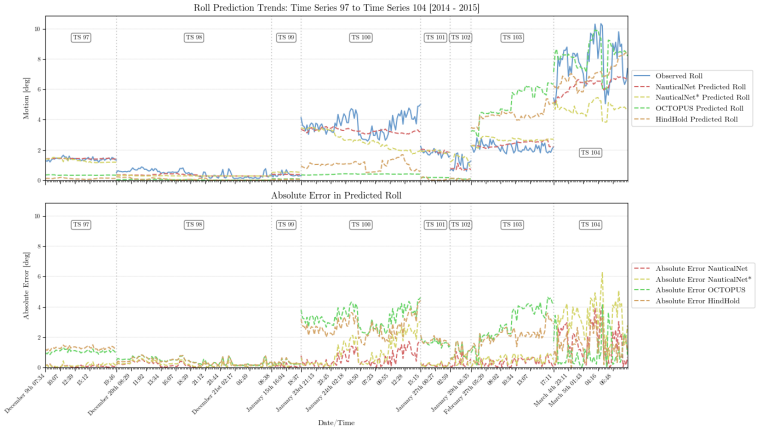


Figure 80: Roll Predictions vs. Observed Roll Motions: Time Series 97 to 104

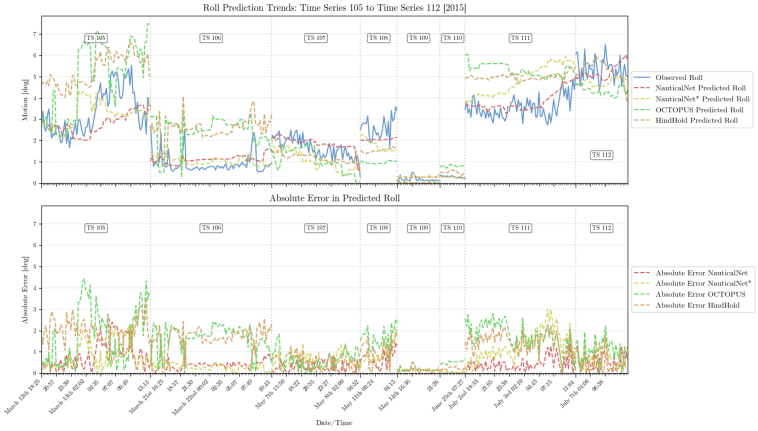


Figure 81: Roll Predictions vs. Observed Roll Motions: Time Series 105 to 112

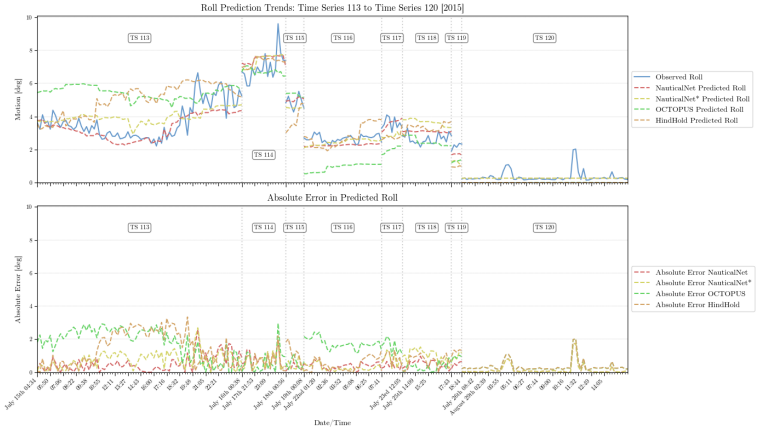


Figure 82: Roll Predictions vs. Observed Roll Motions: Time Series 113 to 120

Figure 83: Roll Predictions vs. Observed Roll Motions: Time Series 121 to 128

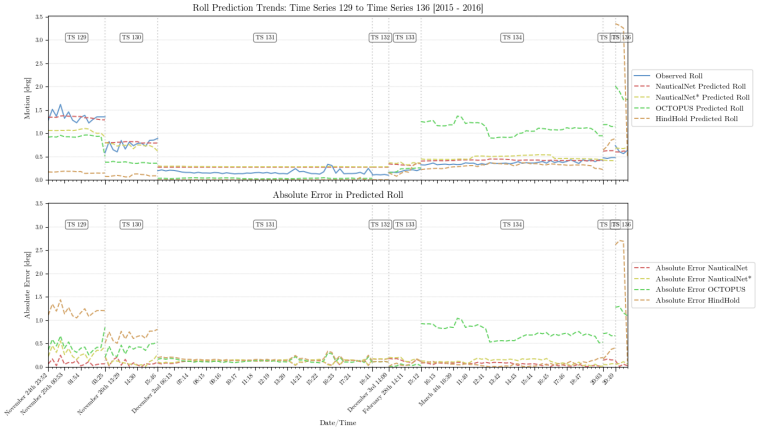


Figure 84: Roll Predictions vs. Observed Roll Motions: Time Series 129 to 136

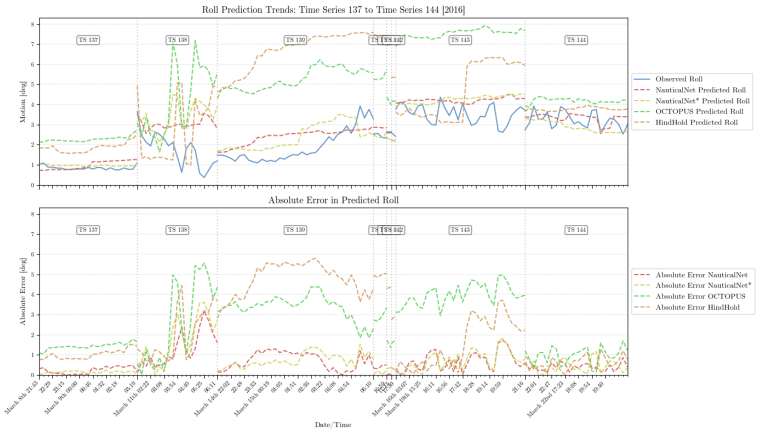


Figure 85: Roll Predictions vs. Observed Roll Motions: Time Series 137 to 144

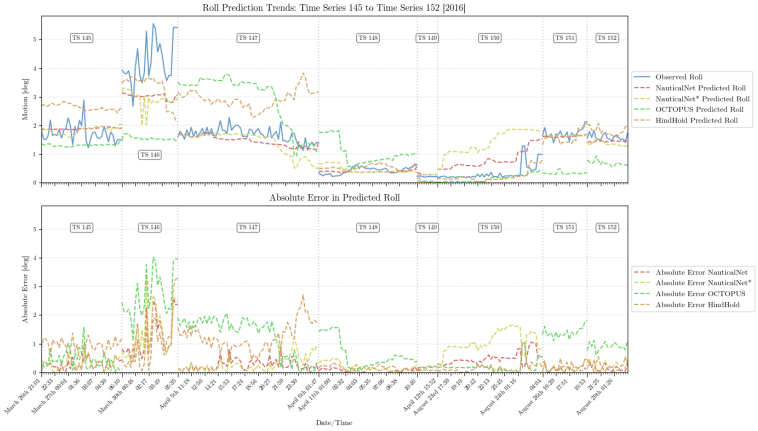


Figure 86: Roll Predictions vs. Observed Roll Motions: Time Series 145 to 152

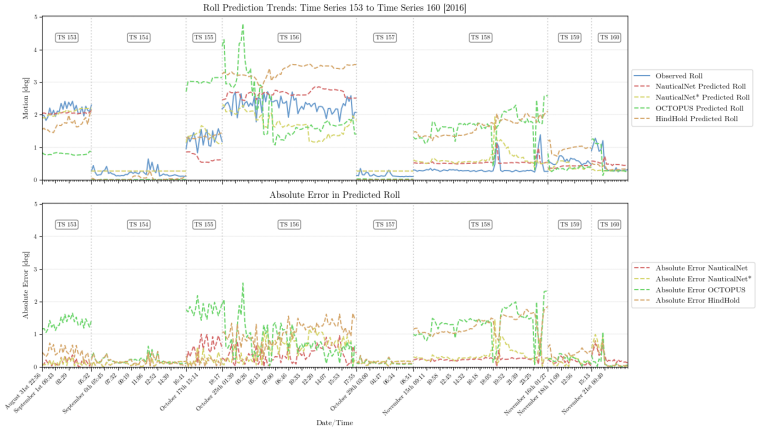


Figure 87: Roll Predictions vs. Observed Roll Motions: Time Series 153 to 160

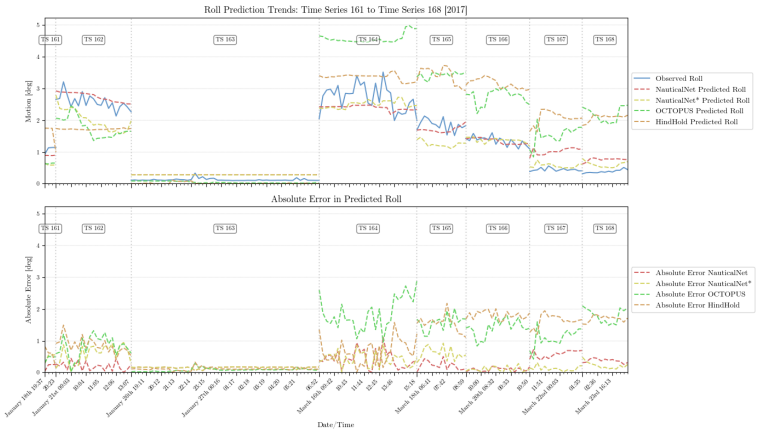


Figure 88: Roll Predictions vs. Observed Roll Motions: Time Series 161 to 168

Figure 89: Roll Predictions vs. Observed Roll Motions: Time Series 169 to 176

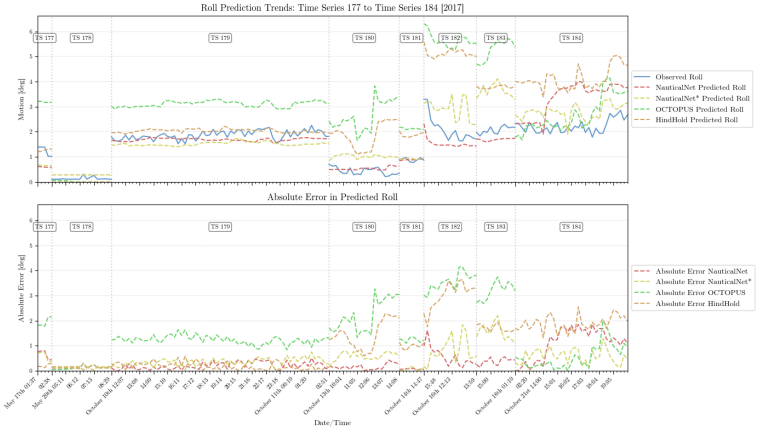


Figure 90: Roll Predictions vs. Observed Roll Motions: Time Series 177 to 184

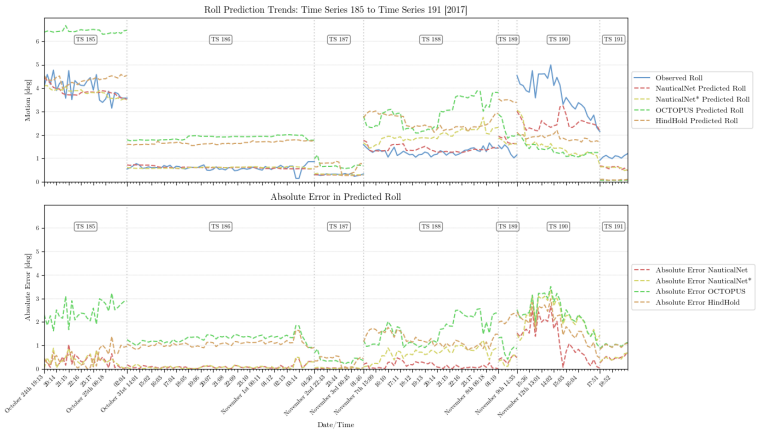


Figure 91: Roll Predictions vs. Observed Roll Motions: Time Series 185 to 191

D. Pitch Prediction Plots

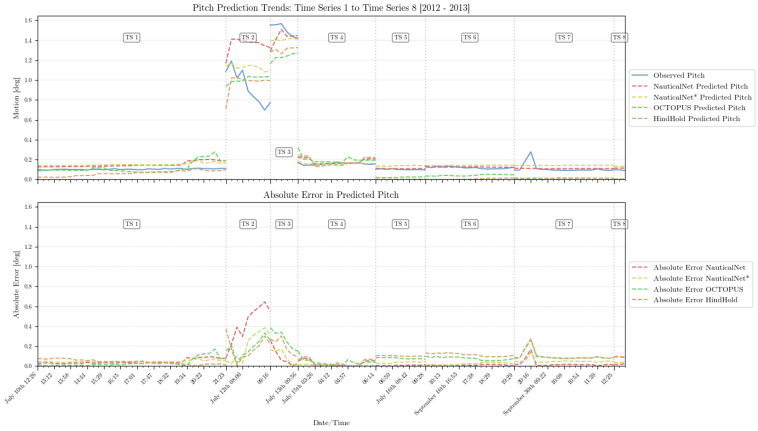


Figure 92: Pitch Predictions vs. Observed Pitch Motions: Time Series 1 to 8

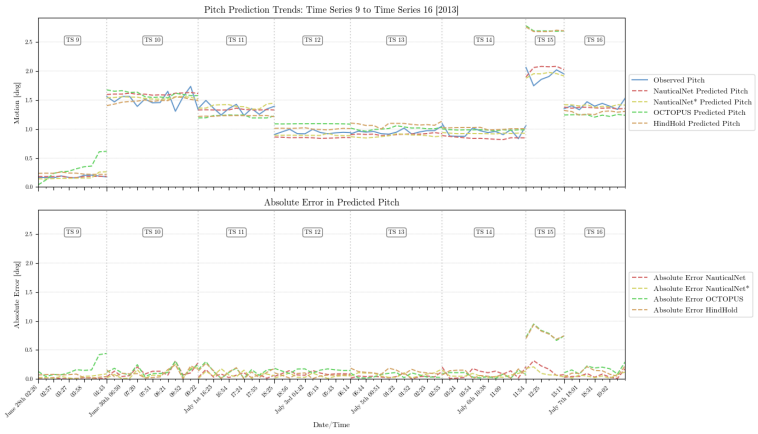


Figure 93: Pitch Predictions vs. Observed Pitch Motions: Time Series 9 to 16

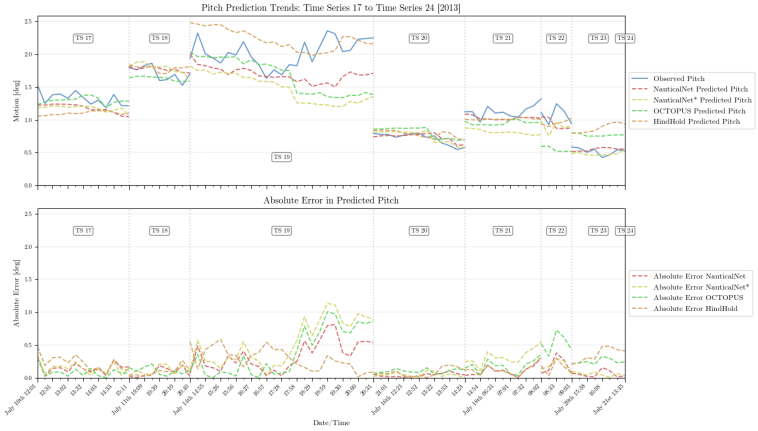


Figure 94: Pitch Predictions vs. Observed Pitch Motions: Time Series 17 to 24

Figure 95: Pitch Predictions vs. Observed Pitch Motions: Time Series 25 to 32

Figure 96: Pitch Predictions vs. Observed Pitch Motions: Time Series 33 to 40

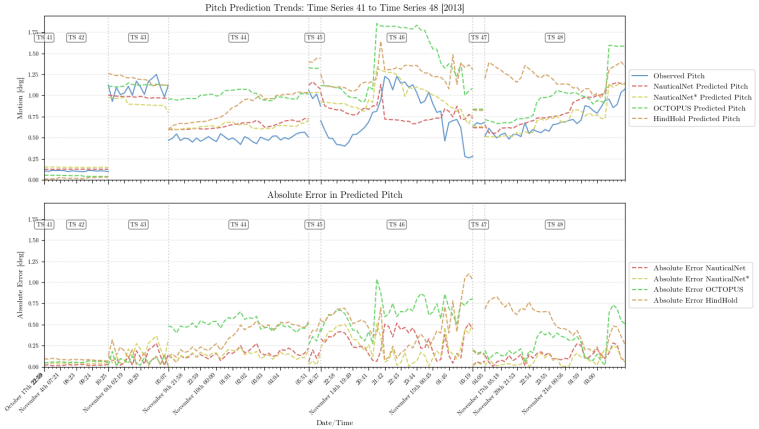


Figure 97: Pitch Predictions vs. Observed Pitch Motions: Time Series 41 to 48

Figure 98: Pitch Predictions vs. Observed Pitch Motions: Time Series 49 to 56

Figure 99: Pitch Predictions vs. Observed Pitch Motions: Time Series 57 to 64

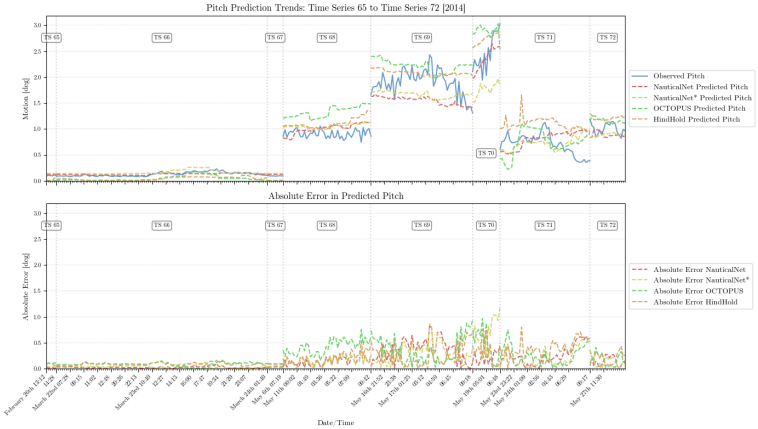


Figure 100: Pitch Predictions vs. Observed Pitch Motions: Time Series 65 to 72

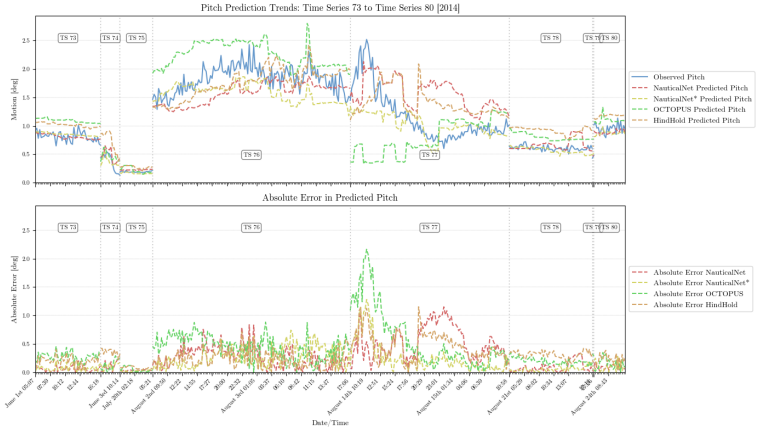


Figure 101: Pitch Predictions vs. Observed Pitch Motions: Time Series 73 to 80

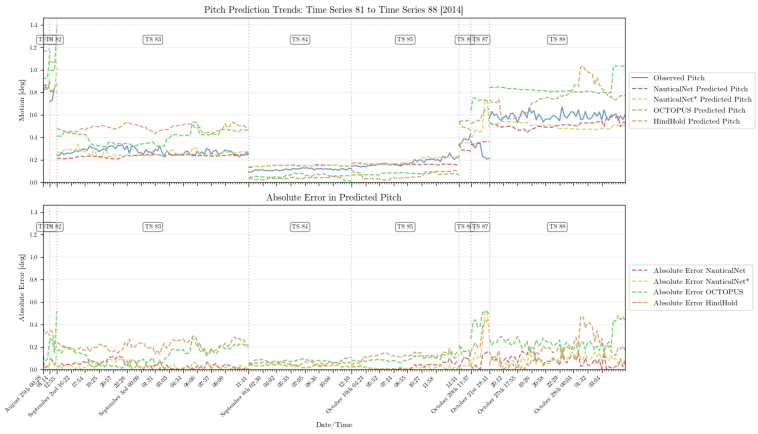


Figure 102: Pitch Predictions vs. Observed Pitch Motions: Time Series 81 to 88

Figure 103: Pitch Predictions vs. Observed Pitch Motions: Time Series 89 to 96

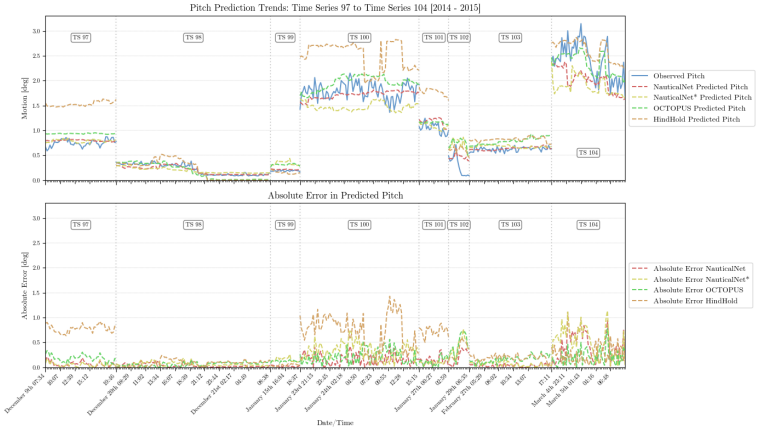


Figure 104: Pitch Predictions vs. Observed Pitch Motions: Time Series 97 to 104

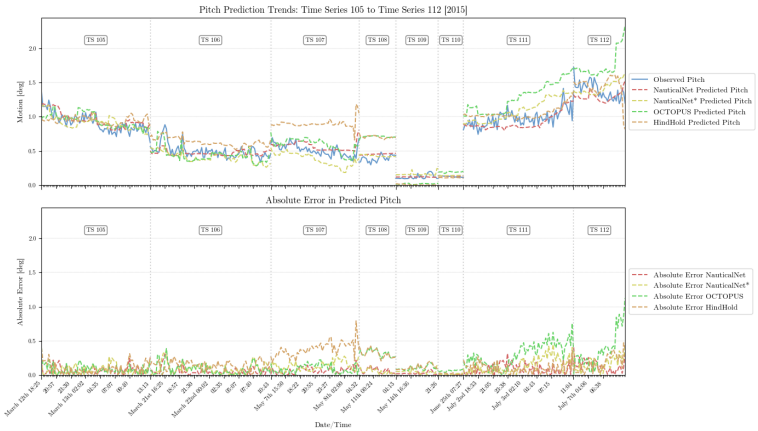


Figure 105: Pitch Predictions vs. Observed Pitch Motions: Time Series 105 to 112

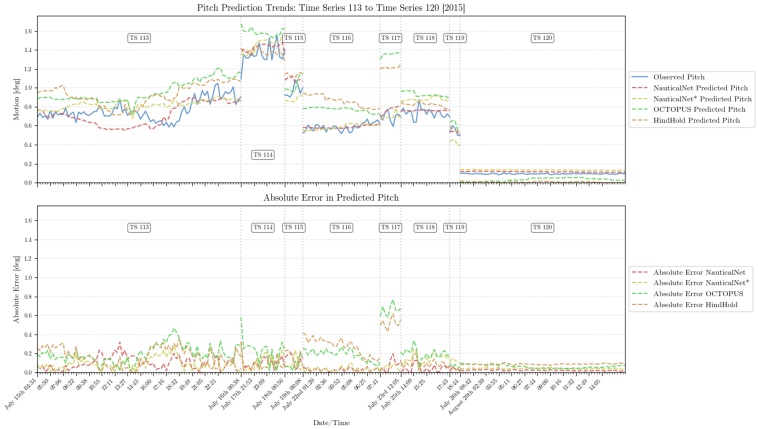


Figure 106: Pitch Predictions vs. Observed Pitch Motions: Time Series 113 to 120

Figure 107: Pitch Predictions vs. Observed Pitch Motions: Time Series 121 to 128

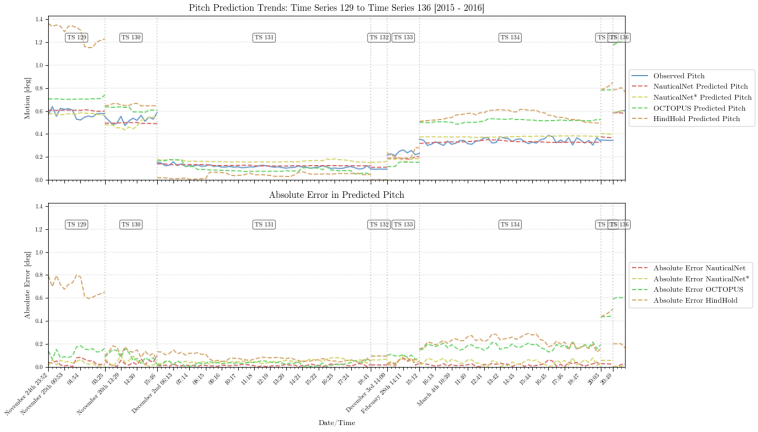


Figure 108: Pitch Predictions vs. Observed Pitch Motions: Time Series 129 to 136

Figure 109: Pitch Predictions vs. Observed Pitch Motions: Time Series 137 to 144

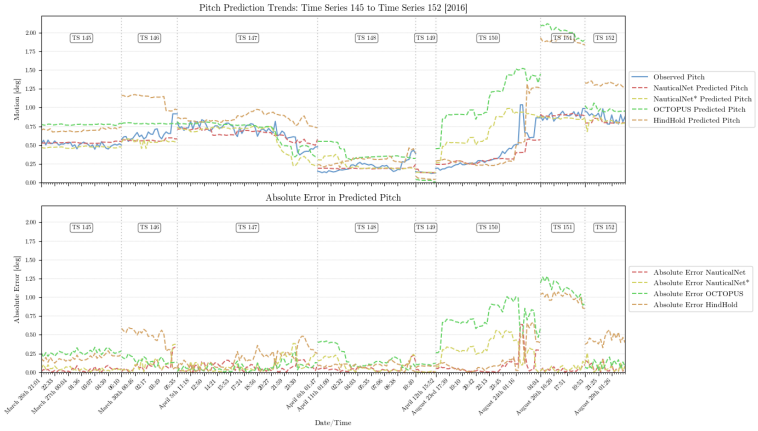


Figure 110: Pitch Predictions vs. Observed Pitch Motions: Time Series 145 to 152

Figure 111: Pitch Predictions vs. Observed Pitch Motions: Time Series 153 to 160

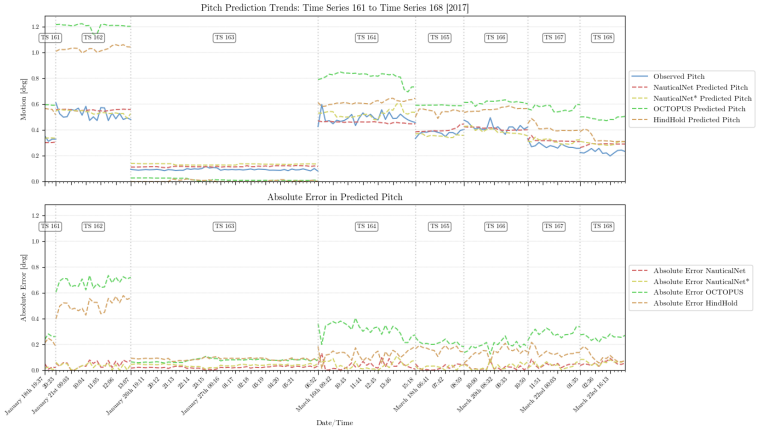


Figure 112: Pitch Predictions vs. Observed Pitch Motions: Time Series 161 to 168

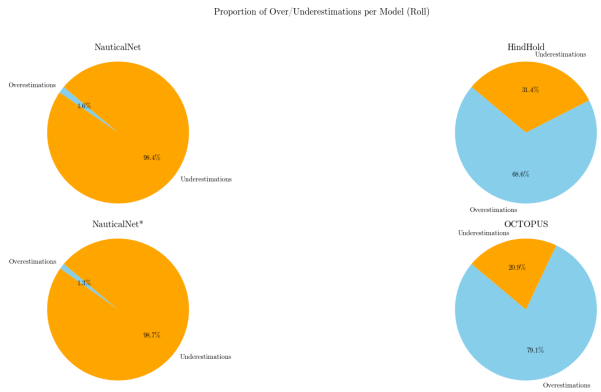
Figure 113: Pitch Predictions vs. Observed Pitch Motions: Time Series 169 to 176

Figure 114: Pitch Predictions vs. Observed Pitch Motions: Time Series 177 to 184

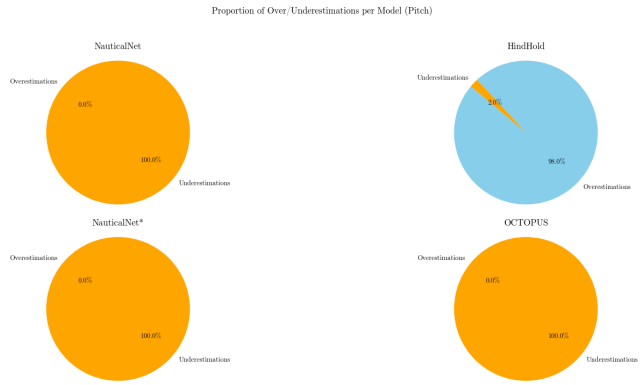
Figure 115: Pitch Predictions vs. Observed Pitch Motions: Time Series 185 to 191

E. Additional Research Case Plots

E.1 Pie Charts



(a)



(b)

Figure 116: Over- and Underestimation Characteristics Maximum Roll and Pitch Motion while excluding loading condition 7 from training

E.2 RMSE Iterative Approach

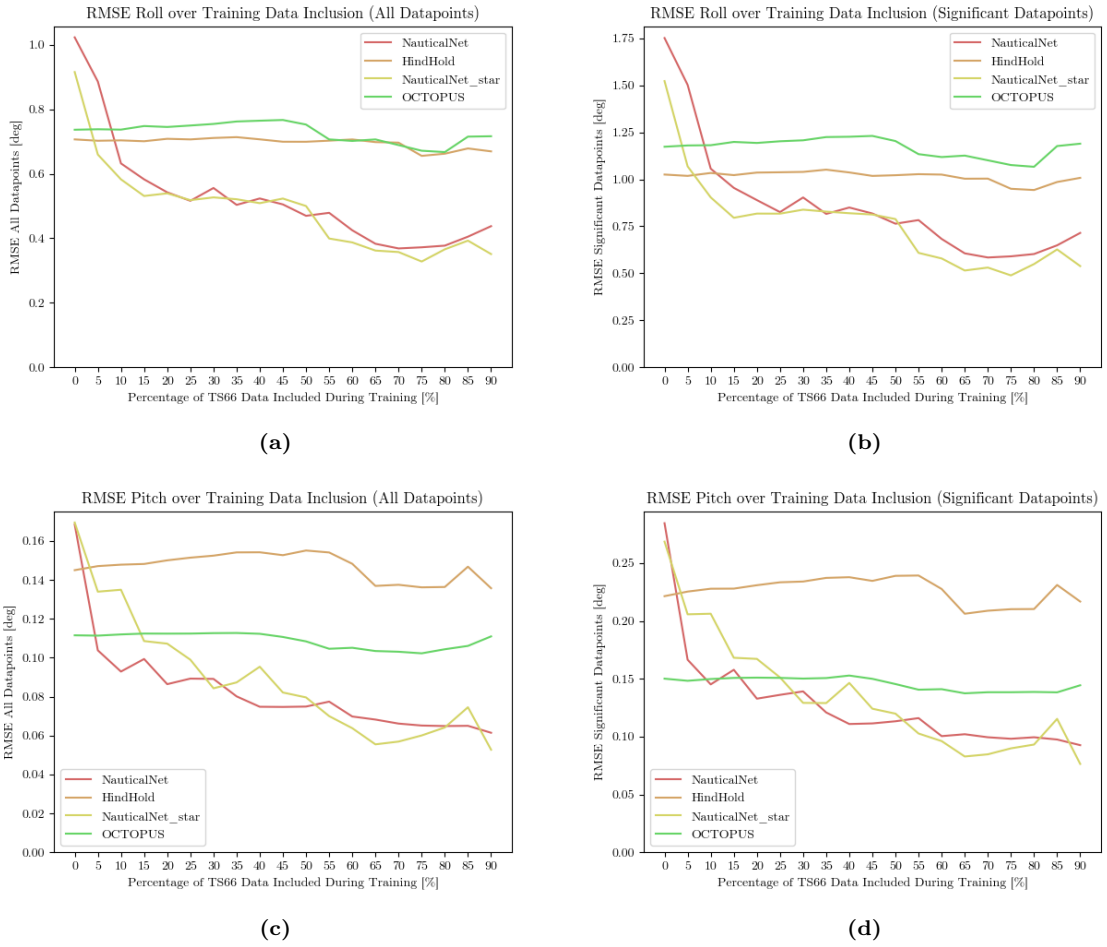


Figure 117: RMSE over % of data included during training and validation of NauticalNet and NauticalNet*

E.3 Prediction Plots Iterative Approach

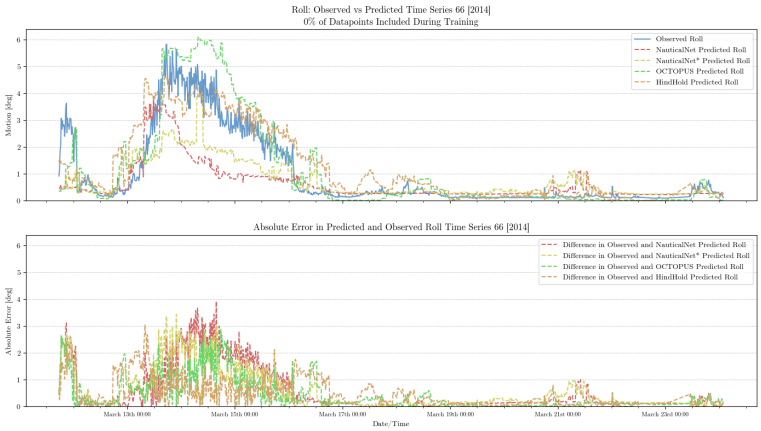


Figure 118: Roll Predictions Additional Research Case [0% of Data Included]

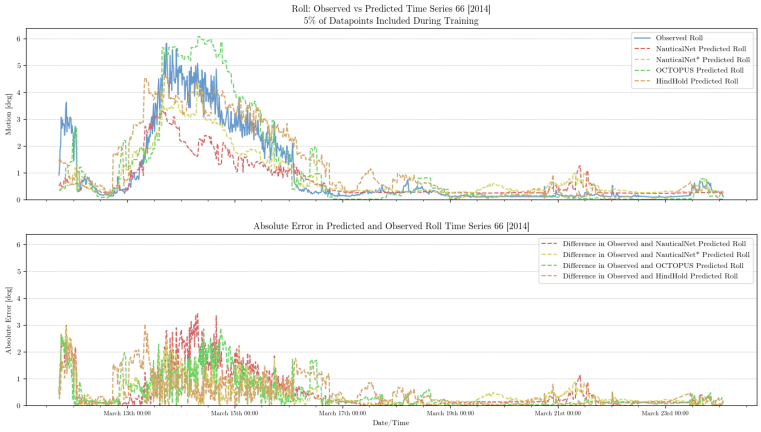


Figure 119: Roll Predictions Additional Research Case [5% of Data Included]

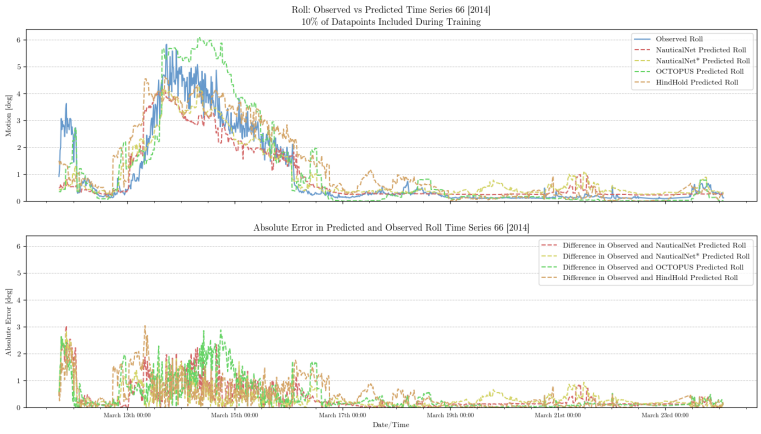


Figure 120: Roll Predictions Additional Research Case [10% of Data Included]

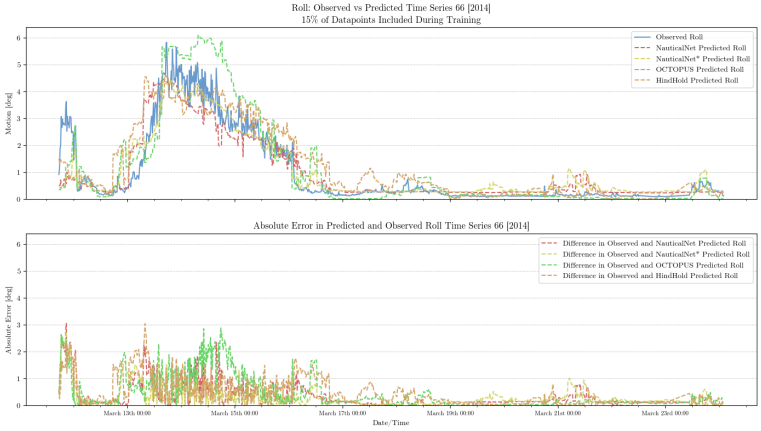


Figure 121: Roll Predictions Additional Research Case [15% of Data Included]

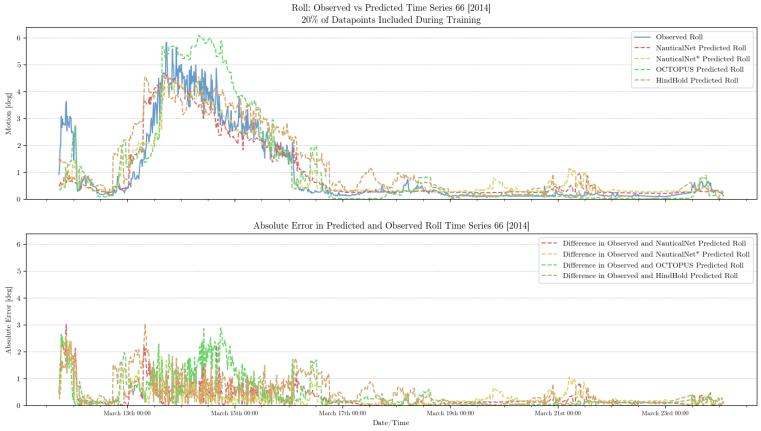


Figure 122: Roll Predictions Additional Research Case [20% of Data Included]

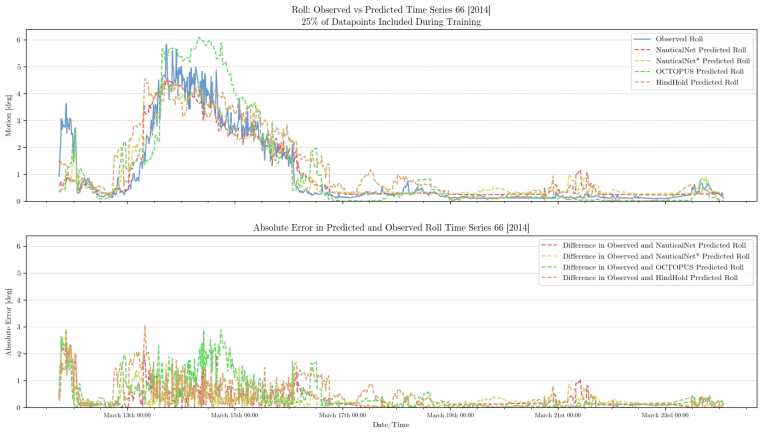


Figure 123: Roll Predictions Additional Research Case [25% of Data Included]

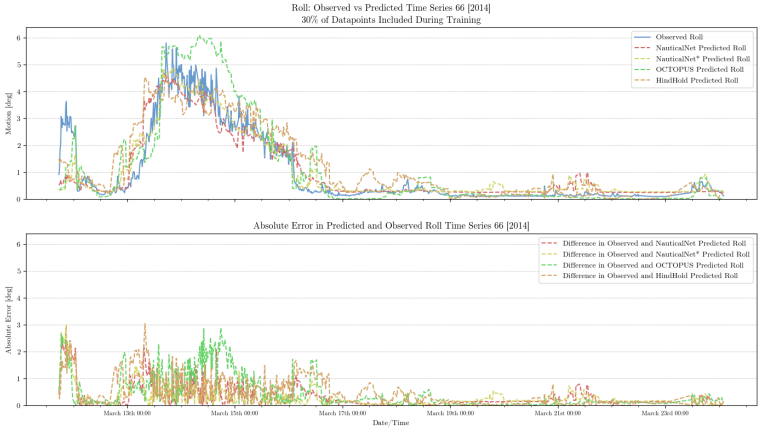


Figure 124: Roll Predictions Additional Research Case [30% of Data Included]

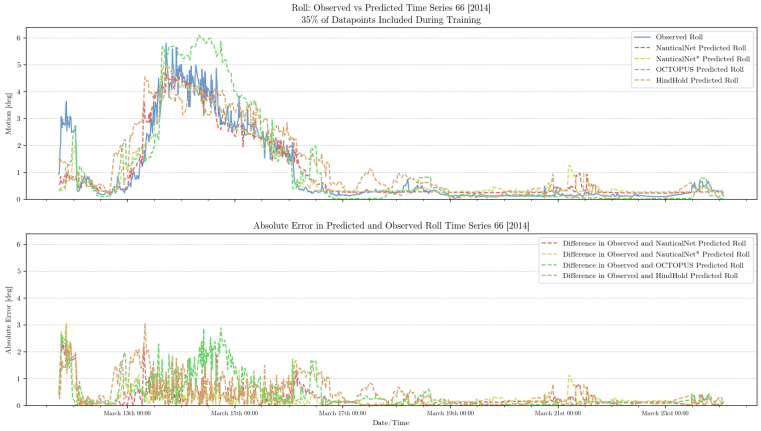


Figure 125: Roll Predictions Additional Research Case [35% of Data Included]

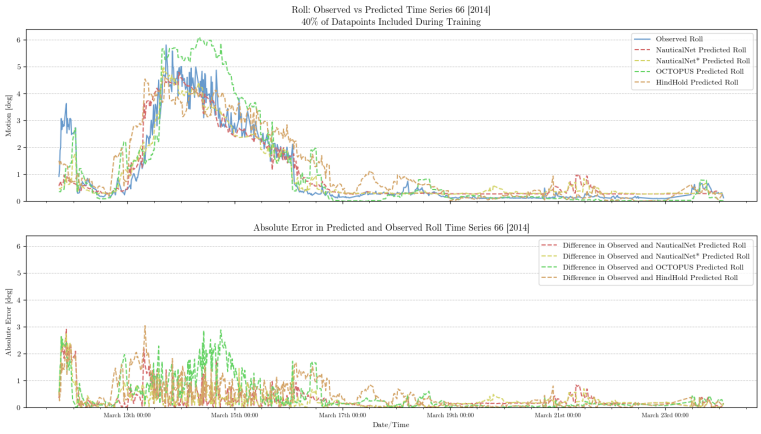


Figure 126: Roll Predictions Additional Research Case [40% of Data Included]

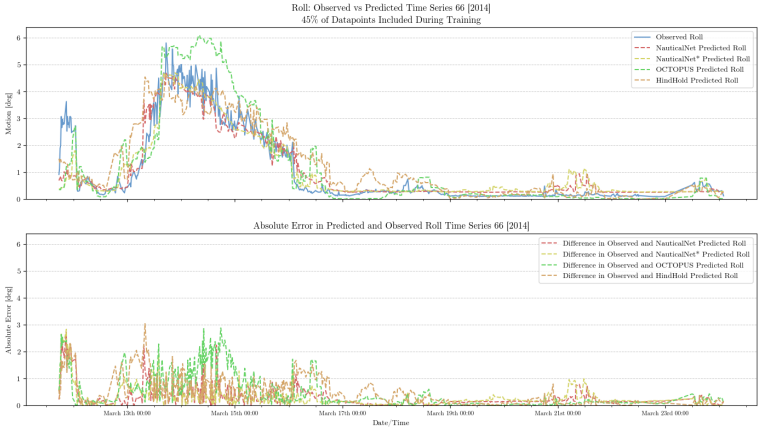


Figure 127: Roll Predictions Additional Research Case [45% of Data Included]

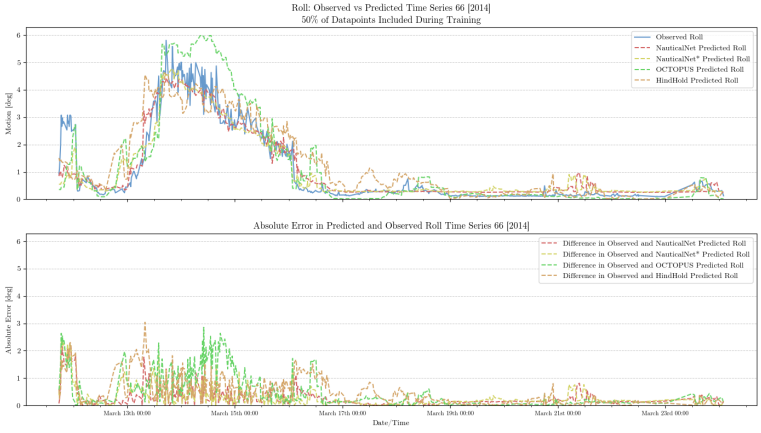


Figure 128: Roll Predictions Additional Research Case [50% of Data Included]

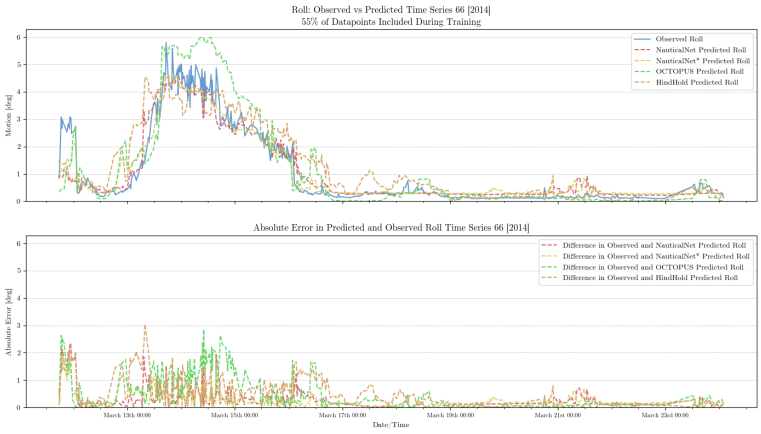


Figure 129: Roll Predictions Additional Research Case [55% of Data Included]

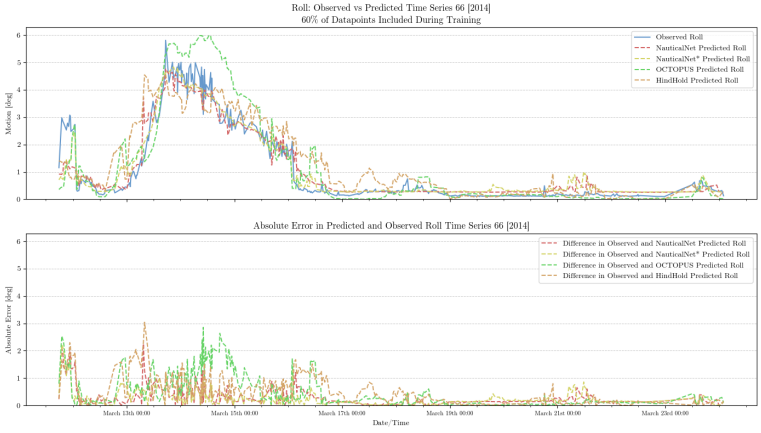


Figure 130: Roll Predictions Additional Research Case [60% of Data Included]

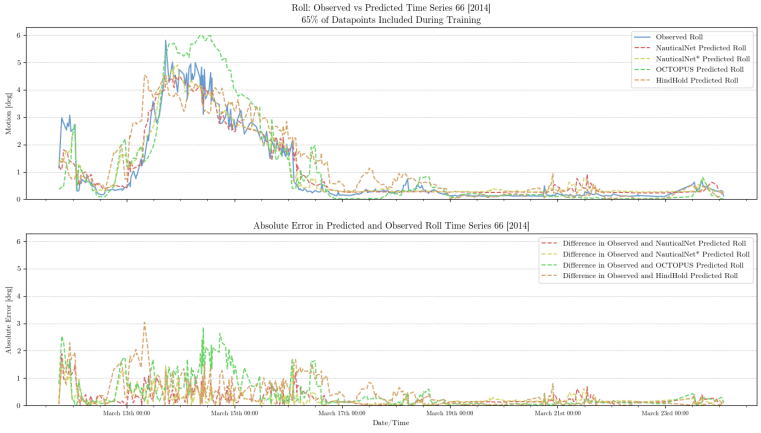


Figure 131: Roll Predictions Additional Research Case [65% of Data Included]

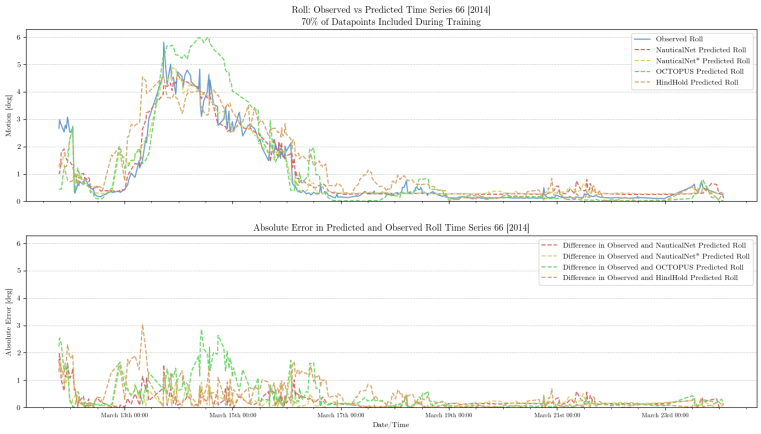


Figure 132: Roll Predictions Additional Research Case [70% of Data Included]

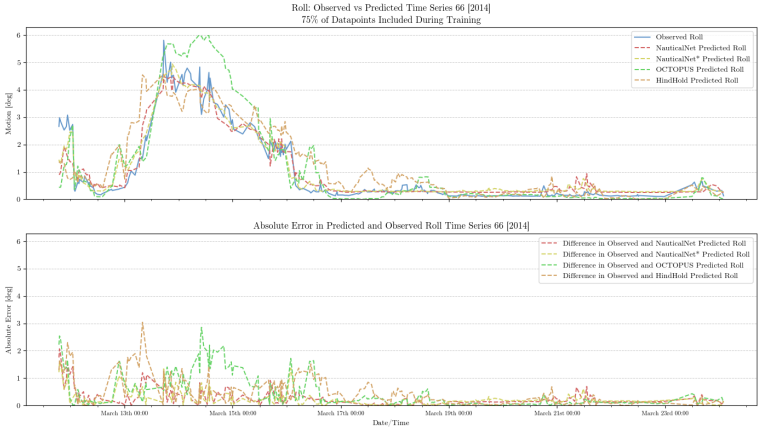


Figure 133: Roll Predictions Additional Research Case [75% of Data Included]

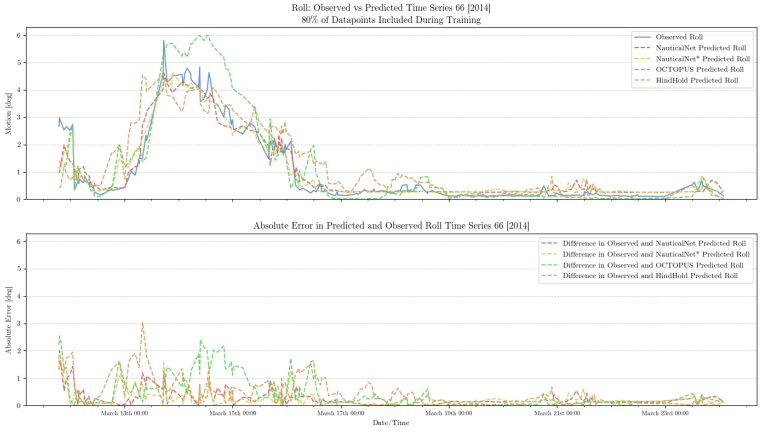


Figure 134: Roll Predictions Additional Research Case [80% of Data Included]

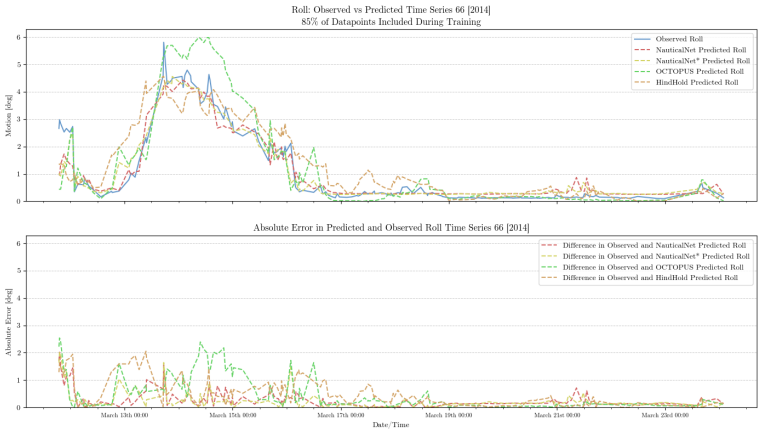


Figure 135: Roll Predictions Additional Research Case [85% of Data Included]

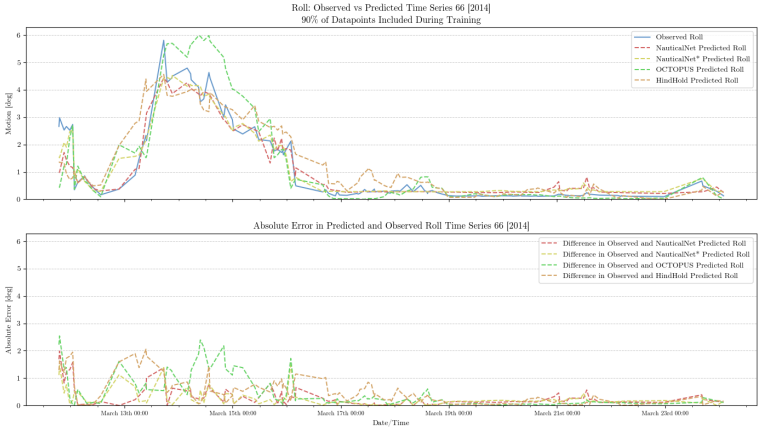


Figure 136: Roll Predictions Additional Research Case [90% of Data Included]

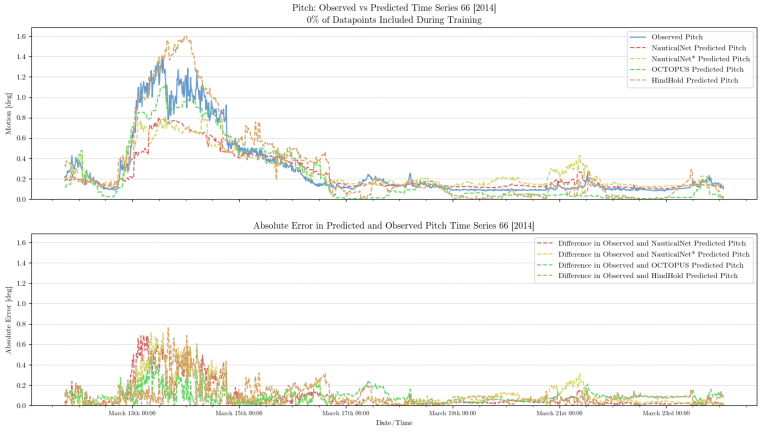


Figure 137: Pitch Predictions Additional Research Case [0% of Data Included]

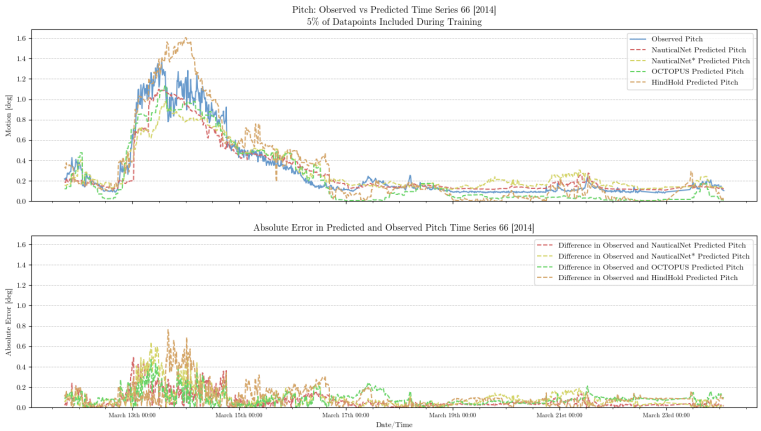


Figure 138: Pitch Predictions Additional Research Case [5% of Data Included]

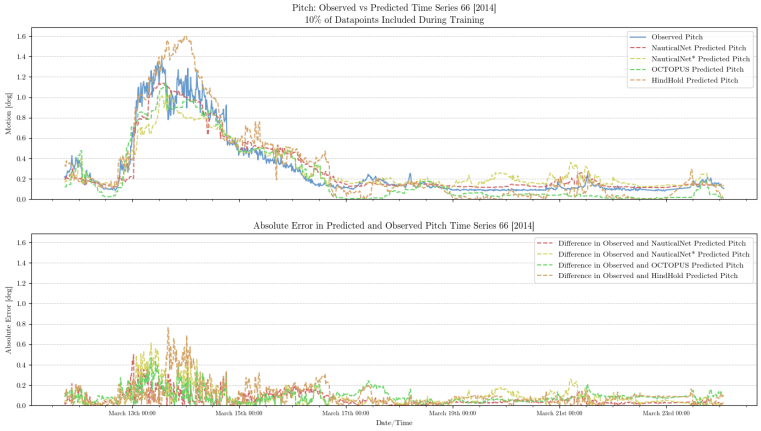


Figure 139: Pitch Predictions Additional Research Case [10% of Data Included]

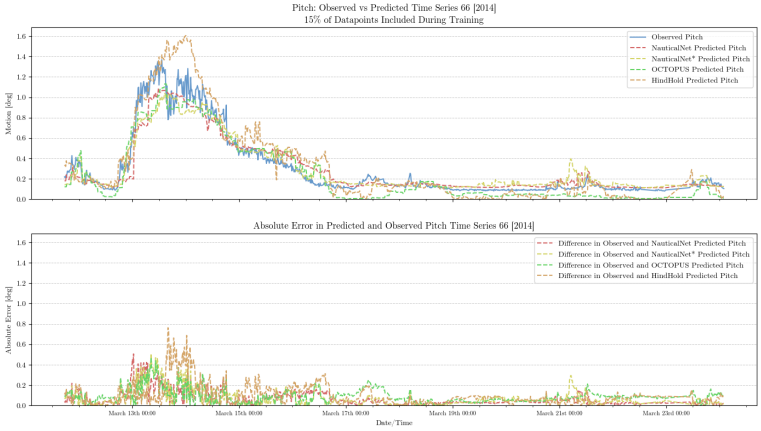


Figure 140: Pitch Predictions Additional Research Case [15% of Data Included]

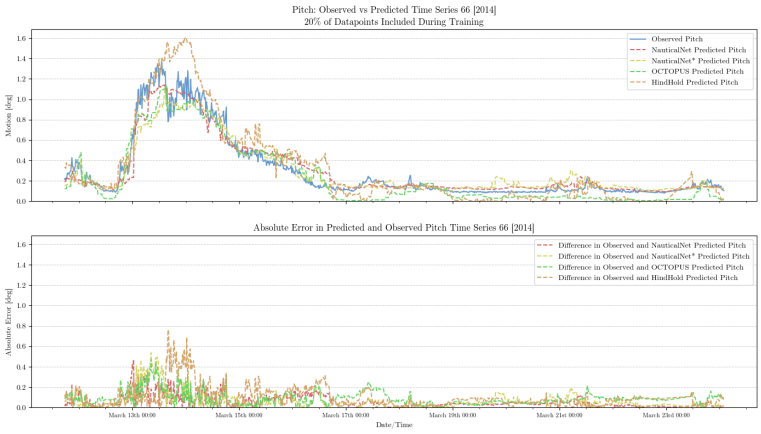


Figure 141: Pitch Predictions Additional Research Case [20% of Data Included]

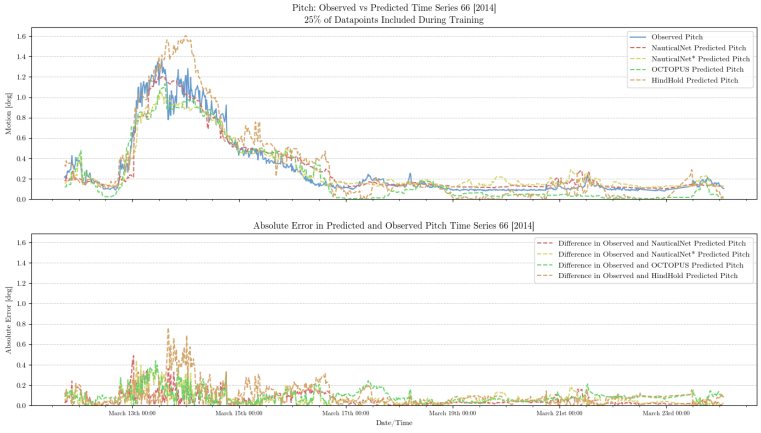


Figure 142: Pitch Predictions Additional Research Case [25% of Data Included]

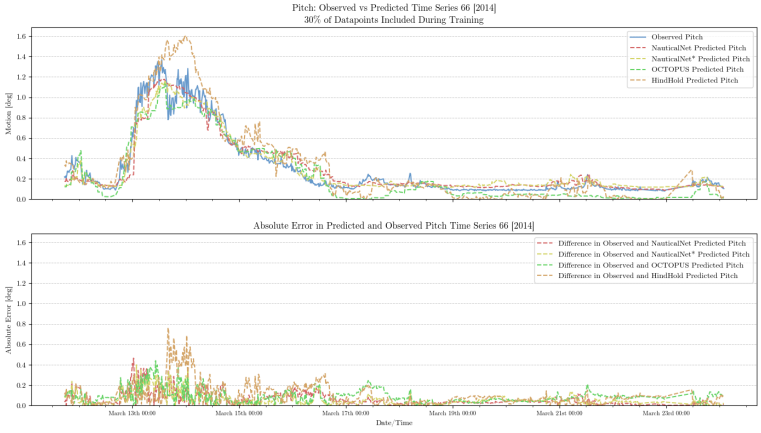


Figure 143: Pitch Predictions Additional Research Case [30% of Data Included]

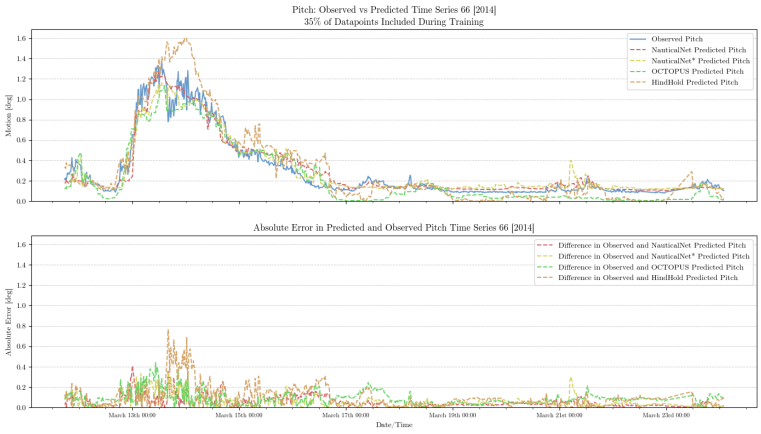


Figure 144: Pitch Predictions Additional Research Case [35% of Data Included]

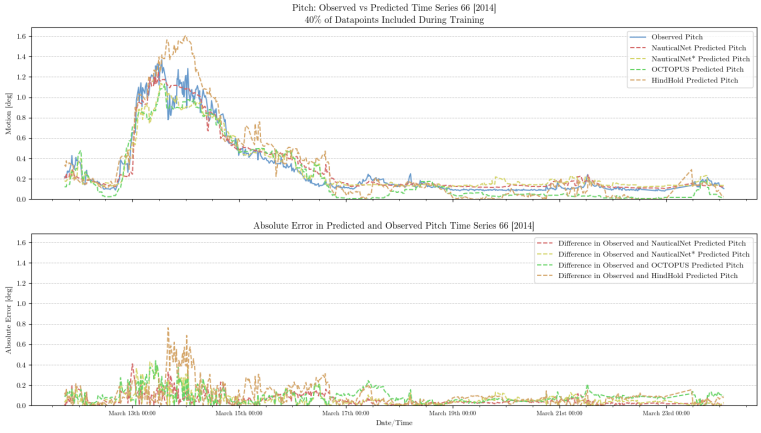


Figure 145: Pitch Predictions Additional Research Case [40% of Data Included]

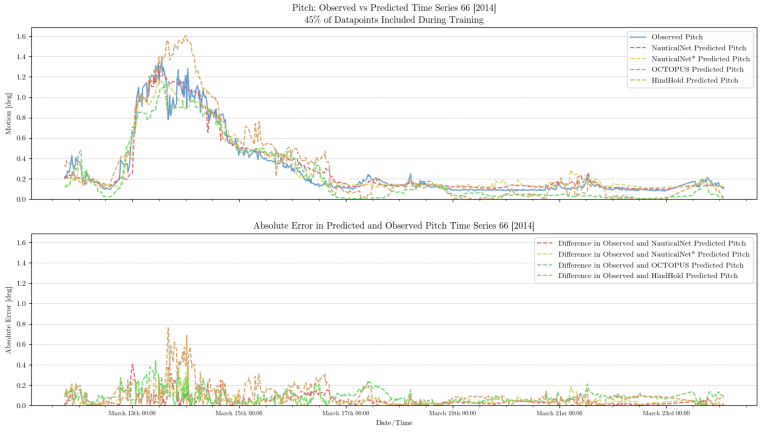


Figure 146: Pitch Predictions Additional Research Case [45% of Data Included]

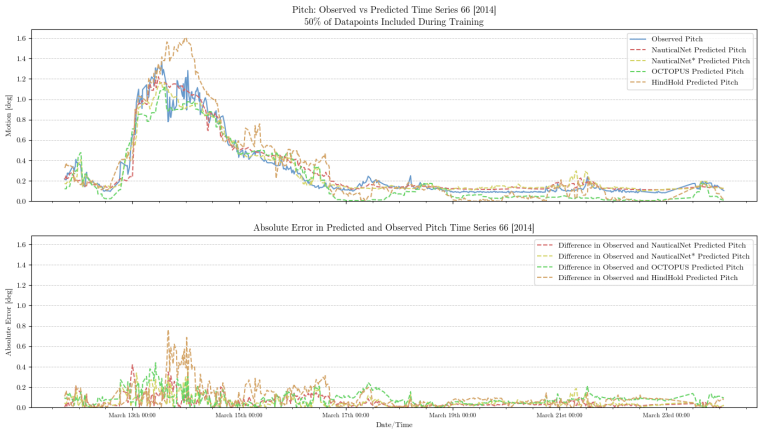


Figure 147: Pitch Predictions Additional Research Case [50% of Data Included]

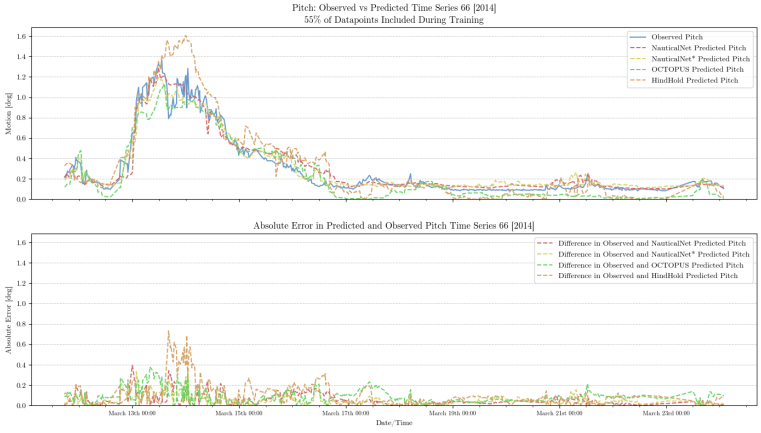


Figure 148: Pitch Predictions Additional Research Case [55% of Data Included]

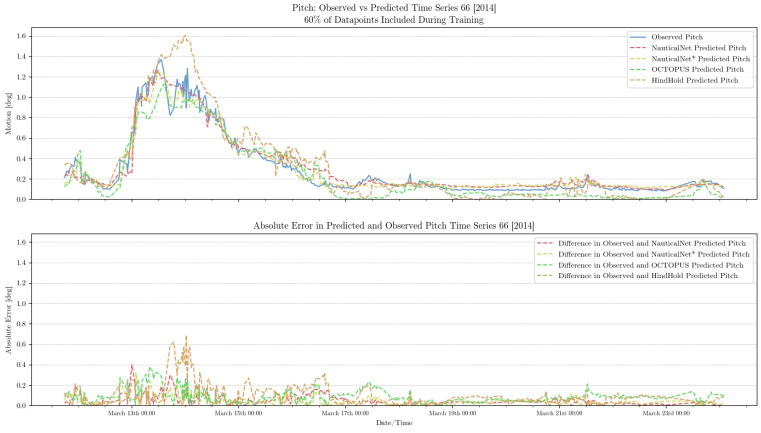


Figure 149: Pitch Predictions Additional Research Case [60% of Data Included]

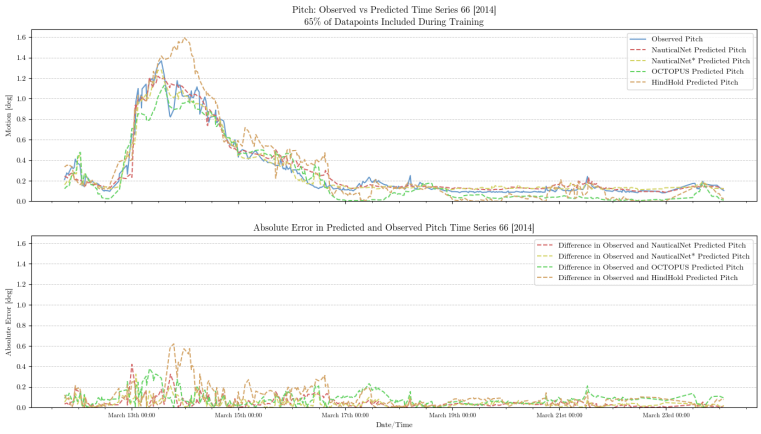


Figure 150: Pitch Predictions Additional Research Case [65% of Data Included]

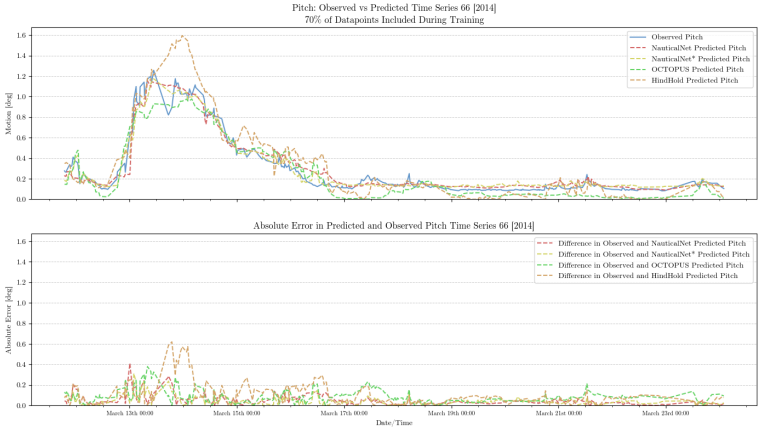


Figure 151: Pitch Predictions Additional Research Case [70% of Data Included]

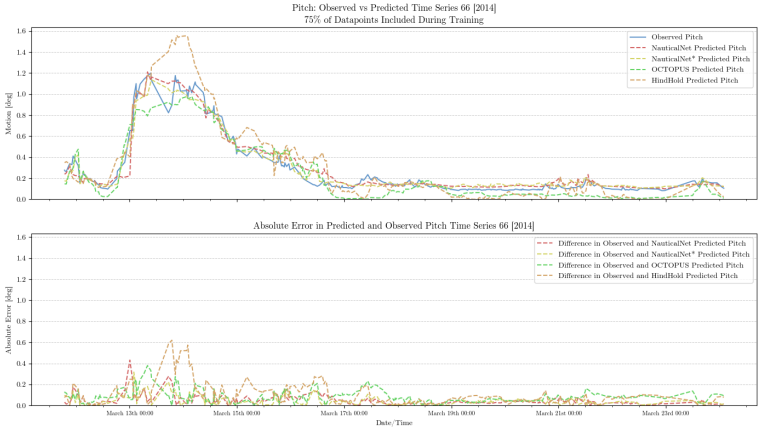


Figure 152: Pitch Predictions Additional Research Case [75% of Data Included]

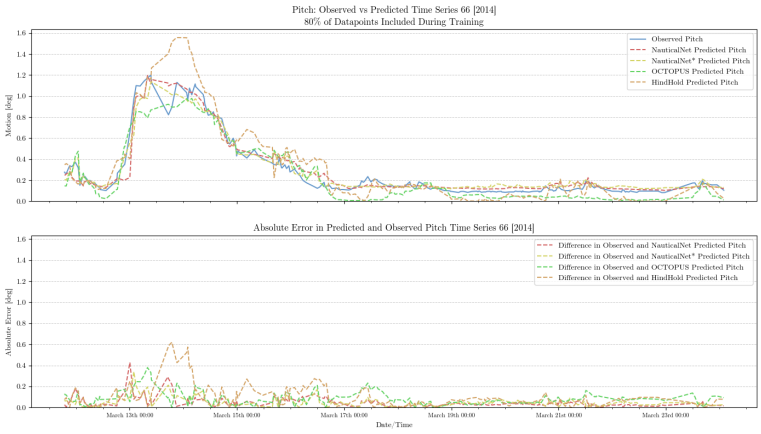


Figure 153: Pitch Predictions Additional Research Case [80% of Data Included]

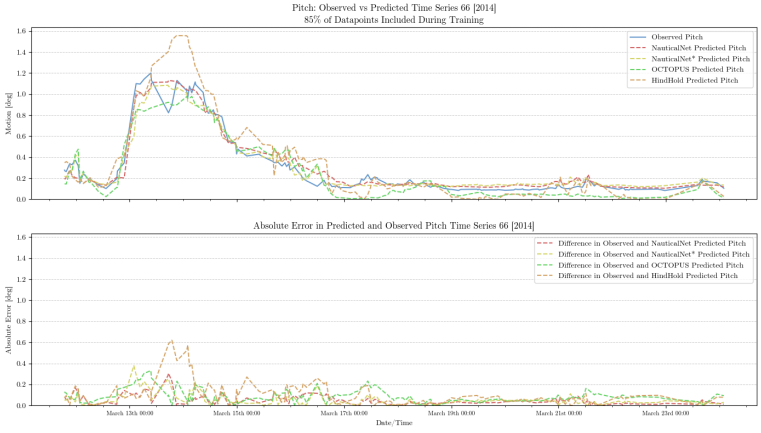


Figure 154: Pitch Predictions Additional Research Case [85% of Data Included]

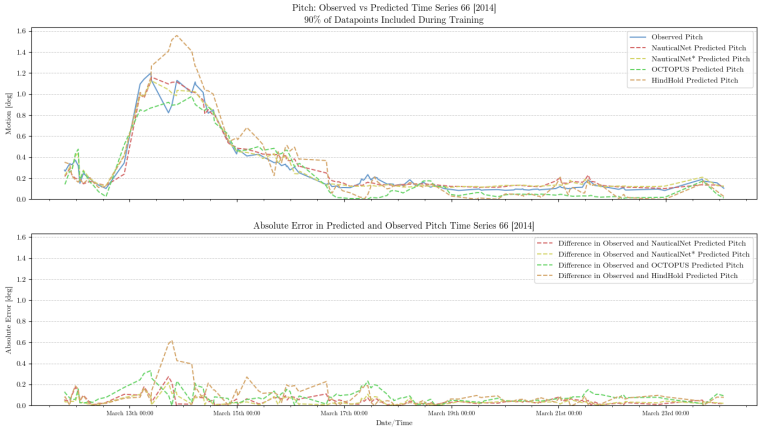


Figure 155: Pitch Predictions Additional Research Case [90% of Data Included]

F. Feature Importance

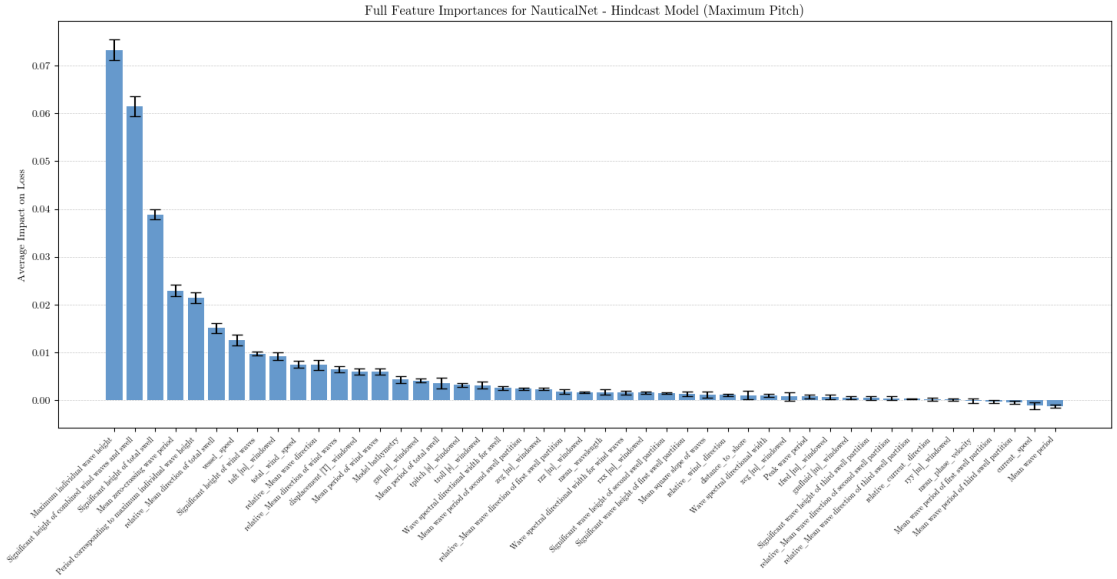


Figure 156: Feature Importance NauticalNet Maximum Pitch Motion

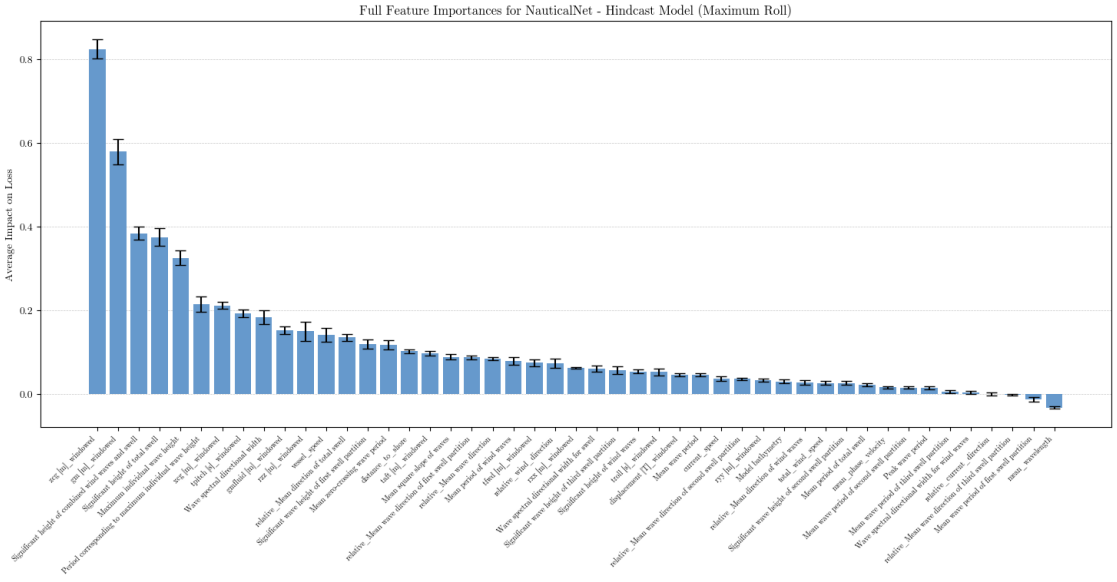


Figure 157: Feature Importance NauticalNet Maximum Roll Motion

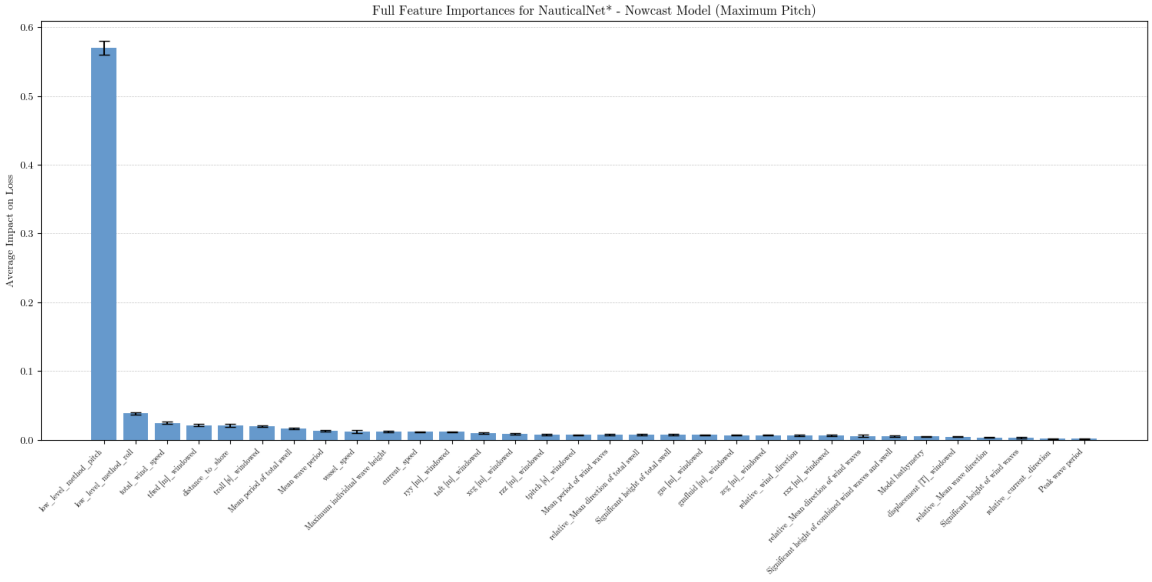


Figure 158: Feature Importance NauticalNet* Maximum Pitch Motion

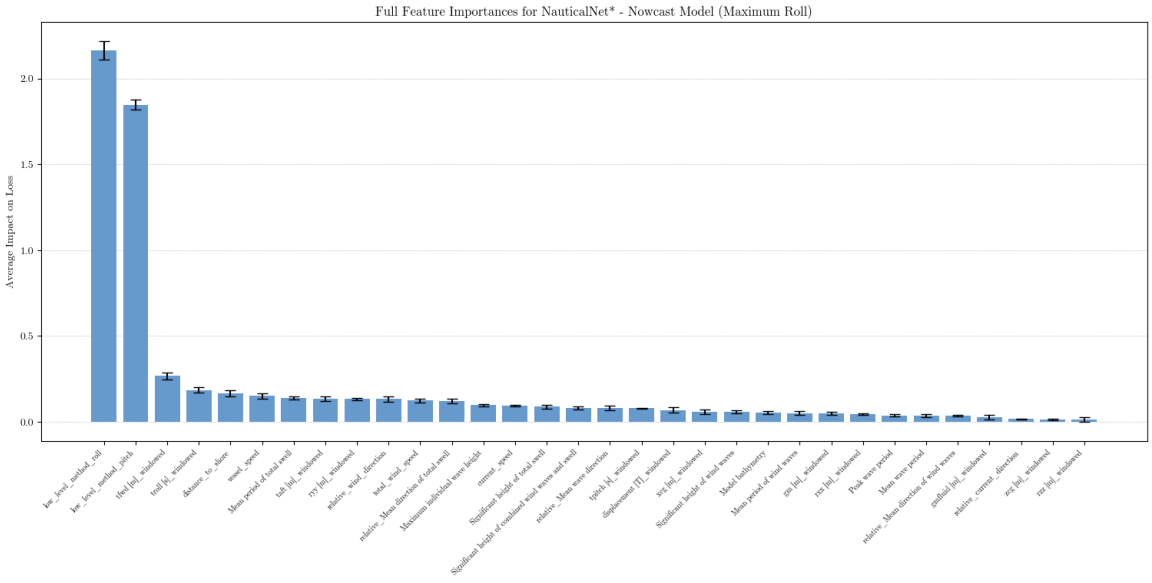


Figure 159: Feature Importance NauticalNet* Maximum Roll Motion

G. Additional Suggestions for Refining ML-Based Prediction Approaches

- **Utilization of Finer Environmental Data Grids for NauticalNet**: The resolution of environmental data can be improved by utilizing finer grids. For example, Copernicus offers more precise grids for certain areas. This could result in higher data quality and improved predictive capabilities.
- **Incorporation of Conventional Prediction Approach Predictions into NauticalNet**: At present, only NauticalNet* makes use of OCTOPUS' predictions in predict vessel motions. It was observed that the prediction accuracy of NauticalNet* was enhanced by incorporating these predictions. Nevertheless, conventional approach predictions were not included in NauticalNet. For practical purposes, it could be included in the input features to further improve the predictions. However, for academic and research purposes, the comparison between the approaches would become more complicated, as discussed in the previous section. By including these predictions, it is almost certain that ML-based approaches will outperform conventional approaches, thus the aim of this research, which was to directly compare the approaches, is no longer achievable.
- **Re-evaluation of Feature Selection and Data Cleaning**: At present, although feature selection and data cleaning were carefully considered, as discussed in Chapter 3, these processes usually require several attempts before the data set is finalized. In this research, the initial selection and procedure were not changed. ML-based approaches could be enhanced by optimizing these factors. In Section 6.1.3, the least influential features are discussed. One could begin with these. Additionally, other, new features could be explored, or techniques such as Principal Component Analysis could be examined.
- **Dedicated Optimization for NauticalNet***: The ML-based approach of NauticalNet was the focus of the current methodology. After evaluating the unequal comparison of hindcast and nowcast-based strategies, NauticalNet* gained attention. A potential way to improve the short-term performance would be to customize the architecture and hyperparameters of NauticalNet* independently, which could lead to improved results for NauticalNet*.
- **Implementation of Extensive Grid Searches**: At present, the architecture of NauticalNet has only been identified through small grid searches. It would be preferable to also use grid

searches to find the hyperparameters, which could lead to a more effective training setup. This should be done for NauticalNet, NauticalNet*, and HindHold. If more computing power is available, it would be beneficial to conduct more extensive grid searches for both the architecture and hyperparameters of all three models.

- **Enhancement of Task-Specific Learning**: It is possible to optimize task-specific learning even further. For instance, one task could be given more priority than the other (the more complex one over the simpler one). In the current system, both tasks are considered equally important, but in reality, the maximum roll motion may be more significant, so more emphasis could be placed on it. Additionally, individual loss functions could be designed for each task. This would guarantee tailored learning for both roll and pitch motions.

- **Adjustment Validation Strategy I**: If one still wishes to employ the validation strategy, they could opt to randomly select 10% of the data for testing instead of using the last 10% of each time series. This would create a more representative set of the entire time series, as opposed to the test set being heavily biased towards the end of the time series and the voyages.

- **Integration of Real-Time Data**: The potential usefulness of ML-based approaches could be increased if they are able to be re-trained during transport operations. The second validation strategy showed that the ML-based approach could adapt quickly after being exposed to data from a new loading condition. It would be beneficial if the model could be automatically updated during a voyage.

Bibliography

Abiodun, O. I., Jantan, A., Omolara, A. E., Dada, K. V., Mohamed, N. A., & Arshad, H. (2018). State-of-the-art in artificial neural network applications: A survey. *Heliyon*, *4*(11), e00938. https://doi.org/https://doi.org/10.1016/j.heliyon.2018.e00938

Alwosheel, A., van Cranenburgh, S., & Chorus, C. G. (2018). Is your dataset big enough? sample size requirements when using artificial neural networks for discrete choice analysis. *Journal of Choice Modelling*, *28*, 167–182. https://doi.org/10.1016/j.jocm.2018.07.002

Bishop, C. M. (2016). *Pattern recognition and machine learning*. Springer.

Black, I. M., Richmond, M., & Kolios, A. (2021). Condition monitoring systems: A systematic literature review on machine-learning methods improving offshore-wind turbine operational management. *International Journal of Sustainable Energy*, *40*(10), 923–946. https://doi.org/10.1080/14786451.2021.1890736

Bzdok, D., Altman, N., & Krzywinski, M. (2018). Statistics versus machine learning. *Nature News*. https://www.nature.com/articles/nmeth.4642

Chakrabarti, S. (2001). Empirical calculation of roll damping for ships and barges. *Ocean Engineering*, *28*, 915–932. https://doi.org/10.1016/S0029-8018(00)00036-6

Chollet, F., et al. (2015). Keras.

Dong, G., & Liu, H. (2018). *Feature engineering for machine learning and data analytics*. CRC press.

Feng, Z., Hu, P., Li, S., & Mo, D. (2022). Prediction of significant wave height in offshore china based on the machine learning method. *Journal of Marine Science and Engineering*, *10*(6). https://doi.org/10.3390/jmse10060836

Fernando, K. R. M., & Tsokos, C. P. (2022). Dynamically weighted balanced loss: Class imbalanced learning and confidence calibration of deep neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, *33*(7), 2940–2951. https://doi.org/10.1109/TNNLS.2020.3047335

Foldesi, L., & Valdenegro-Toro, M. (2022). Comparison of uncertainty quantification with deep learning in time series regression.

Hasselmann, K., Barnett, T., Bouws, E., Carlson, H., Cartwright, D., Enke, K., Ewing, J., Gienapp, H., Hasselmann, D., Kruseman, P., & et al. (1973). Measurements of wind-wave growth and swell decay during the joint north sea wave project (jonswap). *Ergänzungsheft 8-12*. https://repository.tudelft.nl/islandora/object/uuid%5C%3Af204e188-13b9-49d8-a6dc-4fb7c20562fc

Héas, P., Cérou, F., & Rousset, M. (2023). Chilled sampling for uncertainty quantification: A motivation from a meteorological inverse problem.

Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., Nicolas, J., Peubey, C., Radu, R., Rozum, I., Schepers, D., Simmons, A., Soci, C., Dee, D., & Thépaut, J.-N. (2023). ERA5 hourly data on single levels from 1940 to present. https://doi.org/10.24381/cds.adbb2d47

Hersbach, H., Pell, B., & Berrisford, P. (2020). The era5 global reanalysis. https://rmets.onlinelibrary.wiley.com/doi/full/10.1002/qj.3803

Ibrahim, R., & Grace, I. (2010). Modeling of ship roll dynamics and its coupling with heave and pitch. *Mathematical Problems in Engineering*, *2010*. https://doi.org/10.1155/2010/934714

Ikeda, Y., Himeno, Y., & Tanaka, N. (1978). Components of roll damping of ship at forward speed. *Journal of the Society of Naval Architects of Japan*, *143*, 113–125.

Ilyas, I. F., & Chu, X. (2019). *Data cleaning*. Association for Computing Machinery.

Jiang, Y., Hou, X.-R., Wang, X.-G., Wang, Z.-H., Yang, Z.-L., & Zou, Z.-J. (2021, May). Identification modeling and prediction of ship maneuvering motion based on lstm deep neural network - journal of marine science and technology. https://link.springer.com/article/10.1007/s00773-021-00819-9

J.M.J. Journée, L. A. (2012). Octopus technical note (1st ed.).

Journée, J., & Massie, W. (2008). *Offshore hydromechanics*. Delft University of Technology.

Karola, A., Hirdaris, S., Matusiak, J., & Mikkola, T. (2022, June). *Wave loads and ship motions evaluated by linear strip and panel methods* (Vol. Volume 5A: Ocean Engineering). https://doi.org/10.1115/OMAE2022-79702

Khan, A., Bil, C., & Marion, K. E. (2005). Theory and application of artificial neural networks for the real time prediction of ship motion (R. Khosla, R. J. Howlett, & L. C. Jain, Eds.), 1064–1069.

Kotsiantis, S. B., Kanellopoulos, D., & Pintelas, P. (2006, January). Data preprocessing for supervised learning. https://www.researchgate.net/profile/P-Pintelas/publication/228084519_Data_Preprocessing_for_Supervised_Learning/links/0c960517fefa258d0d000000/Data-Preprocessing-for-Supervised-Learning.pdf

Li, G., Kawan, B., Wang, H., & Zhang, H. (2017). Neural-network-based modelling and analysis for time series prediction of ship motion. *Ship Technology Research*, *64*(1), 30–39. https://doi.org/10.1080/09377255.2017.1309786

Liu, Y., Cai, L., Chen, Y., & Wang, B. (2022). Physics-informed neural networks based on adaptive weighted loss functions for hamilton-jacobi equations. *Mathematical Biosciences and Engineering*, *19*(12), 12866–12896. https://doi.org/10.3934/mbe.2022601

Masud, M., Alhamid, M. F., & Zhang, Y. (2022). A convolutional neural network model using weighted loss function to detect diabetic retinopathy. *ACM Trans. Multimedia Comput. Commun. Appl.*, *18*(1s). https://doi.org/10.1145/3470976

McTaggart, K. (2007). Shipmo3d version 1.0 user manual for frequency domain analysis of ship seakeeping in a seaway. https://apps.dtic.mil/sti/citations/AD1005169

McTaggart, K. (2021). Shipmo3d version 4.3 user manual for evaluating ship motions in the time and frequency domains. https://cradpdf.drdc-rddc.gc.ca/PDFS/unc367/p813574_A1b.pdf

McTaggart, K., Datta, I., Stirling, A., Gibson, S., & Glen, I. (1997). An updated strip theory program for predicting ship motions and sea loads in waves. https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=9cab1c06893f3dff64d2847bf361757bc52b0226

McTaggart, K. A. (2011). Verification and validation of shipmo3d ship motion predictions in the time and frequency domains. *International Journal of Naval Architecture and Ocean Engineering*, *3*(1), 86–94. https://doi.org/https://doi.org/10.2478/IJNAOE-2013-0049

MO4. (2018). Mo4 technical note (1st ed.).

Moreira, M., & Fiesler, E. (1995). Neural networks with adaptive learning rate and momentum terms [URL: http://publications.idiap.ch/downloads/reports/1995/95-04.pdf]. *Neural Networks*.

Nayfeh, A. H., Mook, D. T., & Marshall, L. R. (1973). Nonlinear coupling of pitch and roll modes in ship motions. *ARC*. https://arc.aiaa.org/doi/10.2514/3.62949

Nielsen, M. A. (2015, January). *Neural networks and deep learning*. Determination Press. http://neuralnetworksanddeeplearning.com/

Oktavian, M. W., Yudistira, N., & Ridok, A. (2022). Classification of alzheimer's disease using the convolutional neural network (cnn) with transfer learning and weighted loss.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., ... Chintala, S. (2019). Pytorch: An imperative style, high-performance deep learning library. In *Advances in neural information processing systems 32* (pp. 8024–8035). http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Rebala, G., Ravi, A., & Churiwala, S. (2019). *An introduction to machine learning*. Springer International Publishing.

Ribadeneira, B., Benalcázar, S., & Guachamin-Acero, W. (2022). Feasibility study of flopper stoppers as a passive roll compensation system for offshore construction vessels. *Ocean Engineering*, *264*, 112380. https://doi.org/https://doi.org/10.1016/j.oceaneng.2022.112380

Salvesen, N., Tuck, E. O., & Faltinsen, O. M. (1970). Ship motions and sea loads. https://api.semanticscholar.org/CorpusID:109444905

Sayed, M., & Hamed, Y. S. (2010). Stability and response of a nonlinear coupled pitch-roll ship model under parametric and harmonic excitations - nonlinear dynamics. *SpringerLink*. https://link.springer.com/article/10.1007/s11071-010-9841-0#citeas

Seyffert, H. (2022). Lecture 1: Introduction, hydrostatics, & buoyancy. *TU Delft Lecture - Introduction to Ship and Offshore Hydromechanics (MT44045)*. https://brightspace.tudelft.nl/d2l/le/content/500870/viewContent/2718364/View

Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data*, *6*(1), 60. https://doi.org/10.1186/s40537-019-0197-0

Silva, K. M., & Maki, K. J. (2022). Data-driven system identification of 6-dof ship motion in waves with neural networks. *Applied Ocean Research*, *125*, 103222. https://doi.org/https://doi.org/10.1016/j.apor.2022.103222

Skulstad, R., Li, G., Fossen, T. I., Vik, B., & Zhang, H. (2021). A hybrid approach to motion prediction for ship docking—integration of a neural network model into the ship dynamic model. *IEEE Transactions on Instrumentation and Measurement*, *70*, 1–11. https://doi.org/10.1109/TIM.2020.3018568

Song, Y., Wang, T., Mondal, S. K., & Sahoo, J. P. (2022). A comprehensive survey of few-shot learning: Evolution, applications, challenges, and opportunities.

Tan, C., Sun, F., Kong, T., Zhang, W., Yang, C., & Liu, C. (2018). A survey on deep transfer learning. In V. Kůrková, Y. Manolopoulos, B. Hammer, L. Iliadis, & I. Maglogiannis (Eds.), *Artificial neural networks and machine learning – icann 2018* (pp. 270–279). Springer International Publishing.

thung, C., & Huang, Z. (1991). Comparison of the strip theory and the panel method in computing ship motion with forward speed. https://repository.tudelft.nl/islandora/object/uuid:d68a19a0-b790-4102-b103-6b0ea0a1d22c/datastream/OBJ/download

Thung, K.-H., & Wee, C.-Y. (2018). A brief review on multi-task learning. *Multimedia Tools and Applications*, *77*, 29705–29725. https://doi.org/10.1007/s11042-018-6463-x

Vuttipittayamongkol, P., Tung, A., & Elyan, E. (2021). Towards machine learning-driven practices for oil and gas decommissioning – introduction of a new offshore pipeline dataset, 111–116. https://doi.org/10.1145/3479162.3479179

Willard, J., Jia, X., Xu, S., Steinbach, M., & Kumar, V. (2022). Integrating scientific knowledge with machine learning for engineering and environmental systems. *arXiv.org*. https://arxiv.org/abs/2003.04919

Xu, Y., & Goodacre, R. (2018). On splitting training and validation set: A comparative study of cross-validation, bootstrap and systematic sampling for estimating the generalization performance of supervised learning. *Journal of Analysis and Testing*, *2*(3), 249–262. https://doi.org/10.1007/s41664-018-0068-2

Yeter, B., Garbatov, Y., & Soares, C. G. (2023). Review on artificial intelligence-aided life extension assessment of offshore wind support structures - journal of marine science and application. *SpringerLink*. https://link.springer.com/article/10.1007/s11804-022-00298-3