



Sea state identification using machine learning

A comparative study based on in-service data from a container vessel

Mittendorf, Malte; Nielsen, Ulrik Dam; Bingham, Harry B.; Storhaug, Gaute

Published in:
Marine Structures

Link to article, DOI:
[10.1016/j.marstruc.2022.103274](https://doi.org/10.1016/j.marstruc.2022.103274)

Publication date:
2022

Document Version
Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

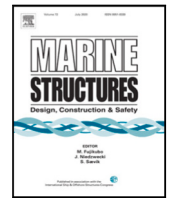
Citation (APA):
Mittendorf, M., Nielsen, U. D., Bingham, H. B., & Storhaug, G. (2022). Sea state identification using machine learning: A comparative study based on in-service data from a container vessel. *Marine Structures*, 85, Article 103274. <https://doi.org/10.1016/j.marstruc.2022.103274>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.



Sea state identification using machine learning—A comparative study based on in-service data from a container vessel

Malte Mittendorf^{a,*}, Ulrik Dam Nielsen^{a,c}, Harry B. Bingham^a, Gaute Storhaug^b

^a DTU Civil and Mechanical Engineering, Technical University of Denmark, DK-2800, Kgs. Lyngby, Denmark

^b Maritime Advisory, DNV AS, NO-1363, Høvik, Norway

^c Centre for Autonomous Marine Operations and Systems, NTNU AMOS, NO-7052, Trondheim, Norway

ARTICLE INFO

Keywords:

Sea state estimation
Wave buoy analogy
Sensor data
Wave radar
Deep learning
Multi-task learning

ABSTRACT

This paper is concerned with a machine learning-based approach for sea state estimation using the wave buoy analogy. In-situ sensor data of an advancing medium-size container vessel has been utilized for the prediction of integral sea state parameters. The main novelty of this contribution is the rigorous comparison of time and frequency domain models in terms of accuracy, robustness and computational cost. The frequency domain model is trained on sequences of spectral ordinates derived from cross response spectra, while the time domain model is applied to 5-minute time series of ship responses. Multiple deep neural networks were trained and the sensitivity of individual sensor recordings, sample length, and frequency discretization on estimation accuracy was analysed. An Inception Architecture adapted for sequential data yields the highest out of sample performance in both considered domains. Additionally, multi-task learning was employed, as it is known for increased generalization capability and diminished uncertainty. Overall, it was found that the frequency domain method provides both superior performance and significantly less computational effort for training.

1. Introduction

1.1. Motivation

Ship safety and operational efficiency depend to a large degree on the prevalent sea state. As such, on-board identification of the ambient wave system may assist the crew in the decision making processes for minimizing risks related to critical wave encounters. For instance, large roll amplitudes are presently an immense concern due to the increasing number of lost containers at sea over the last couple of years resulting from wave-related impact, Meister et al. [1]. Furthermore, estimates of the sea state experienced throughout the vessel's operational profile are of significant importance for shore-based vessel performance analysis for the assessment of general energy efficiency and scheduling maintenance. Therefore, an accurate and reliable estimate of the prevailing wave energy density spectrum is the key aspect of on-board decision support systems. In addition, with an ever-increasing interest in autonomous ships, the significance of real time estimates of the wave environment and the corresponding ship response grows further as the expertise of seafarers may not necessarily be available, Jalonen et al. [2]. For all of this, it is thought that ship response-based sea state estimation (SSE) could be an essential building block. Hereby, it is understood that data in terms of wave-induced responses from the ship is processed, thus facilitating a real-time identification of the sea state at the ship's exact position. In fact, this is the underlying idea of the wave buoy analogy, as presented by Nielsen [3]. Broadly speaking, the wave buoy

* Corresponding author.

E-mail address: mamit@mek.dtu.dk (M. Mittendorf).

analogy considers the ship as a wave rider buoy and establishes an inverse mathematical relationship between measured responses and the encountered directional wave spectrum or the corresponding integral parameters.

1.2. Literature review

The methods for ship-based sea state estimation are manifold and an overview is given in Nielsen [4]. Initial studies addressing the wave buoy analogy were carried out in the 1970s by e.g. Takekuma and Takahashi [5], but without considering forward speed, i.e. the Doppler shift. Iseki and Ohtsu [6] and Nielsen [7] present methodologies based on Bayes theorem for the calculation of directional wave spectra using both complex-valued transfer functions and cross response spectra under forward speed conditions. Following the convention of Nielsen [3], the techniques for ship-board SSE are split into non-parametric and parametric approaches: The former provides the directional or 2D wave spectrum, while parametric techniques by e.g. [8–10] yield input values to parameterized wave energy density spectra with the possible addition of spreading functions. The applicability and accuracy of the aforementioned estimation methods inherently depend on the availability of reliable transfer functions, also referred to by Response Amplitude Operators (RAO). Hence, Nielsen et al. [11] propose a correction or rather *calibration* methodology for the pitch RAO using met-ocean hindcast data and in-service ship motion recordings of a container vessel. Mounet et al. [12] extend this work and merge the correction technique into a sea state estimation approach using a network of ships as observation platforms. Nevertheless, in individual cases, transfer functions may simply be unknown to the ship operator due to a lack of detailed hull shape information. Moreover, the presented procedures are premised on the assumptions of linearity for which transfer functions can be applied. These potential disadvantages of techniques relying on RAOs subsequently motivate data-driven machine learning studies for the estimation of sea state conditions.

Nowadays, machine learning techniques are universally applied as so-called surrogate models in general ship hydrodynamics, i.e. the model approximates data of computationally expensive methods in a regression task, e.g. Mittendorf and Papanikolaou [13]. In the field of ship-based SSE, machine learning is increasingly applied as well: Åvist and Pyörre [14], for instance, apply traditional machine learning – non-parametric regression techniques in particular – for the prediction of the significant wave height and encountered wave direction based on frequency domain features. Furthermore, Han et al. [15] predict the sea state parametrically using a research vessel as a case study and compare three different machine learning methods. In addition, the data preprocessing part is another key aspect and the feature space comprises elements from multiple domains. More importantly, deep learning methods, i.e. artificial neural networks with more than one hidden layer, achieved significant results in a variety of tasks ranging from image to speech recognition. The increased accuracy, versatility and scalability compared to traditional machine learning methods is credited to special layer-types, such as convolutional or recurrent layers. For the theoretical intricacies of deep learning consult Goodfellow et al. [16]. Its rapid development over the last decade subsequently sparked several studies in the ship-based SSE domain. Cheng et al. [17] develop an end-to-end classification approach via a multi-channel convolutional network for sea state estimation using the Beaufort sea scale for dynamic positioning, i.e. without forward speed. Instead of using raw time series data, Cheng et al. [18] convert the ship motions to spectrograms and train a neural network in an image recognition task predicting the sea state scale. The work is extended for forward speed cases in Cheng et al. [19] and an advanced architecture is employed. Their so-called SSENET features attention mechanisms and residual skip connections for enhanced performance. Moreover, Düz et al. [20] present a real-time multivariate time series *regression* approach for integral sea state parameters applying multiple deep architectures on 2.5-min motion samples of a frigate-type ship. One distinct aspect of their work is the procedure of transfer learning: Initially, the model is trained on simulated data obtained from time domain potential theory calculations and then re-trained on in-situ measurement data. The displayed results show sufficient accuracy for the prediction of the sea state parameters under forward speed conditions. Kawai et al. [21] present a simulation-based study of a container carrier in the frequency domain by extracting sequences of spectral values from a set of cross response spectra. The convolutional neural network predicts the parameters for an Ochi–Hubble type spectrum, i.e. for both wind and swell waves. Scholcz and Mak [22] extend the work of Düz et al. [20] and present a deep learning methodology for the non-parametric estimation of the directional wave spectrum based on wave radar data using a convolutional encoder–decoder network applied to in-service time series data. Lastly, Han et al. [23] provide an investigation for non-parametric SSE and establish an approach based on a generative adversarial network, in which the generator predicts the 2D spectrum relying on cross response spectra, and the discriminator classifies the validity of the prediction. This iterative approach is then compared to a traditional model-based approach and shows satisfactory accuracy in non-forward speed scenarios; emphasizing that simulated time series data has been considered exclusively.

1.3. Objective

This paper focuses on deep learning methods exclusively, whereas model-based approaches, dependent on availability of RAOs, are not considered. In view of state of the art literature, it is concluded that the majority of deep learning studies use time domain data. Thus, the main novelty of this paper is the parallel application of deep neural networks in both the time and frequency domains. In addition, a large part of the literature's deep learning studies are based on simulated ship motion data. In the present paper, however, in-service sensor recordings from a container vessel sailing in the Northern Atlantic are studied. A regression approach is proposed using sequential data from the time and frequency domain. The inverse mapping from several ship responses to prevalent sea conditions, i.e. significant wave height, peak period and mean encounter wave direction, is achieved by advanced deep neural networks, such as residual networks. Additionally, a novel approach for obtaining the optimal combination of ship responses for sea state identification is demonstrated in a feature importance study. Moreover, the trade-off between sample length



Fig. 1. Case study in service taken from [24].

Table 1

Main particulars of the Panamax container vessel, cf. [24].

Dimension	Unit	Magnitude
L_{pp}	[m]	232.0
B	[m]	32.2
T	[m]	10.78
C_B	[-]	0.685
dwt	[t]	40,900
P_B (@ 93 rpm)	[kW]	25,786

and frequency discretization will be determined in the frequency domain in another sensitivity study. Initially, four multi-output-regressor architectures are compared in terms of prediction accuracy, and two different multi-task configurations will be applied to the best-performing model in the first iteration. The application of multi-task learning for sea state estimation is another novelty of this investigation, as the recent literature is only focused on multi-output estimators. Lastly, a contribution of this work is the development of data-driven methods for the mapping of in-situ ship motion recordings to sea state parameters derived from data of a directional wave radar.

1.4. Composition

The remaining sections of the article are organized in the following way: In the upcoming Section 2, the case study, its sensor infrastructure and the data filtering methodology will be presented. Section 3 has its focus on the proposed methodology and conveys the applied neural network architectures as well as the concept of multi-task learning. Furthermore, the obtained results are shown and discussed in Section 4. In the final Section 5, the described work is summarized and suggestions for future work are presented.

2. Data analysis

The employed case ship of the present study is a 2800 TEU Panamax container vessel built in 1998. The vessel is displayed in Fig. 1 and its main particulars are listed in Table 1. During a four-year period between August 2007 and July 2011, the vessel was equipped with an extensive sensor framework and sailed in cross-Atlantic service. Even though the actual data acquisition period was conducted over 4 years, the Miros wave radar Wavex was only installed until March 2009. Hence, the span of the data used herein is 1.5 years and the GPS position of the vessel during that time is depicted in Fig. 2. Additional details of the measurement campaign including a study pertaining to structural fatigue due to wave excitations, are described in Storhaug et al. [24].

It is noteworthy that draft and trim were not measured. This sort of epistemic (i.e. systematic) uncertainty within the data may be influential on the following machine learning approach. In the paper of Storhaug et al. [24], it is stated that the typical transit draft is approximately 9.5 m and the transit trim lies between 0.5 to 1.0 m. In addition, the loading condition, and thus the transverse metacentric height GM_T , is not included in the dataset adding further uncertainty due to the widespread operational profile of container vessels, in general.

In Fig. 2, the GPS position history of the case ship is visible and the effect of seamanship as well as weather routing stands out since several routes deviate notably from the shortest distance. In one particular case, the container vessel even sails around the British islands in order to circumnavigate possible adverse weather conditions. The case ship trades between Western Europe (France, Belgium, Great Britain and Germany) and Quebec in Canada. Moreover, the ship obviously operates in coastal and restricted waters, such as the St. Lawrence river; however, the focus of the present study is on deep water conditions. For this reason, samples possibly influenced by shallow water or refraction from the coastline are disregarded by enforcing spatial boundaries at -55 and -5 degrees of Longitude as indicated in Fig. 2.

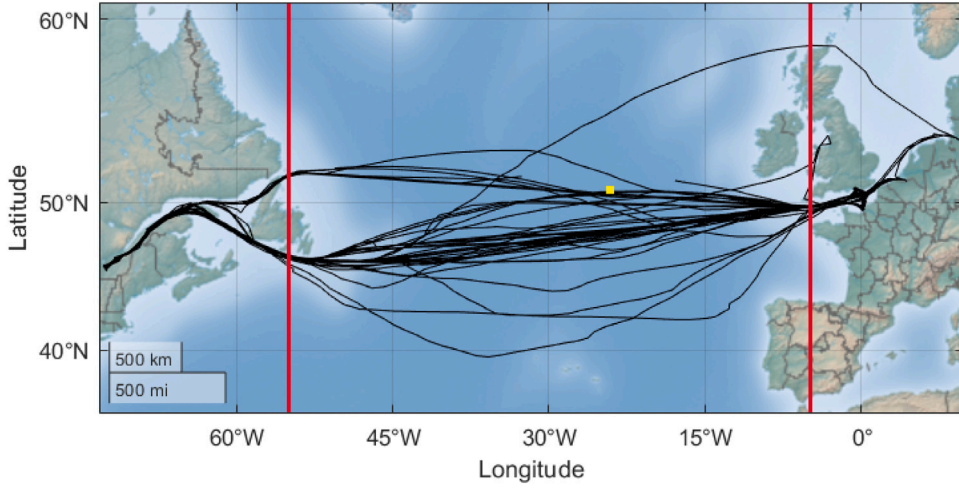


Fig. 2. Overview of the vessel's GPS position history (black) and the applied spatial boundaries (red). The yellow dot indicates one particular measurement sample.

The on-board measurement system comprises: (1) A Miros Wavex directional wave radar providing an averaged spectrum in a 30 min interval based on 1-min sample periods. The directional wave spectrum $E(f, \mu)$ is discretized into 31 frequencies f [Hz] and 36 directions μ [rad]. X-band marine radars are nowadays frequently applied as ocean remote sensors onboard of vessels due to their high spatio-temporal resolution. However, the accurate estimation of the significant wave height is a matter of empirical corrections (cf. Gangeskar [25]) and also the effect of rain clutter on data quality is noteworthy (cf. Chen et al. [26]). The accuracy of the Wavex radar is in the range of $\pm 10\%$ in case of the three sea state parameters, cf. Miros [27]. The working principles of wave radars are conveyed in Barstow et al. [28]. (2) The Motion Response Unit (MRU) was located 78.5 m forward of the aft perpendicular (AP), 11.7 m above base line (BL), and at the centre line (CL). The MRU was installed in a socket filtering high frequency noise in the range of 50–100 Hz caused by e.g. thrusters or pumps. (3) The bow accelerometer is installed on a vertical pillar in the bosun store at the forward perpendicular and measures the vertical bow acceleration. (4) Moreover, four strain gauges are attached to longitudinal stiffeners at port and starboard in the aft section (50.3 m from AP) and amidships (118.7 m from AP). As described in [24], the port and starboard strain sensors are aggregated into two artificial or virtual sensors indicating vertical bending and axial stress in the aft and midship section. Lastly, (5) the propeller revolutions (rpm), (6) the rudder angle δ , (7) the GPS position, as well as (8) the Speed Over Ground (SOG) were obtained during data acquisition period.

The present study is focused on the prediction of the three integral sea state parameters: The significant wave height H_s , peak period T_p and mean encounter wave direction β . The parameters were obtained from the directional wave spectrum provided by the Wavex wave radar measurements. The significant wave height H_s relates to the zeroth order spectral moment m_n (with $n = 0$) and is defined using Eqs. (1)–(3). It is noted that the angular wave frequency is $\omega = 2\pi f$ and the directional wave energy density spectrum is denoted as $E(\omega, \mu)$.

$$H_s = 4\sqrt{m_0} \quad (1)$$

$$m_n = \int_0^\infty \omega^n F(\omega) d\omega \quad (2)$$

$$F(\omega) = \int_{-\pi}^{\pi} E(\omega, \mu) d\mu \quad (3)$$

The peak period T_p is extracted from the integrated wave spectrum $F(\omega)$ and corresponds to the period at which the wave energy density is highest, cf. Eq. (4).

$$T_p = \frac{2\pi}{\omega_p}, \quad F(\omega_p) \equiv \max_{\omega} F(\omega) \quad (4)$$

The mean encounter wave direction β is calculated as the circular mean according to Longuet-Higgins et al. [29] in Eqs. (5)–(7). It is noted that the direction of the measured wave spectrum is relative to the ship's heading and, thus, no transformation is necessary.

$$\beta = \arctan(d/c) \quad (5)$$

$$d = \int_{-\pi}^{\pi} \int_0^\infty E(\omega, \mu) \sin(\mu) d\omega d\mu \quad (6)$$

$$c = \int_{-\pi}^{\pi} \int_0^\infty E(\omega, \mu) \cos(\mu) d\omega d\mu \quad (7)$$

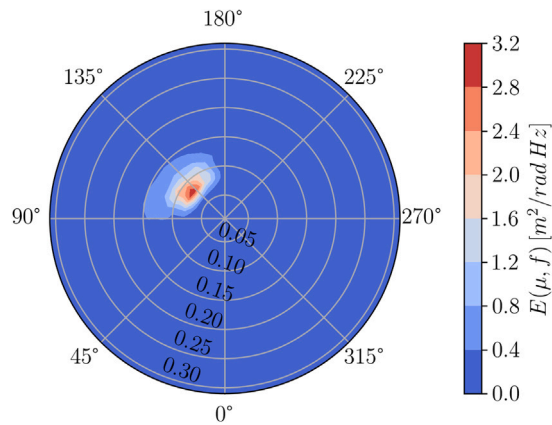


Fig. 3. Approaching directional wave spectrum E obtained at the yellow dot in Fig. 2. It is stressed that $\beta = 180\text{deg}$. indicates head waves and $\beta = 0\text{deg}$. stands for following waves; that is, in the polar diagram, the ship can be imagined to have its centre line aligned with the chord (diameter) from 0 to 180 deg.

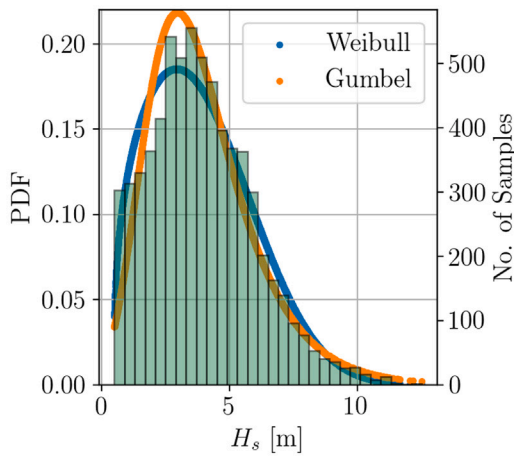
In Fig. 3, a sample directional wave spectrum is depicted in its approaching form and it was obtained at the yellow dot in Fig. 2 on the 9th April 2008 (01:30 UTC). It is noted that $\beta = 180\text{deg}$. refers to head wave conditions. As can be seen, the ship encountered bow oblique waves and the calculated parameters using the aforementioned equations are: $H_s = 4.0\text{ m}$, $T_p = 12.1\text{ s}$ and $\beta = 124.8\text{deg}$.

The available data is preprocessed and filtered in four individual steps: (1) Initially, the wave radar data was merged with the available GPS data and filtered for deep sea conditions using the enforced geographic boundaries (cf. Fig. 2). This reduced the wave radar samples from 10 790 to 7051. (2) In addition, the wave radar data was synchronized with the entirety of sensor readings, but samples with missing or corrupted recordings were disregarded. The majority of the readings had different sample frequencies ranging from 100 to 1 Hz and hence 25 min time series samples sampled with a consistent frequency of 5 Hz were extracted for each timestamp, i.e. in a 30 min interval. It is noted that 25 min are chosen as the maximum sample length since recordings close to the 30 min threshold were frequently missing. In addition, the focus of the herein presented work is on using smaller time frames, noting that other studies use longer durations, say, from 30 min in [23] to 60 min in [21]. The synchronization step decreased the size of the dataset further to 5182 samples. (3) Lastly, the data was cleaned manually from erroneous time series and (4) samples with significant wave heights $H_s < 0.5\text{m}$ were excluded, as the vessel's response is negligible in this case. Altogether, the final dataset comprises 4779 samples, and the distributions of the resulting sea state parameters as well as the ship advance speed U_s are presented in Fig. 4.

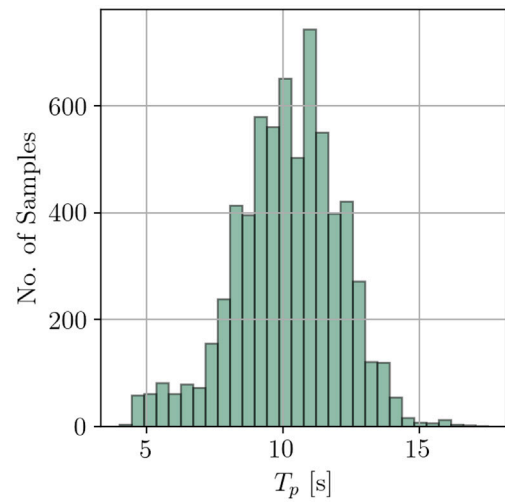
As can be inferred from Fig. 4(a), the probability density of the significant wave height follows an exponential distribution. The theoretical probability density functions (PDF) Weibull and Gumbel are fitted to the measured data distribution using the Python library `scipy`.¹ The conclusions drawn in Nielsen and Ikonomakis [30] are also observed in this case: The Gumbel distribution matches the actual data with higher accuracy than the Weibull PDF. In addition, it is visible that the ship is experiencing harsh conditions despite seamanship and the utilization of weather routing. The distribution of the peak period, which is depicted in Fig. 4(b), is symmetrical and centred around 10 s. The histogram of the relative wave direction as shown in Fig. 4(c) conveys that predominantly head and following wave conditions are encountered which results from the fact that the Northern Atlantic is known for dominant westerly wind conditions. Furthermore, the speed variation including involuntary and voluntary speed loss due to waves can be seen in Fig. 4(d), where the data distribution arranges around an mean advance speed of 19 knots. In addition, the measured joint distribution of H_s and T_p is compared to long term wave climate statistics in the Northern Atlantic provided by Söding [31] in Fig. 5. It is noted that Fig. 5(a) is a combined scatter, kernel density and two-dimensional histogram plot for demonstrating the overall data distribution as well as possible outliers in parallel.

As can be inferred from the comparison of Figs. 5(a) and 5(b), the joint distributions are in good agreement, in general. However, the covariance in case of the measurements is slightly larger compared to the modelled long term statistics. It is noted that the latter is based on hindcast wave spectra determined from wind fields during 10 years, [31]. Hence, different sources and magnitudes of epistemic (i.e. systematic) measurement uncertainty may be responsible for the minor deviation of both joint distributions. On the other hand, it is stated in [24] that the calibration of the Wavex wave radar is based on earlier studies and has not been performed for this particular vessel. Even in case of a slightly biased wave radar, it has no direct impact on the machine learning methodology itself, as the offset remained the same throughout the measurement campaign. Furthermore, transfer learning enables neural networks to be retrained and adjusted on newly obtained sea state data — possibly even from different vessels.

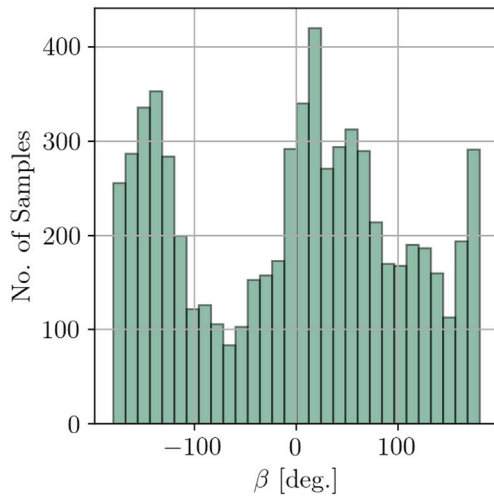
¹ <https://scipy.org/>.



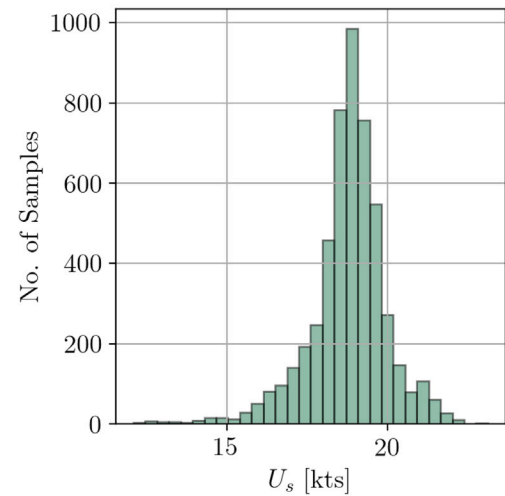
(a) Histogram of the significant wave height H_s including two parameterized PDF.



(b) Histogram of the peak period T_p .



(c) Histogram of the relative wave direction β , where $|180|$ deg. indicates head waves and 0 deg denotes following waves.



(d) Histogram of the ship advance speed U_s according to GPS data.

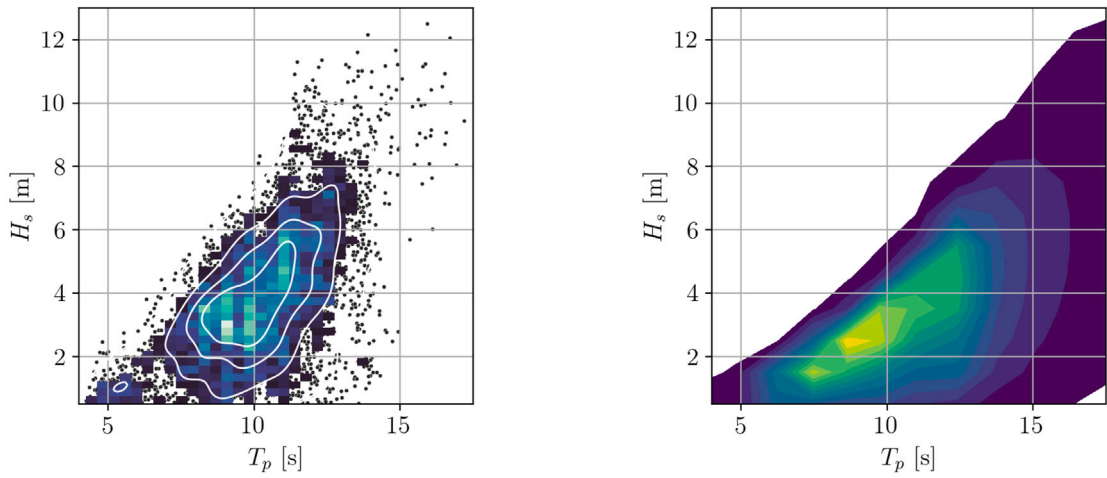
Fig. 4. Data distributions of integral sea state parameters and ship advance speed.

3. Methodology

The surface wave elevation in realistic seaways is often assumed to be stationary, ergodic, and a Gaussian random process throughout the observation time. Following St. Denis and Pierson [32], a ship can be seen as a linear and time-invariant filter in the frequency domain under these assumptions. Specifically, this leads to the underlying principle of the wave buoy analogy, emphasizing a number of associated characteristics and concerns: (1) The filtering characteristics of ships are governed by their hull geometries and the relative size compared to waves. (2) Generally, a ship acts as a low-pass filter which means that it is irresponsive to higher wave frequencies. Furthermore, (3) when sailing with non-zero forward speed the Doppler shift is introduced and (4) the assumption of the ship as a linear and time-invariant filter is subject to uncertainty under severe sea states or changing wave conditions — especially in case of forward speed. All of these characteristics have an impact on a machine learning-based sea state identification methodology.

3.1. Data processing

Initially, Mittendorf et al. [33] suggest that transformations to non-linear and skewed target variables positively affect prediction accuracy even for non-linear regression algorithms, such as neural networks. For this reason, the logarithm is applied to the significant wave height H_s as it leads to a more symmetrical shape of the resulting distribution. It is noted that the application



(a) Joint distribution of H_s and T_p in deep sea conditions for 1.5 years.

(b) Northern Atlantic wave climate statistics from Söding [31].

Fig. 5. Comparison of the wave statistics in the Northern Atlantic.

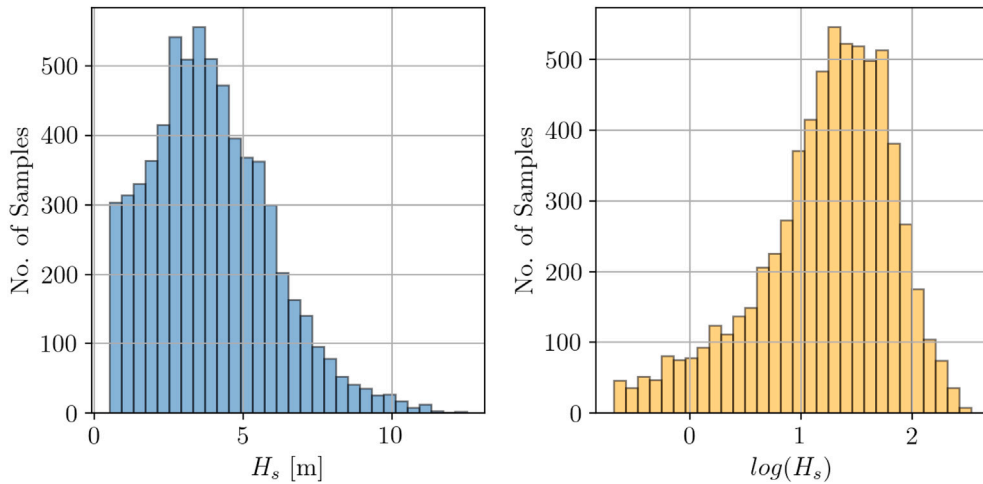


Fig. 6. Effect of the logarithmic transformation on the H_s distribution.

of the log-transform is considered as standard practice in statistics and time series analysis for forcing a more symmetrical shape onto an exponential distribution.

As shown in Fig. 6, applying the logarithm to the H_s distribution leads, in fact, to a more symmetrical shape, even though it still shows minor asymmetry in this case. Moreover, the relative mean wave direction β is decomposed into corresponding sine and cosine values, in order to circumvent the circular ambivalence in the machine learning approach. The peak period remains unchanged. Ultimately, the output vector in both the time and frequency domain methodologies has the shape of $\{\log(H_s), T_p, \sin(\beta), \cos(\beta)\}$ and is passed through a linear activation function. It is noted that input and output data are normalized beforehand, which will be described at a later stage.

In an earlier study of Mittendorf et al. [34], a time and frequency domain approach were applied based on simulated data of a vessel in unimodal and unidirectional sea states. In that study, a frequency domain model was trained on spectral moments of the auto cross response spectra and the peak spectral ordinates, as well as the corresponding peak frequencies obtained from the off-diagonal cross response spectra. The temporal method, on the other hand, was based on raw heave, pitch and roll time series. In the present work, raw 5 min acceleration time series with a sample frequency of 5 Hz are fed into the time domain models based on the findings of [34]. This leads to an input matrix of $1500 \times M$, where M is the number of features. A sample plot of 5 min acceleration time series resulting from the encountered sea state in Fig. 3 is depicted in Fig. 7. It is noted that a coordinate system is adopted herein, where the x_1 -axis coincides with the ship's centreline (positive forward) and the x_2 -coordinate points upwards.

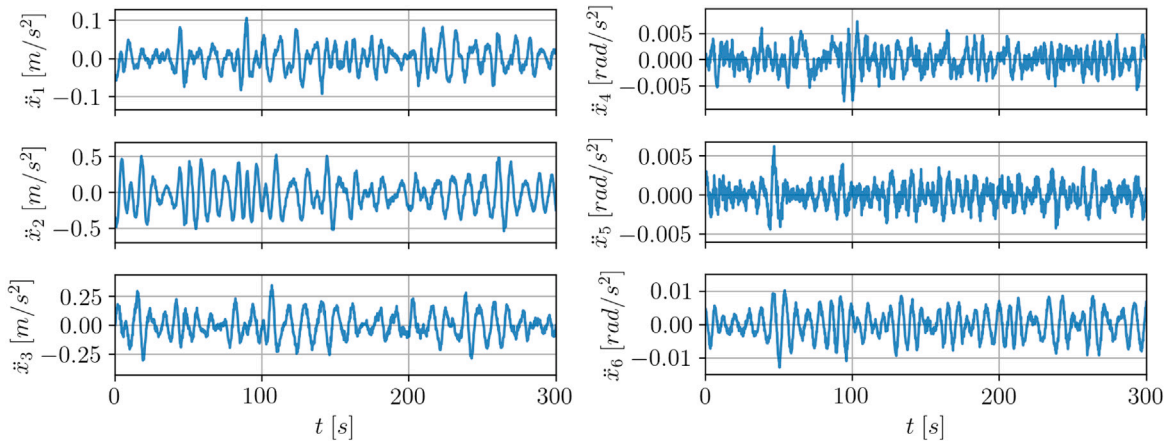


Fig. 7. Measured accelerations corresponding to Fig. 3 in 6 degrees of freedom, namely surge, heave, sway, roll, yaw and pitch (starting at the upper left plot).

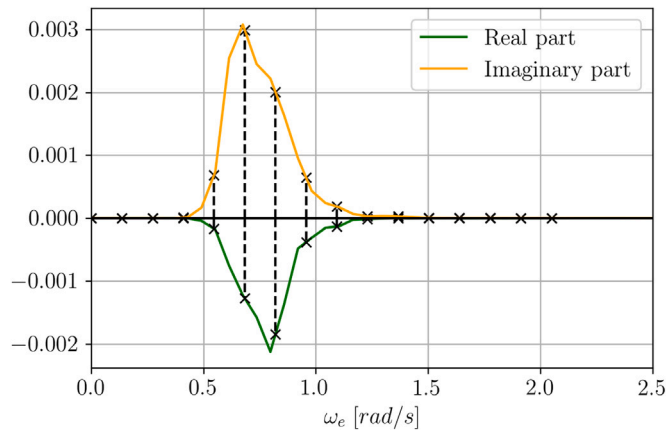


Fig. 8. Pitch-heave acceleration cross response spectrum demonstrating the frequency discretization with 16 spectral ordinates (black) resulting from the conditions in Fig. 3.

It is stressed that not only the 6 Degree of Freedom (DOF) accelerations – shown in Fig. 7 – are considered in the feature space, but also the bow acceleration and the strain measurements at the aft and midship sections. The underlying reasons are that both a second reference point on the ship and an indication of the hull flexure might add value to the sea state identification approach. The optimal number of features M will be determined in sensitivity studies in both the temporal and spectral domain.

The prediction of the relative wave direction is a function of the individual phase differences of accelerations in several degrees of freedom, which are obviously directly available in the time domain, cf. Fig. 7. In the frequency domain, however, cross spectral analysis of two individual time series, is used as predictor for β . In contrast to [34], sequences of spectral ordinates extracted from cross response spectra are fed into the neural network. This leads to an $N \times N$ complex-valued matrix, where N is the number of the considered DOF. It is noted that only one side of the off-diagonals are taken into account, as the matrix is complex conjugate symmetric. It is noted that the sensor recordings of the bow acceleration and the strain measurements are also examined in the sensitivity study and may be part of the final feature space. The matrix of cross-spectral densities is determined using the Welch algorithm (cf. Welch [35]) with a Hanning window as well as a segment length of 512. The employed software package for the digital signal processing procedures is scipy. In Fig. 8, the pitch-heave acceleration cross spectrum and the chosen frequency discretization are visible for the same conditions as in Fig. 3.

In view of Fig. 8, it is appreciated that the discrete spectral ordinates are extracted in the range of $\omega_e \in [0, 2.05]$ rad/s with a number of discrete frequencies denoted as Ω . The cut off frequency is selected for filtering the measurement noise and for responses due to excitations which are not related to waves, e.g. hull-structural vibrations. From Fig. 8, it appears that the peak value is well captured in case of the imaginary part of the cross response spectrum, but not in case of the real part. Thus, the number of discrete frequencies Ω is part of a second sensitivity study, in which also the effect of the sample length on prediction accuracy will be investigated. For reference, 71 and 100 discrete spectral ordinates were extracted from individual cross response spectra in the work of [21,23], respectively. Lastly, the shape of the input matrix for the machine learning model in the spectral domain has a shape of $\Omega \times M$, where M takes a maximum value of 43. Firstly, 30 real and imaginary sequences from the off diagonals of the

cross response spectra and 6 sequences from the auto cross response spectra, i.e. the diagonal elements, will be part of the largest possible feature space (with $N = 6$). Moreover, 2 response spectra from the strain measurements in the aft and midship sections may be included in the input matrix. In addition, the real-valued response spectrum of the bow acceleration as well as 4 real and imaginary sequences from the cross spectra of bow and heave/pitch accelerations may be also considered as features. It is stressed that only real valued numbers are permitted as input to the models. The optimal number of features M and thus the shape of the input matrices in both domains will be obtained in sensitivity studies presented in the further course of this contribution.

Before feeding the data into the neural networks, the dataset is split into training and validation sets, where the latter is exclusively used for the model assessment. The validation set makes up 20% of the initial dataset and comprises 956 samples. The training set, on the other hand, has a sample size of 3823. Neural networks are not scale invariant, thus, the individual sequences as well as the elements of the output vector were normalized according to the extreme values of the training set.

3.2. Model architectures

The machine learning task itself is considered as a supervised regression approach and the models are trained on sequences from the temporal and spectral domain. As has been stated in Section 1.3, the focus of this contribution is on deep learning methods. Artificial neural networks are utilized as universal function estimators and are considered as composite functions \mathcal{F} of two essential building blocks: The weight matrix and the non-linear activation function. Concise theoretical considerations of artificial neural networks are presented in Mittendorf et al. [33] and more detailed information is delivered by Goodfellow et al. [16].

Broadly speaking, deep learning architectures are capable of coping with high dimensional input tensors and have the ability of implicit feature construction, i.e. derive meaningful data *themselves*. Additional advantages of deep neural networks over traditional machine learning algorithms include higher scalability and increased generalization capability. The caveats of deep neural networks are on one hand the relatively high computational cost of their training process, and on the other hand the tendency of overfitting on small datasets due to their relatively large amount of trainable parameters. While, one particular problem has been most influential on the development of deep learning methods: As the name suggests, deep learning achieves its superior performance by stacking individual layers. However, very deep models suffer from the *vanishing gradient problem*. The parameters of neural networks are updated under backpropagation in a gradient-based optimization problem of an arbitrary loss function \mathcal{L} . However, with an increasing number of hidden layers, the magnitude of the gradient, i.e. the impact on parameter adjustment, reduces. Batch normalization as proposed by Ioffe and Szegedy [36] standardizes the output of hidden layers before applying the activation function mitigating vanishing gradients by resetting the parameter distribution. In addition, batch normalization leads to robust neural networks reducing the need for hyperparameter optimization. The simplified activation function Rectified Linear Unit ($ReLU = \max(0, x)$) not only decreases the chance of vanishing gradients, but also speeds up the training process, as it is faster to differentiate in comparison to traditional activation functions, such as the hyperbolic tangent function. Moreover, numerous advanced architectures, such as the LSTM (Long short-term memory) by Hochreiter and Schmidhuber [37] or residual networks by He et al. [38], were proposed for alleviating the occurrence of vanishing gradients. In this paper, the residual network as well as the Inception model proposed by Szegedy et al. [39] are applied.² The herein employed models feature one-dimensional convolutional layers, since they are best-suited to sequential data. Convolutional layers utilize spatially shared weights and are often followed by a pooling or subsampling procedure. In theory, the convolution procedure has a multidimensional tensor as input, which is modified by a kernel whose parameters are trainable. For more elaborate details consult the work of Krizhevsky et al. [40]. The concepts and structures of the herein applied models are described in the following paragraphs. Moreover, the model architectures and hyperparameters are the same in both the time and frequency domain for consistency.

3.2.1. Residual network

The Residual Network (ResNet) features a block-wise architecture and the blocks are bypassed by identity mappings or gates. The ResNet was proposed by He et al. [38] and its concept draws inspiration from the pyramidal cells of the cerebral cortex and its skip connections. The feedforward residual network is made out of multiple residual blocks as shown in Fig. 9(a) and the skip connections allow the development of very deep models without vanishing gradients. The internal shape of the proposed residual block comprises three convolutional layers of constant filter size, but with 8, 4 and 2 kernels, respectively. Filter size refers to the number of output dimensions, whereas the kernel size denotes the length of the receptive field, i.e. of the convolutional window. Lastly, the output of the block is the sum of the last layer's output and the input matrix gated through the skip connection activated by ReLU, as can be seen in Fig. 9(a). It is noted that the identity mapping is multiplied by a linear projection for expanding the channels of the skip gate, in order to match the shape of the residual block's output. The 4 blocks of the chosen model have filters with the size of 32, 64, 64 and 64, respectively. The tail part of the ResNet consists of an average pooling layer and a fully connected layer with a width of 50 neurons. It is noted that batch normalization is applied after each convolutional layer.

² An early version of the Inception Network, namely the GoogLeNet, and a residual network were the winners of the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) in 2014 and 2015, respectively.

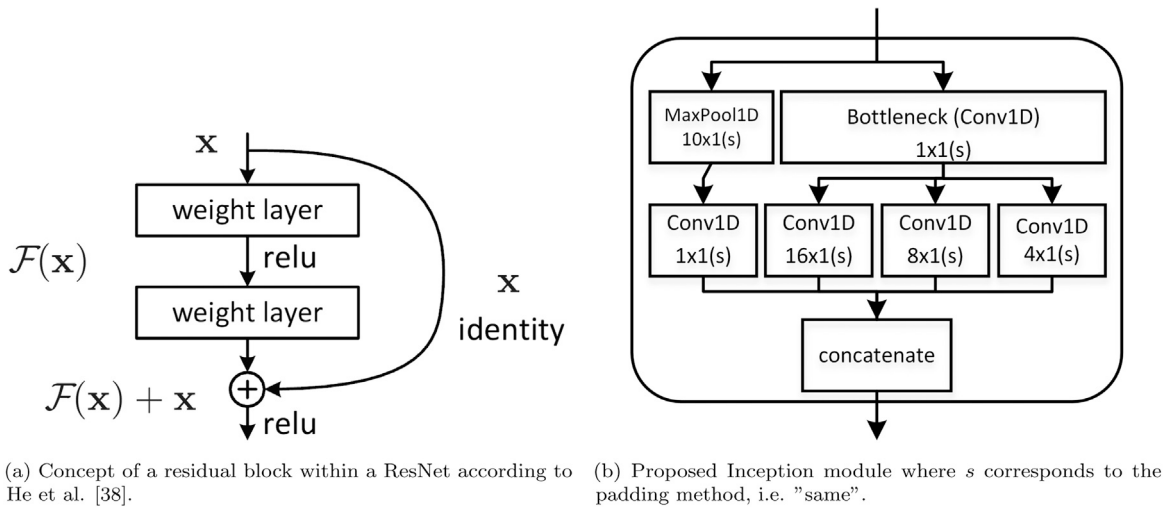


Fig. 9. Underlying concepts of the ResNet (a) and Inception (b).

3.2.2. Inception network

The Inception architecture was initially proposed by Szegedy et al. [39] in an end-to-end image classification procedure. The model follows a similar block-wise methodology as the ResNet employing the novel Inception module as its building block. The fundamental idea is that several convolutional layers of constant filter size of 16, but with different kernel sizes, i.e. different receptive fields, are employed simultaneously at the same level. In doing so, the individual receptive fields capture different patterns in the feature maps at varying scales. This concept is inspired by principles of the visual cortex in the human brain. The Inception architecture is named after the eponymous movie³ indicating the common premise of embedding either networks within networks in case of the model or dreams within dreams as in the movie. The architecture evolved in multiple iterations to Inception-v4, where the model was extended with residual connections for increased performance, Szegedy et al. [41]. The proposed Inception module has been adapted for sequential data following Ismail Fawaz et al. [42], i.e. one-dimensional convolutional layers are employed. As can be inferred from Fig. 9(b), the module splits into four elements: (1) The bottleneck or dimensionality reduction element is introduced in front of the computationally intensive parallel layers and takes cross-channel patterns into account. The bottleneck layer reduces computational cost as well as the number of trainable parameters. (2) The second element is made out of the three parallel convolutional layers with different kernel sizes, i.e. the core part of the Inception module. In the present work, kernel sizes $l \in \{16, 8, 4\}$ are applied. (3) The third part of the module is the skip connection, which is inspired by the ResNet, and consists of a pooling layer and another dimensionality reduction layer. (4) The two gates are concatenated along the depth dimension in the last element. It is stressed that the Inception architecture has additional outer residual skip connections bypassing entire Inception modules. The general structure of the proposed Inception network can be inferred from Fig. 9(a) when replacing the residual block with the Inception module shown in Fig. 9(b). The herein proposed Inception network is defined with filter size of 16 and 8 inception modules in total with individual residual connections. In front of the output layer, the activations are fed through an average pooling layer and batch normalization is also applied in the entire model prior to all activation functions.

3.2.3. Benchmark models

For comparison of ResNet and Inception to state of the art literature, two models are adopted as benchmarks: The convolutional neural network (CNN) from Kawai et al. [21] and the multichannel convolutional LSTM (MLSTM) network as proposed by Düz et al. [20]. The former network was applied in the frequency domain on cross spectral sequences, whereas the latter one used raw time series as input for the prediction of sea state parameters.

3.3. Multi-task learning

Multi-Task Learning (MTL) has been pioneered by Caruana [43] and is considered as *parallel* transfer learning. As opposed to a multioutput-regression task, each output is now considered as a separate task and has, thus, its own dedicated branch of fully connected hidden layers and corresponding output layers. In the present paper, it is thought that considering sea state estimation as a multi-task learning approach may be beneficial as there are obviously physical interdependencies between H_s and T_p . According to Ruder [44], a hard parameter sharing configuration is selected herein, as can be seen in Fig. 10, which mitigates overfitting and improves generalization immensely.

³ Inception is a science fiction movie from 2010 written and directed by Christopher Nolan.

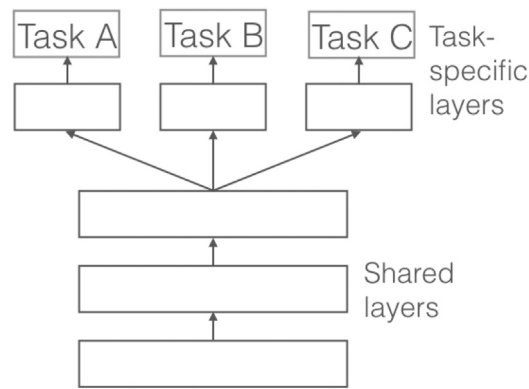


Fig. 10. Hard parameter sharing concept from Ruder [44].

In view of Fig. 10, it is appreciated that the model is said to learn several tasks at once using a shared latent data representation. Following [44], MTL improves the model's generalization capability based on eavesdropping, i.e. the network learns complex features and the relationships between other dependent variables, instead of receiving them as inputs. In addition, MTL leads to increased regularization for improving the model's robustness by reducing Rademacher complexity, i.e. the model's tendency to fit to stochastic noise. In this work, each task-related branch consists of two hidden layers with 100 and 50 neurons and an output layer. Two different MTL architectures are examined in this contribution: (1) The *MTL* architecture has three branches for each of the sea state parameters, however, the output layer of the third branch has a length of two, as both the sine and cosine of β are considered. (2) The *MTL+* architecture, however, takes the advance speed U_s as a fourth task into account. It has been demonstrated by Caruana and de Sa [45] that potential input features with low variance may be better predicted as an additional task instead, when the overall input data is subject to large variance — as in the present case. In addition, the inclusion of the mean advance speed in the input space is impractical due to the mismatching dimensions. It is thought that the model is able to derive an understanding of the Doppler shift and thus the forward speed for the correct prediction of the sea state. Hence, it will be studied herein, whether the prediction of the forward speed enhances the prediction of the sea state parameters and could act simultaneously as a reliability indicator. Lastly, the tasks of the two investigated architectures are related or even complementary, however, adversarial training as in [23] is also feasible using multi-task learning.

3.4. Training setup

In the present work, the ADAM optimizer – proposed by Kingma and Ba [46] – is employed as the learning algorithm. The ADaptive Momentum estimation (ADAM) optimizer uses the exponential moving average of the gradient and scales the learning rate, i.e. the step size, according to the squared gradient. Moreover, the applied loss function is the mean squared error, the chosen batchsize is 64, and the initial learning rate is set to 10^{-3} . Mak and Düz [47] suggest that k-fold cross validation with shuffled samples leads to increased performance in a sea state estimation methodology. Hence, we apply shuffled 5-fold cross validation in 300 epochs. The employed model-checkpoint-callback stores the model's parameters with the smallest cross validation loss, in order to save the model with the highest generalization capability before overfitting occurs with an increasing number of epochs. The computations of the training procedures were carried out using a GPU node of the DTU computing center equipped with two Nvidia© Volta-100 GPUs, each with 16 GB of memory (RAM) and multiple Intel© Xeon© Gold 6126 CPU with 2.60 GHz were used. Moreover, the utilized programming language is Python 3.6 and the deep learning framework is TensorFlow 2.6 as proposed by Abadi et al. [48]. For GPU parallelization of the computations CUDA (Compute Unified Device Architecture) was employed.

4. Results and discussion

In the following, the obtained results are presented and discussed. The used metrics are the root mean squared error (*RMSE*) and the mean absolute error (*MAE*), and their definitions are given in Eqs. (8) and (9). It is noted that \hat{y}_i indicates the model's prediction, whereas y_i refers to the wave radar data.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2} \quad (8)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |\hat{y}_i - y_i| \quad (9)$$

The *RMSE* attributes more weight to outliers or variance and the *MAE* indicates the magnitude of the error without considering the sign. Due to the circular ambiguity in case of the mean encounter wave direction, a similar approach as in Nielsen [49] is

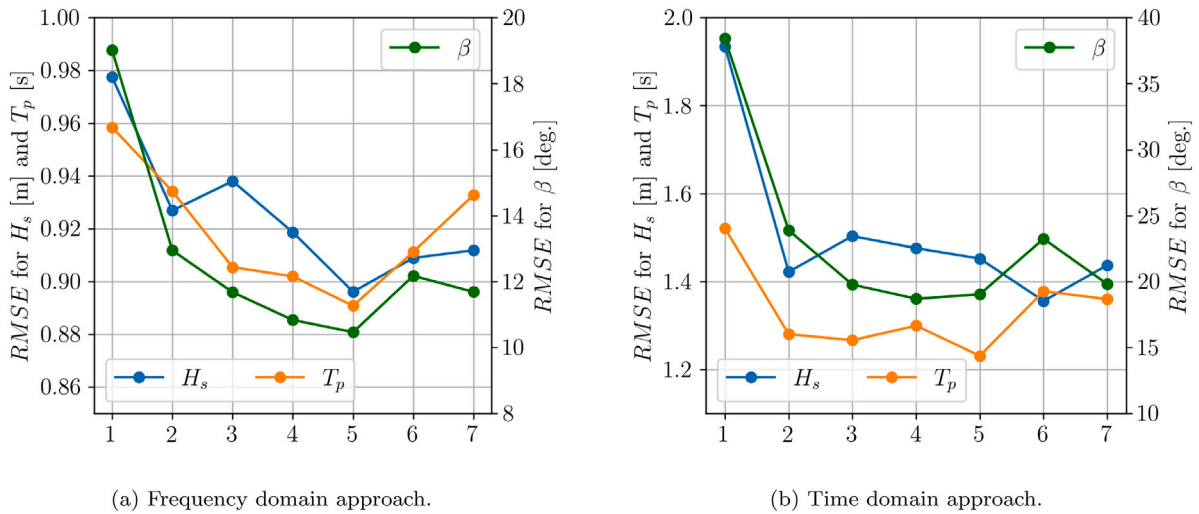


Fig. 11. Optimal selection of the considered ship responses according to the performance on the validation set. Note that the individual scales on the ordinates are different.

adopted for error calculation, i.e. absolute values of β are considered when calculating the performance metrics. It is stressed that the accuracy on the *validation* dataset is assessed exclusively in this section and that the model performance on the *training* set is presented in the [Appendix](#).

4.1. Sensitivity studies

Model-based approaches of the wave buoy analogy, e.g. Nielsen [7], traditionally consider heave and pitch as well as one additional asymmetric motion, such as roll or sway for satisfactory results with a distinguishing between waves approaching from port or starboard side. The use of surge and yaw, however, is generally not pursued, since the RAOs provided by, say, strip theory often turn out to be unreliable due to the missing restoring term. However, as the present study is not reliant on RAOs, all 6 DOF accelerations are considered in the sensitivity study. Mittendorf et al. [33] show in multiple iterations that there is an optimal number of features M to be found in a machine learning approach. In this work, the trade-off between accuracy and complexity will be determined in a convergence study, i.e. the number of features and indirectly the number of measurement devices are increased incrementally. Therefore, multiple models are trained on varying, expanding input matrices, and their *RMSE* on the validation dataset is taken as the underlying score for all three sea state parameters. Based on the findings of Mittendorf et al. [34], a ResNet was chosen as the baseline model using a frequency discretization with $\Omega = 42$ components. In Fig. 11, the following combinations are considered in individual scenarios: (1) heave and pitch acceleration, (2) + roll acceleration, (3) + sway acceleration, (4) + surge acceleration, (5) + yaw acceleration, (6) + vertical bow acceleration and (7) + deck strain measurements in the aft and mid-ship positions.

When comparing Figs. 11(a) and 11(b), it is appreciated that the shape of the individual curves of the scores for H_s , T_p and β is generally in qualitative agreement — even across domains. Moreover, it stands out that the consideration of all 6 DOF accelerations yields the lowest scores, i.e. the highest out-of-sample accuracy. Thus, it is concluded from a practical point of view that, in the present case, only the motion recordings from the MRU are needed; in the sole event of sea state identification. In fact, bow acceleration and strain measurements do not lead to increased accuracy, but add complexity and noise into the machine learning model. This may be attributed to the fact that the contained information in the mentioned features is redundant, but a larger input matrix leads to a larger number of model parameters (i.e. complexity). Interestingly, the magnitude of the *RMSE* of the targets in the time domain are roughly twice the ones in the frequency domain model, which is likely to be caused by the shorter considered sample length. As expected, the inclusion of roll acceleration in the second scenario leads to the most significant increase in accuracy for all sea state parameters. In theory, H_s and T_p , i.e. the energy content of the spectrum and its distribution, may be determined by just considering heave and pitch. However, the consideration of an asymmetric acceleration is essential for the sufficient estimation of β . Lastly, an aggregated sensitivity indicator is used, i.e. irrespective of different conditions, in this study. However, Montazeri et al. [50] as well as Andersen and Storhaug [51] indicate that the sensitivity of individual responses is not constant, but variable under different sea state conditions. For instance, the relative importance of considering the vertical bending moment or its axial stress in the hull girder increases in relatively low period waves, due to increased bandwidth of these RAOs compared to the motion response RAOs, [50]. Therefore, [51] suggest an adaptive selection of the optimal responses in case of a model-based methodology. Yet, it is stressed that a dynamic methodology is not feasible in this case, as the underlying computation graph of neural networks is static. Ultimately, the optimal number of features is identified as $M = N = 6$ in both domains.

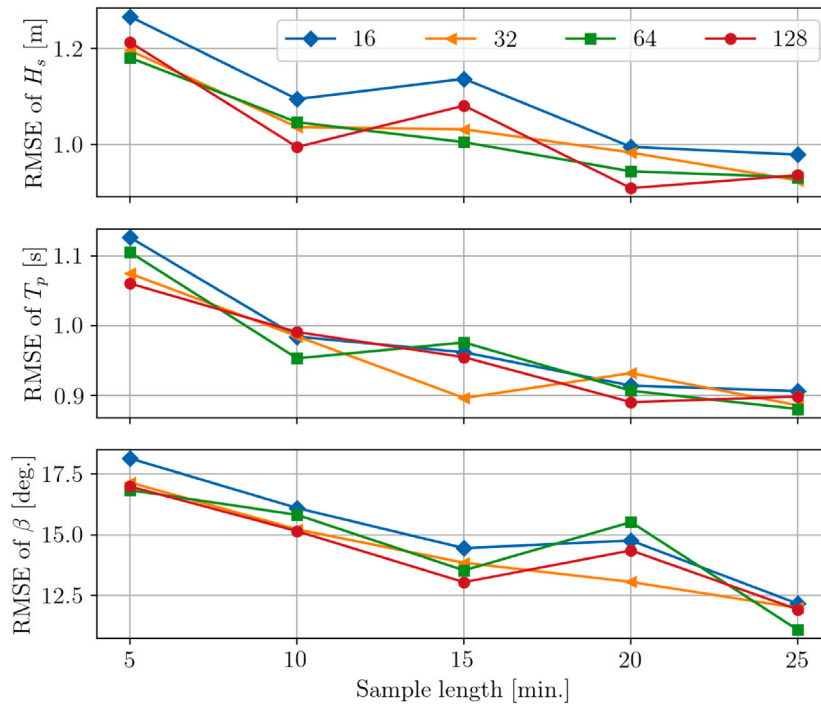


Fig. 12. Sensitivity study for sample length L and frequency discretization with Ω indicated in the legend.

Linear spectral analysis is based on the assumption of stationarity, however, it has been shown by Brodtkorb et al. [52] that the time-invariance of sea state parameters in 30 min samples is subject to uncertainty — even in a simulation-based study under zero-forward speed conditions. It is thought that forward speed and in-situ measurement data result in additional uncertainty regarding stationarity. In their work [52], it is deduced that the relative wave direction is most susceptible to instationarity, whereas H_s and T_p are seen as more stable. For minimizing this kind of uncertainty, it is the objective of this work to use shorter sample lengths as compared to 30 min in [23] and 60 min in [21]. However, the uncertainty regarding stable spectral analysis is a counteracting contribution, which will increase with shorter time series samples. Hence, the upcoming sensitivity study is about obtaining the compromise of the smallest possible sample length with the most accurate frequency resolution at the same time. The sample length L is discretized as $L \in \{5, 10, 15, 20, 25\}$ min. Obviously, the frequency discretization is interrelated and, thus, the number of frequencies Ω is considered as the second discrete variable in the study, i.e. $\Omega \in \{16, 32, 64, 128\}$. For reference, [21] use 71 and [23] employ 100 discrete frequencies. It is noted that the $RMSE$ of the model applied to the validation dataset is taken as the score.

In view of Fig. 12, the conclusions drawn in Iseki and Nielsen [53] are confirmed, i.e. the longer the time frame, the clearer the frequency resolution, and thus the higher the model score. For H_s and T_p , monotone curves with minor variance are apparent, which is expected as the accuracy obviously will be higher with less variance in the frequency representation. However, the same is visible in case of β , even though a local minimum is located at $L = 15$ min. This is in contradiction to [52] and results, on one hand, from the considered conditions, i.e. in deep water and without any major heading changes due to steering (mostly stationary conditions). On the other hand, the presented sensitivity score is highly biased by the model and its learning algorithm. Thus, the model not only reflects physical interdependencies, but also model uncertainty, since the loss function, which is minimized during training, is the sum of the errors of all three target variables. Moreover, it is trivially concluded that, in general, a finer frequency discretization results in lower error values, i.e. higher model performance or accuracy. However, in Fig. 12 it is visible that $\Omega = 64$ leads to the smallest validation error, which may be explained by the increased complexity of using 128 discrete frequencies. In the latter case, no additional information is provided to the model, but the large feature space leads to more trainable parameters, i.e. complexity. Ultimately, 25 min are identified as the best suited sample length and 64 discrete frequencies are chosen as the dimension of the feature space for the frequency domain model. Subsequently, it can be seen that the uncertainty pertaining to the spectral analysis is of greater relevance than the uncertainty of the time invariance assumption. The X-band wave radar provides samples in a 30-min interval at time t_{radar} . Thus, for consistency, we choose the starting point $t_1 = t_{radar} - 25$ min in the frequency domain approach, and in the time domain approach $t_2 = t_{radar} - 5$ min. On a side note, it can be inferred that the higher validation error in the time domain, cf. Fig. 11(b), results from the smaller sample length as the errors using $L = 5$ min are of similar magnitude in both time and frequency domains.

Table 2

Metrics of model performance on the validation set in the frequency and time domains. The lowest error values are indicated in bold font.

		Frequency domain				Time domain			
		ResNet	Inception	MLSTM	CNN	ResNet	Inception	MLSTM	CNN
H_s	<i>RMSE</i>	0.932	0.927	1.089	1.273	1.702	1.345	2.702	1.358
	<i>MAE</i>	0.620	0.626	0.767	0.868	0.909	0.916	1.979	0.991
T_p	<i>RMSE</i>	0.881	0.878	0.925	1.051	1.253	1.266	2.379	1.300
	<i>MAE</i>	0.539	0.529	0.596	0.714	0.935	0.847	2.122	0.975
β	<i>RMSE</i>	11.09	10.50	12.12	14.87	19.03	16.67	95.34	26.00
	<i>MAE</i>	12.34	11.29	77.60	16.67	12.33	11.29	77.60	16.67

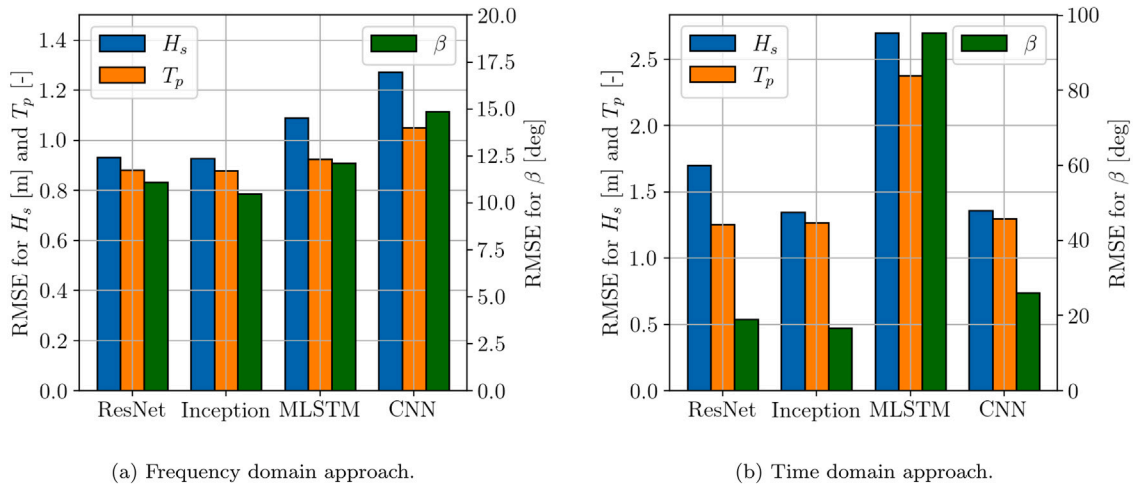


Fig. 13. Histograms of the model performance on the validation set using the *RMSE*. Note that β is presented on the secondary axis.

4.2. Model assessment

The performance of frequency and time domain models is presented in the following in parallel. In Table 2, the metrics of the prediction error for the validation dataset are presented for all four models in both domains. Moreover, the lowest error values, i.e. the highest accuracy, are indicated in bold font. For the sake of clarity, corresponding histograms of the *RMSE* taken from Table 2 are provided for all models in Figs. 13(a) and 13(b).

In view of Table 2, it is stated that the Inception network yields the lowest errors, i.e. the highest accuracy, on the validation dataset in either domain. However, it turns out that the error values of the ResNet are comparable and even lower in certain cases indicating similar model performance compared to the Inception architecture. The performance of MLSTM and CNN in the frequency domain are considered as sufficient. However, the MLSTM shows a considerable drop-off in accuracy in the time domain. It is thought that this results from the parallel architecture, i.e. splitting the model into a convolutional and a recurrent branch, leading to decreased robustness. In contrast, convolutional layers are frequently used as feature construction layers prior to LSTM cells in a sequential fashion. Moreover, it is obvious in the spectral approach (cf. Fig. 13(a)) that out of sample accuracy is dependent on model complexity, as the Inception network comprises the most trainable parameters and the CNN the least. Conversely, in the time domain, the overall picture is scattered, as the approach is subject to more variance and stochastic noise. The variance results from the short considered sample lengths and is amplified by the larger number of parameters due to the increased input matrix compared to the frequency domain approach. Thus, the temporal model is more susceptible to perturbations because of its complexity. Ultimately, the Inception network is used as the underlying model for the multi-task learning methodology and the results in both domains are conveyed numerically in Table 3 and visually using histograms in Fig. 14.

In view of Table 3, it is stated that the MTL-Inception models show satisfactory accuracy on the unseen validation dataset in both configurations. In general, they achieve higher performances than the corresponding multi-output approach as presented in Table 2. It can be seen in Fig. 14 that multi-task learning, indeed, may have a beneficial impact on model performance in a sea state identification approach, however, this applies predominantly to the time domain. In Section 3.3 it has been described that MTL reduces the model's tendency to fit to stochastic noise, i.e. Rademacher complexity, (cf. Ruder [44]) and that the MTL+ architecture performs well under noisy conditions, [45]. As can be seen in Table 3, the prediction of advance speed U_s in the MTL+ architecture is considered as satisfactory and, in fact, facilitates the prediction of sea state parameters despite the inherent uncertainty in the time domain approach. The *RMSE* is a measure of the prediction's variance and decreased consistently applying both MTL approaches. In turn, the effect of MTL in the frequency domain is not as pronounced, since there is not as large variance as in the temporal

Table 3

Metrics of model performance on the validation set in the frequency and time domains using multi-task learning. The lowest error values are indicated in bold font.

		Frequency domain			Time domain		
		Inception	MTL	MTL+	Inception	MTL	MTL+
H_s	RMSE	0.927	0.883	0.871	1.345	1.228	1.248
	MAE	0.626	0.582	0.580	0.953	0.885	0.858
T_p	RMSE	0.878	0.855	0.903	1.266	1.124	1.087
	MAE	0.529	0.472	0.501	0.916	0.752	0.744
β	RMSE	10.50	12.82	11.57	16.67	18.75	16.84
	MAE	6.446	6.916	6.767	11.29	11.56	10.33
U_s	RMSE			0.322			0.441
	MAE			0.227			0.331

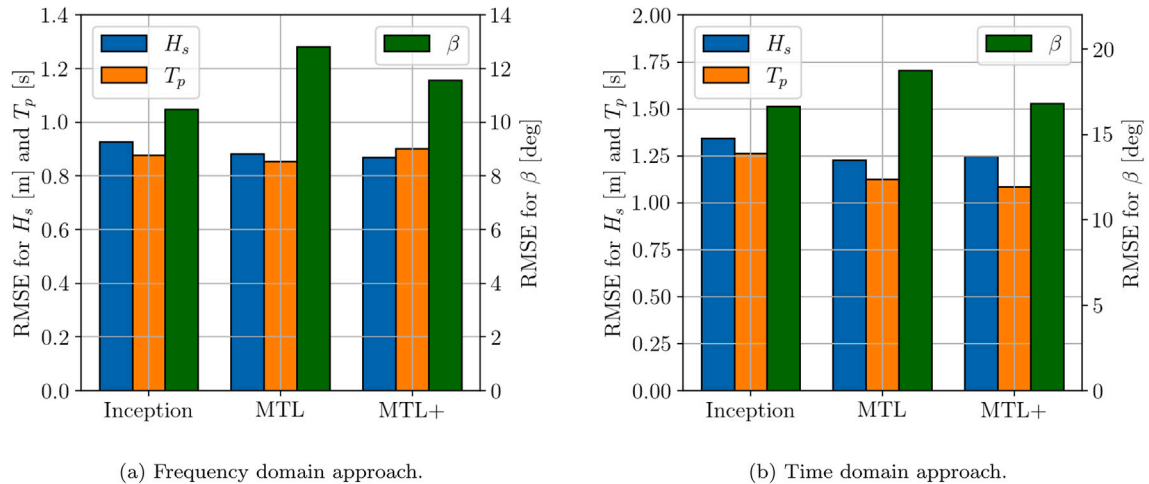
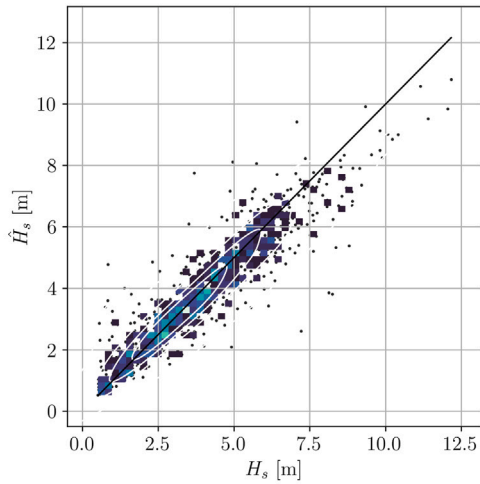


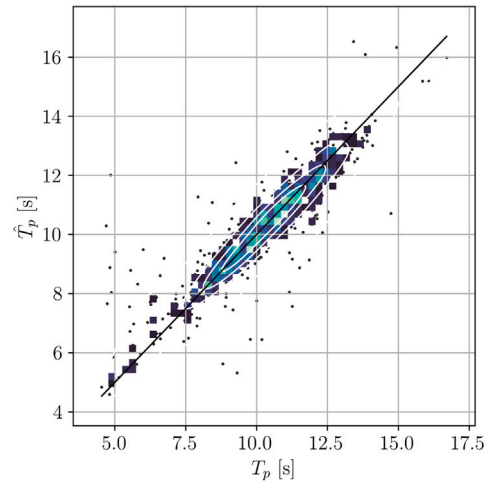
Fig. 14. RMSE histogram indicating the performance on the validation dataset of different multi-task architectures. Note that β is presented on the secondary axis.

approach. Hence, the spectral approach is considered as more robust and it is concluded that the MTL methodology increases robustness as well as accuracy for the temporal approach notably. Altogether, the MTL version is taken in the frequency domain as the final model and the MTL+ architecture is adopted in the temporal approach. In Fig. 15, the predictions on the validation set of the spectral MTL-Inception model are presented for ground truth values. Additionally, the loss curves of training and cross validation set are presented on a logarithmic scale. It is stressed that Figs. 15(a)–15(c) are of similar type to Fig. 5(a). Moreover, corresponding correlation and loss plots of the temporal MTL+Inception model are depicted in Fig. 16 for comparison.

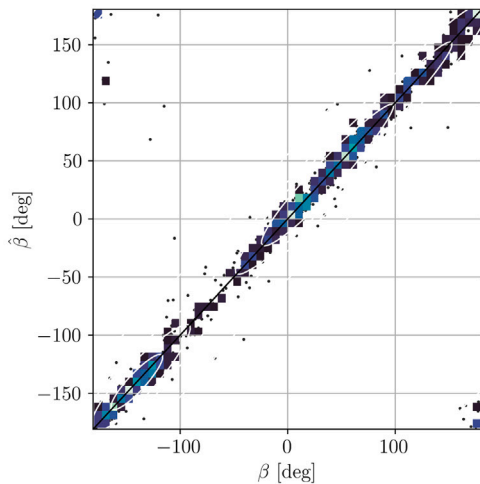
Generally, it is seen from Figs. 15 and 16 that both model types are capable of providing accurate predictions of the governing sea state. Moreover, it proves again that the temporal approach is characterized by higher variance compared to the spectral approach. This applies not only to the prediction of the three target variables, but also to the learning procedure, i.e. the behaviour of the loss functions. The loss curves are already volatile in the frequency domain due to the application of 5-fold cross validation, however, the variance is even larger in the time domain. From Fig. 16(d), it can be inferred that the time domain model was trained for only 200 epochs in order to save computational resources, as it was evident in the spectral approach (cf. Fig. 15(d)) that the model generalization converged before reaching 100 epochs anyhow. Therefore, implementing an early stopping callback is a key aspect of extending the proposed training procedure. In case of the peak period T_p , minor heteroskedastic behaviour of the predictions is visible, i.e. the variance is not constant for the entire definition range, cf. Figs. 15(b) and 16(b). Instead, the variance increases, as the magnitude of T_p decreases. It is assumed that this is due to the low pass filtering effect of the vessel, i.e. there is no vessel response in low-period waves introducing uncertainty. However, in case of higher H_s and T_p values, both spectral and temporal models exhibit reduced accuracy, which is due to the unbalanced training data and amplified by possible non-linear behaviour in more severe sea states. Furthermore, when comparing both H_s and T_p for both models, it shows that the time domain model is subject to larger uncertainty due to the more widespread joint distributions arranged around the identity line in the figures. In turn, the variance of β reveals only a small increase in the time domain model, when investigating Figs. 15(c) and 16(c). However, several outliers are visible in Figs. 15(c) and 16(c) in beam-sea conditions, i.e. $\beta = \pm 90$ deg. It is noted that the scattered squares in Fig. 16 are part of the 2D histogram and their variance stems from the uncertainty of the predictions in the time domain. Obviously, the sea state identification approach is subject to uncertainty resulting from e.g. the unknown loading conditions, as has been mentioned in Section 2. Moreover, possible unfiltered measurement errors in both wave radar and MRU, i.e. corrupted samples, increase the



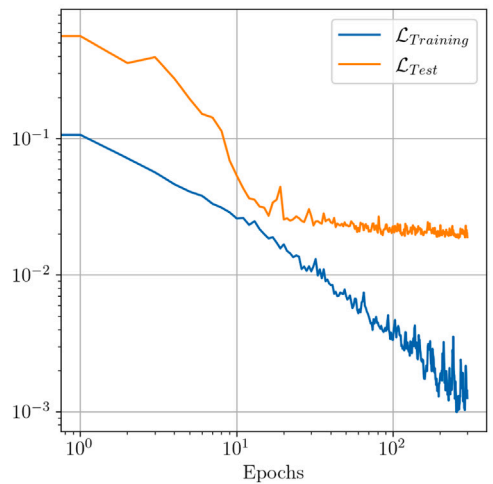
(a) Predictions of the significant wave height H_s based on the validation set.



(b) Predictions of the peak period T_p based on the validation set.



(c) Predictions of the relative wave direction β based on the validation set.



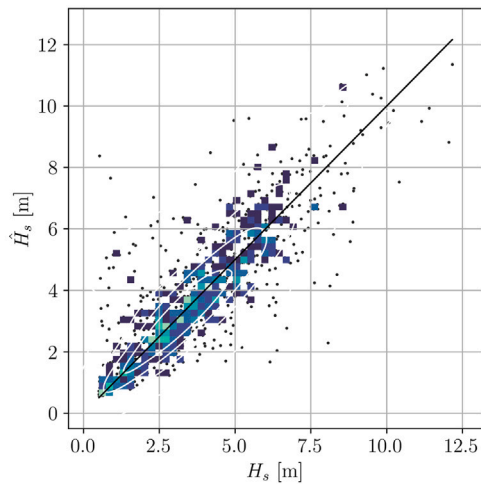
(d) Loss curves for both training and cross-validation set. Note that both axes are in logarithmic scale.

Fig. 15. Correlation plots of the target variables and the loss curves in the frequency domain. Note that overhat denotes the model predictions.

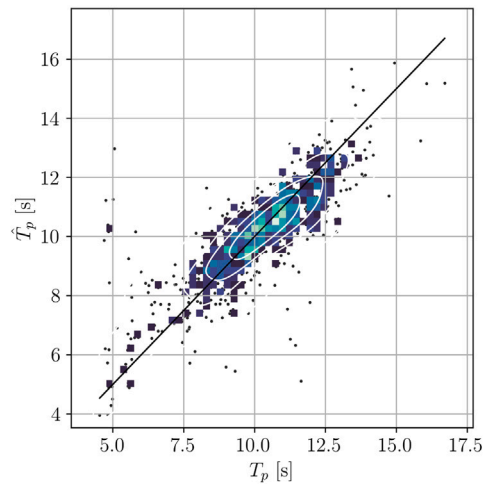
aleatoric or statistical uncertainty. As discussed in Chen et al. [26], wave radars exhibit significant inaccuracies under precipitation due to rain clutter.

4.3. General discussions

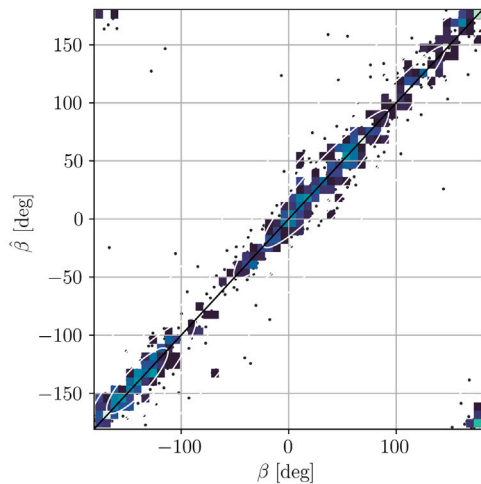
In the present work, it has been shown that the frequency domain models achieve higher out-of-sample accuracy, i.e. generalization capability, which is mainly due to the consideration of longer sample lengths as compared to the time domain. It is stressed that both the computational effort and model complexity in the time domain are proportional to the considered sample length. Conversely, both of these remain constant in the frequency domain. The computational time for one epoch – using the exact same hardware and model architecture (ResNet) – was 56 s in the time domain and in the frequency domain one epoch took 13 s. Specifically, the frequency domain requires approximately 77% less computing time, which is generally in accordance with the ratio of the numbers of elements of the two input feature spaces, i.e. 1500×6 in the time domain and 64×36 in the frequency domain, having obviously a direct impact on the trainable parameters of the model. Ultimately, it is concluded that the spectral approach is characterized by better accuracy, robustness and computational efficiency. This finding is somewhat consistent with other recent research on machine learning-driven sea state estimation that focuses on frameworks formulated in the frequency



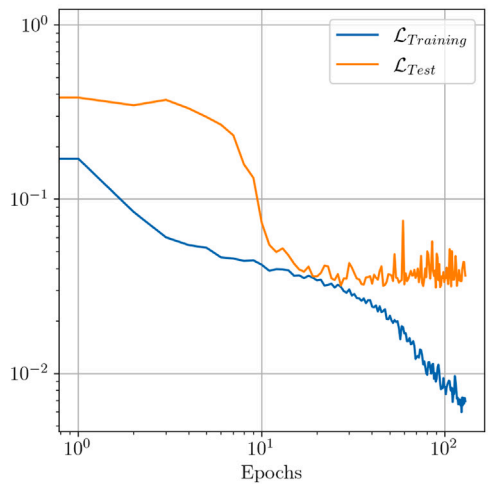
(a) Predictions of the significant wave height H_s for the validation set.



(b) Predictions of the peak period T_p for the validation set.



(c) Predictions of the mean encounter wave direction β for the validation set.



(d) Loss curves for both training and cross-validation set. Note that both axes are in logarithmic scale.

Fig. 16. Correlation plots of the target variables and the loss curves in the time domain. Note that overhat denotes the model predictions.

domain, e.g., [21,23]. In this context, it is also noteworthy that the generalization and extension towards methods for estimating the full directional wave spectrum is an important aspect of future work, such as in [22,23].

The main drawback of frequency domain approaches is considered to be the dependency on relatively long time windows, up to 25–30 min as found in this study, for a reasonable frequency resolution. Inherently, this can compromise results due to problems related to nonstationary conditions, emphasizing that even the sea state itself, as encountered from a sailing ship going maybe +20 knots, can vary because of spatial and temporal progression [30,49]. On the other hand, the literature provides promising studies that could be considered for mitigation of the drawback. For instance, Takami et al. [54] present the use of the prolate spheroidal wave functions for spectral analysis. Its main advantage over, say, Fast Fourier Transformation (FFT) is to provide a higher frequency resolution when applied on shorter time series, and with no need for manual choices with respect to smoothing. Alternatively, Cheng et al. [18] employ a hybrid approach, based on time–frequency representations of ship motions using spectrograms also indicating promising results.

Generally, sea state estimation is subject to substantial inherent uncertainty of both aleatoric and epistemic type. As such, the extension of this work towards uncertainty quantification and minimization of sea state estimation seems promising. On one hand, Mounet et al. [12] provide a methodology using a model-based sea state estimation approach while minimizing epistemic uncertainty by both reducing the uncertainty of RAOs by tuning, and spatial uncertainty by considering a network of ships at the same time. On the other hand, Han et al. [55] present an approach of visualizing and expressing the surrounding aleatoric or statistical

uncertainty in a hybrid approach with a model-based approach and Gaussian process regression. It is thought that the herein presented methodology may be extended towards both directions, however, the focus will be on the latter as aleatoric uncertainty was more prominent in the presented results. Thus, two individual approaches show potential for the transparent representation of uncertainty: (1) Following Mittendorf et al. [56], a quantile regression approach, i.e. training the neural network on quantile loss functions, seems worthwhile in a deterministic attempt. (2) Conversely, establishing a Bayesian or probabilistic model using the Monte Carlo dropout method as proposed by Gal and Ghahramani [57] shows great potential for future work, but may also be limited by increased computational effort. Lastly, Bitner-Gregersen et al. [58] conclude that the identification of the sea state with its surrounding uncertainty is and will be a vital research area in the maritime sector; and the changing climate contributes additional relevance to the topic.

5. Conclusions

Real time monitoring of the governing wave environment during a vessel's passage is directly linked to aspects of safety and economy. For this reason, the present contribution established a non-linear mapping from in-situ ship responses to prevalent sea state parameters using machine learning. The case ship was a midsize container vessel trading in Northern Atlantic. The ship was equipped with a wave radar from which data was collected during 1.5 years. Machine Learning frameworks were formulated in both the time and frequency domains using four different deep neural network architectures. The Inception model gave the highest performance and was applied in a multi-task learning setting. The model assessment suggested satisfactory performance of the multi-task frequency domain model on unseen validation data. The time domain models, on the other hand, exhibited substantial aleatoric (or statistical) uncertainty due to the considered short sample length. Lastly, the frequency domain model also showed superior characteristics in terms of computational effort and robustness. From a practical perspective, it was demonstrated that a machine learning methodology for sea state identification is applicable under realistic conditions, exclusively using response measurements from a cost effective MRU equipment.

Initially, it was thought that additional sensor recordings, such as relative wind speed and direction, could act as well-suited predictors for the sea state and for reducing uncertainty. Moreover, the consideration of propeller revolutions and rudder angle (or rather their variances) could be beneficial in a machine learning-based sea state estimation approach. These sensor readings were disregarded herein, since they are mostly instationary and filtering these samples reduces the sample size even more. Additionally, it stood out that a data acquisition period of 1.5 years only leads to approximately 5000 valid samples, i.e. complying to the enforced boundary conditions defined in Section 2. The major drawback of deep learning methods is definitely the required large amount of training data as well as limited applicability for extreme events. In turn, this motivates a parallel hybrid approach using a model-based method relying on RAOs and a machine learning model simultaneously, as investigated by Han et al. [55]. Model-based methods are not dependent on data availability, except for making the actual estimate, and may be considered as *self-supervised* in machine learning terms as they generate their directional wave spectrum estimate from an RAO database, but without any ground truth information. Therefore, it seems appealing to train a baseline model on results of model-based techniques and re-train an advanced model on e.g. wave radar data in a transfer learning approach, similar to the work of Düz et al. [20]. Hence, model-based and machine learning approaches are not seen as competitors, but as complementary — particularly when facing scarcity of sea state data. Moreover, the methodology is not dependent on wave radar data, but may also be applied to hindcast metocean data, e.g. from ERA 5 [59] being part of the EU Copernicus programme. Lastly, it is thought that the herein developed model may be considered as a baseline and could be adapted to other vessels using transfer learning in case of limited data availability. Possibly, the transfer learning procedure may be carried out in an incremental learning approach, i.e. the model is adapted in an online fashion, as more data becomes available. All of the above are interesting aspects of extending the present work.

CRedit authorship contribution statement

Malte Mittendorf: Conceptualization, Methodology, Software, Formal analysis, Investigation, Data curation, Validation, Visualization, Writing – original draft. **Ulrik Dam Nielsen:** Conceptualization, Methodology, Writing – review & editing, Supervision, Project administration, Resources, Funding acquisition. **Harry B. Bingham:** Conceptualization, Methodology, Writing – review & editing, Supervision, Resources, Project administration. **Gaute Storhaug:** Conceptualization, Data curation, Writing – review & editing, Resources.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that has been used is confidential.

Table 4

Metrics of predictions using the training dataset in both domains. The lowest error values are indicated in bold font.

		Frequency domain				Time domain			
		ResNet	Inception	MLSTM	CNN	ResNet	Inception	MLSTM	CNN
H_s	<i>RMSE</i>	0.413	0.425	0.649	1.033	1.353	0.883	2.568	1.196
	<i>MAE</i>	0.255	0.268	0.492	0.666	1.062	0.631	2.116	0.890
T_p	<i>RMSE</i>	0.393	0.381	0.506	0.893	1.022	1.002	1.653	1.276
	<i>MAE</i>	0.225	0.216	0.329	0.633	0.768	0.755	1.374	0.956
β	<i>RMSE</i>	5.106	4.910	5.486	10.84	13.49	10.20	31.99	21.38
	<i>MAE</i>	2.641	2.372	3.024	6.920	9.261	7.130	20.17	14.16

Table 5

Metrics of predictions using the training dataset in both domains using multi-task learning. The lowest error values are indicated in bold font.

		Frequency domain			Time domain		
		Inception	MTL	MTL+	Inception	MTL	MTL+
H_s	<i>RMSE</i>	0.425	0.341	0.324	0.883	0.945	0.672
	<i>MAE</i>	0.268	0.179	0.171	0.631	0.699	0.461
T_p	<i>RMSE</i>	0.381	0.356	0.371	1.002	0.845	0.639
	<i>MAE</i>	0.216	0.170	0.192	0.755	0.599	0.463
β	<i>RMSE</i>	4.910	5.031	5.082	10.20	12.75	9.022
	<i>MAE</i>	2.372	2.548	2.569	7.130	8.864	5.888
U_s	<i>RMSE</i>			0.167			0.378
	<i>MAE</i>			0.103			0.282

Acknowledgements

The authors highly acknowledge the data provision from DNV. In addition, we express our gratitude towards to Vemund Svanes Bertelsen (Miros) for sharing insights regarding the used wave radar as well as Daniel Schmode (Microsoft) for helpful discussions. The financial support from The Danish Maritime Fund (case number 2019-043), A/S D/S Orient's Fond and the Department of Civil and Mechanical Engineering (DTU) is highly appreciated. The second author has received funding by the Research Council of Norway through the Centres of Excellence scheme, project number 223254 AMOS. The calculations were performed using resources of the DTU computing center (DCC).

Appendix

Model performance on training data

See Table 4 for the results of the four initial models and Table 5 for data pertaining to the multi-task learning approach.

References

- [1] Meister S, Taylordean S, Singer D. Predicting extreme parametric roll in container ships. In: Proc. of the 14th int. symp. on practical design of ships and other floating structures (PRADS'19), Yokohama, Japan. Springer; 2019, p. 703–18.
- [2] Jalonen R, Tuominen R, Wahlström M. Remote and autonomous ships—the next steps: Safety and security in autonomous shipping—challenges for research and development. In: Rolls-Royce, Buckingham Gate, London: The advanced autonomous waterborne applications (AAWA). 2016, p. 56–73.
- [3] Nielsen UD. Sea state estimation based on measurements of wave-induced ship responses (Dr. Techn. thesis), DTU Mechanical Engineering, Kgs. Lyngby, Denmark; 2018.
- [4] Nielsen UD. A concise account of techniques available for shipboard sea state estimation. Ocean Eng 2017;129:352–62.
- [5] Takekuma K, Takahashi T. On the evaluation of sea spectra based on measured ship motions. Trans West-Jpn Soc Nav Archit 1973;45:51–7.
- [6] Iseki T, Ohtsu K. Bayesian estimation of directional wave spectra based on ship motions. Control Eng Pract 2000;8:215–9.
- [7] Nielsen UD. Estimation of on-site directional wave spectra from measured ship responses. Mar Struct 2006;19(1):33–69.
- [8] Pascoal R, Perera LP, Soares CG. Estimation of directional sea spectra from ship motions in sea trials. Ocean Eng 2017;132:126–37.
- [9] Tannuri EA, Sparano JV, Simos AN, Cruz JJD. Estimating directional wave spectrum based on stationary ship motion measurements. Appl Ocean Res 2003;25:243–61.
- [10] Nielsen UD, Stredulinsky DC. Sea state estimation from an advancing ship – A comparative study using sea trial data. Appl Ocean Res 2012;34:33–44.
- [11] Nielsen UD, Mounet REG, Brodtkorb AH. Tuning of transfer functions for analysis of wave-ship interactions. Mar Struct 2021;79:103029.
- [12] Mounet REG, Nielsen UD, Brodtkorb AH. Simultaneous sea state estimation and transfer function tuning using a network of dynamically positioned ships. Appl Ocean Res 2022. (Under review).
- [13] Mittendorf M, Papanikolaou AD. Hydrodynamic hull form optimization of fast catamarans using surrogate models. Ship Technol Res 2021;68(1):14–26. <http://dx.doi.org/10.1080/09377255.2020.1802165>.

- [14] Åvist P, Pyörre J. Modeling the impact of significant wave height and wave vector using an on-board attitude sensor network. In: Proc. of the 12th int. conf. on computer and IT applications in the maritime industries, Cortona; 2013.
- [15] Han P, Li G, Skjong S, Wu B, Zhang H. Data-driven sea state estimation for vessels using multi-domain features from motion responses. In: 2021 IEEE international conference on robotics and automation (ICRA). 2021, p. 2120–6. <http://dx.doi.org/10.1109/ICRA48506.2021.9561261>.
- [16] Goodfellow I, Bengio Y, Courville A. Deep learning. Cambridge: The MIT Press; 2016.
- [17] Cheng X, Li G, Skulstad R, Chen S, Hildre HP, Zhang H. Modeling and analysis of motion data from dynamically positioned vessels for sea state estimation. In: Proc. of 2019 int. conf. on robotics and automation (ICRA). IEEE; 2019, p. 6644–50.
- [18] Cheng X, Li G, Skulstad R, Zhang H, Chen S. SpectralSeaNet: Spectrogram and convolutional network-based sea state estimation. In: Proc. of IECON 2020 the 46th annual conference of the IEEE industrial electronics society; 2020.
- [19] Cheng X, Li G, Ellefsen AL, Chen S, Hildre HP, Zhang H. A novel densely connected convolutional neural network for sea-state estimation using ship motion data. IEEE Trans Instrum Meas 2020;69(9):5984–93.
- [20] Düz B, Mak B, Hageman R, Grasso N. Real time estimation of local wave characteristics from ship motions using artificial neural networks. In: Proc. 14th int. symp. practical design of ships and other floating structures (PRADS'19), Yokohama, Japan; 2019.
- [21] Kawai T, Kawamura Y, Okada T, Mitsuyuki T, Chen X. Sea state estimation using monitoring data by convolutional neural network. J Mar Sci Technol 2021;26:947–62.
- [22] Scholz TP, Mak B. Ship as a wave buoy – Estimating full directional wave spectra from in-service ship motion measurements using deep learning. In: Proc. of 39th int. conf. on ocean, offshore and arctic engineering (OMAE 2020), Fort Lauderdale, USA; 2020.
- [23] Han P, Li G, Skjong S, Zhang H. Directional wave spectrum estimation with ship motion responses using adversarial networks. Mar Struct 2022;83:103159.
- [24] Storhaug G, Moe E, Piedras Lopes TA. Whipping measurements onboard a midsize container vessel operating in the north atlantic. In: Marintec China proceedings (RINA, CMP, and SNAME). 2007.
- [25] Gangeskar R. An algorithm for estimation of wave height from shadowing in X-Band radar sea surface images. IEEE Trans Geosci Remote Sens 2014;52(6):3373–81.
- [26] Chen X, Huang W, Zhao C, Tian Y. Rain detection from X-Band marine radar images: A support vector machine-based approach. IEEE Trans Geosci Remote Sens 2020;58(3):2115–23.
- [27] Miros. Wavex[®] v5.6 - the x-band radar-based wave and current monitoring system. 2017.
- [28] Barstow S, Bidlot J, Caires S, Donelan M, Drennan W, Dupuis H, Graber H, Green J, Grønlie Ø, Guérin C, Gurgel K-W, Günther H, Hauser D, Hayes K, Hessner K, Hoja D, Icard D, Kahma K, Keller W, Wyatt L. Measuring and analysing the directional spectrum of ocean waves. 2005, COST 714.
- [29] Longuet-Higgins MS, Cartwright DE, Smith ND. Observations of the directional spectrum of sea waves using motions of a floating buoy. In: Ocean wave spectra. New York: Prentice Hall; 1963, p. 111–36.
- [30] Nielsen UD, Ikonomakis A. Wave conditions encountered by ships—A report from a larger shipping company based on ERA5. Ocean Eng 2021;237:109584.
- [31] Söding H. Global seaway statistics. Schriftenreihe Schiffbau, report no. 610, Hamburg: Technical University Hamburg-Harburg; 2001.
- [32] St. Denis M, Pierson WJ. On the motion of ships in confused seas. SNAME Transactions; 1953.
- [33] Mittendorf M, Nielsen UD, Bingham HB. Data-driven prediction of added-wave resistance on ships in oblique waves - a comparison between tree-based ensemble methods and artificial neural networks. J Appl Ocean Res 2022;118(102964).
- [34] Mittendorf M, Nielsen UD, Bingham HB. The prediction of sea state parameters by deep learning techniques using ship motion data. In: Proc. of 7th world maritime technology conference (WMTC'22), Copenhagen, Denmark; 2022.
- [35] Welch P. The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. IEEE Trans Audio Electroacoust 1967;15:70–3.
- [36] Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: Bach F, Blei D, editors. Proc. of the 32nd int. conf. on machine learning. Proceedings of machine learning research, vol. 37, Lille, France: PMLR; 2015, p. 448–56.
- [37] Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput 1997;9(8):1735–80.
- [38] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proc. 2016 IEEE conference on computer vision and pattern recognition (CVPR); 2016, p. 770–8.
- [39] Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan V, Rabinovich A. Going deeper with convolutions. In: Proc. of the IEEE conference on computer vision and pattern recognition; 2015, p. 1–9.
- [40] Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. 2012, NIPS 2012: 1106–1114.
- [41] Szegedy C, Ioffe S, Vanhoucke V, Alemi AA. Inception-v4, inception-resnet and the impact of residual connections on learning. In: Proc. of the 31st AAAI conference on artificial intelligence; 2017, p. 4278–84.
- [42] Ismail Fawaz H, Lucas B, Forestier G, Pelletier C, Schmidt DF, Weber J, Webb GI, Idoumghar L, Muller P-A, cois Petitjean F. InceptionTime: Finding AlexNet for time series classification. Data Min Knowl Disc 2020;34:1936–62.
- [43] Caruana R. Multi-task learning. Mach Learn 1997;28:41–75.
- [44] Ruder S. An overview of multi-task learning in deep neural networks. 2017, arXiv:1706.05098v1.
- [45] Caruana R, de Sa V. Promoting poor features to supervisors: Some inputs work better as outputs. In: Advances in neural information processing systems 9: proceedings of the 1996 conference, Vol. 9. 1997.
- [46] Kingma DP, Ba J. Adam: a method for stochastic optimization. In: Proc. of 3rd int. conf. on learning representations (ICLR '15), San Diego; 2015.
- [47] Mak B, Düz B. Ship as a wave buoy – estimating relative wave direction from in-service ship motion measurements using machine learning. In: Proc. 38th int. conf. on ocean, offshore & arctic engineering OMAE 2019, Glasgow, Scotland; 2019.
- [48] Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, Corrado GS, Davis A, Dean J, Devin M, Ghemawat S, Goodfellow I, Harp A, Irving G, Isard M, Jia Y, Jozefowicz R, Kaiser L, Kudlur M, Levenberg J, Mané D, Monga R, Moore S, Murray D, Olah C, Schuster M, Shlens J, Steiner B, Sutskever I, Talwar K, Tucker P, Vanhoucke V, Vasudevan V, Viégas F, Vinyals O, Warden P, Wattenberg M, Wicke M, Yu Y, Zheng X. TensorFlow: Large-scale machine learning on heterogeneous systems. 2015, URL: <https://www.tensorflow.org/>, software available from tensorflow.org.
- [49] Nielsen UD. Spatio-temporal variation in sea state parameters along virtual ship route paths. J Oper Oceanogr 2021.
- [50] Montazeri N, Nielsen UD, Jensen JJ. Selection of the optimum combination of responses for wave buoy analogy - an approach based on local sensitivity analysis. In: Proc. of 13th PRADS, Copenhagen, Denmark; 2016.
- [51] Andersen IMV, Storhaug G. Dynamic selection of ship responses for estimation of on-site directional wave spectra. In: Proc. 31st int. conf. on ocean, offshore and arctic engineering (OMAE'12), Rio de Janeiro, Brazil; 2012.
- [52] Brodtkorb AH, Nielsen UD, Sørensen AJ. Online wave estimation using vessel motion measurements. In: Proc. of 11th IFAC conf. CAMS, Opatija, Croatia; 2018.
- [53] Iseki T, Nielsen UD. Study on short-term variability of ship responses in waves. J Jpn Inst Navig 2015;132:51–7.
- [54] Takami T, Nielsen UD, Jensen JJ. Estimation of autocorrelation function and spectrum density of wave-induced responses using prolate spheroidal wave functions. J Mar Sci Technol 2020;26.
- [55] Han P, Li G, Cheng X, Skjong S, Zhang H. An uncertainty-aware hybrid approach for sea state estimation using ship motion responses. IEEE Trans Ind Inf 2022;18(2):891–900.
- [56] Mittendorf M, Nielsen UD, Bingham HB, Liu S. Towards the uncertainty quantification of semi-empirical formulas applied to the added resistance of ships in waves of arbitrary heading. Ocean Eng 2022;251(111040).

- [57] Gal Y, Ghahramani Z. Dropout as a Bayesian approximation: Representing model uncertainty in deep learning. In: Balcan MF, Weinberger KQ, editors. Proc. of the 33rd int. conf. on machine learning. Proceedings of machine learning research, vol. 48, New York, New York, USA: PMLR; 2016, p. 1050–9.
- [58] Bitner-Gregersen EM, Bhattacharya SK, Chatjigeorgiou IK, Eames I, Ellermann K, Ewans K, Hermanski G, Johnson MC, Ma N, Maisondieu C, Nilva A, Rychlik I, Waseda T. Recent developments of ocean environmental description with focus on uncertainties. *Ocean Eng* 2014;86:26–46.
- [59] 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. Era5 hourly data on single levels from 1979 to present. 2018, Copernicus Climate Change Service (C3S) Climate Data Store (CDS). (Accessed on 24-Feb-2022).