Physics-guided metamodel for vertical bending-induced fatigue damage monitoring in container vessels

Ocean Engineering 312 (2024) 119223

Contents lists available at ScienceDirect

Ocean Engineering

journal homepage: www.elsevier.com/locate/oceaneng

Research paper

Physics-guided metamodel for vertical bending-induced fatigue damage monitoring in container vessels

Xiao Lang [a], Mingyang Zhang [b,*], Chi Zhang [a], Jonas W. Ringsberg [a], Wengang Mao [a,*]

[a] Department of Mechanics and Maritime Sciences, Chalmers University of Technology, Sweden
[b] Department of Mechanical Engineering, Aalto University, Finland

ARTICLE INFO

Keywords:
Fatigue damage
Full-scale measurements
Container vessel
Physics-guided
Machine learning
Metamodel

ABSTRACT

This study proposes a novel physics-guided metamodel to predict vertical bending-induced fatigue damage in a 2800TEU container vessel navigating the North Atlantic, based on data from the vessel's hull monitoring system. The metamodel combines two XGBoost-based base learners: a black-box model utilizing ship heave and pitch motion measurements, and a gray-box model using spectral moments from numerical analysis. Predictions from both models are refined through a meta learner Gaussian process regression to enhance accuracy. The metamodel was evaluated against black-box and gray-box models across various training data volumes. The proposed model adapts to varying data volumes, from months to over 2 years, effectively integrating the strengths of both base learners to provide reliable predictions in both seen and unseen scenarios. The model consistently demonstrated superior performance, enhancing fatigue damage accumulation accuracy by up to 35% over traditional machine learning methods. This advancement can aid the maritime industry in effectively monitoring ship fatigue and implementing predictive maintenance strategies, marking a significant step forward in applying data-driven techniques in shipping.

1. Introduction

The accumulated fatigue damage in maritime vessels during sailing presents a significant challenge to their safety. Various methods for predicting fatigue have been developed and are continuously evolving to offer reliable estimates of the fatigue life for ships. The ship structures are subjected to fatigue loads that are random and uncertain; the development of fatigue prediction methods is mainly performed in two directions: using the fatigue accumulation method or random fatigue load modeling (Mao, 2010).

When structural stress data is readily available, fatigue damage evaluation can be efficiently conducted using the rainflow counting method, which adheres Palmgren-Miner's rule (Miner, 1945; Palmgren, 1924). However, obtaining time series data on stresses for ship fatigue assessment is challenging since very few ships are installed with the requisite sensors to measure structural responses, and these are usually installed only during short monitoring campaigns. Using theoretical methods to simulate stress time series across a broad spectrum of sea states requires substantial computational costs. As a result, fatigue prediction is often carried out in the frequency domain by assuming Gaussian processes, a process referred to as the narrow-band approach (Rychlik, 1987). Given that the real fatigue loads are not typically narrow-banded or Gaussian for ships, various methods have been developed to refine these approximations. Notably, Dirlik (1985), Zhao and Baker (1990), and Tovo (2002) explored simulations or models for stress range probability distributions. Winterstein (1985, 1988) also suggested models that include the 3rd and 4th stress spectral moments. Fig. 1 illustrates a DNV GL (2018) recommended approach to direct fatigue prediction under frequency domain, noting the uncertainties at each assessment stage.

These uncertainties persist across various components in current fatigue prediction methods, as evidenced by ongoing research (Friedman et al., 2000; Lang et al., 2021; ISSC, 2018; Dong et al., 2022). This includes calculations of wave loads and structural stresses (Li et al., 2014; Yosri et al., 2022; Yang et al., 2021), wave modeling (Olsen et al., 2006; Mao et al., 2010), and structural analysis (Yamamoto, 2017; Friedman et al., 2000; Thompson, 2018; Gaidai et al., 2020). Generally, these methodologies—which were founded based on physical principles—are known as white-box models (WBMs). Although these do not require actual sailing measurements, their accuracy is contingent upon

* Corresponding authors.
E-mail addresses: mingyang.0.zhang@aalto.fi (M. Zhang), wengang.mao@chalmers.se (W. Mao).

https://doi.org/10.1016/j.oceaneng.2024.119223
Received 29 May 2024; Received in revised form 7 August 2024; Accepted 7 September 2024
Available online 25 September 2024

the assumptions and simplifications employed in the physical modeling process.

With the booming of digitalization in the present maritime industry, an extensive amount of data has been measured to monitor ship performance, including stress signals pertinent to fatigue accumulation (Storhaug et al., 2007; Mao, 2014). Machine learning techniques have been extensively utilized in the analysis of fatigue and the prediction of structural lifespan (Lang et al., 2023; Masoudi Nejad et al., 2022; Bao et al., 2021; He et al., 2021; Yan et al., 2019). These pure data-driven approaches necessitate the use of substantial datasets and do not require physical knowledge; therefore, they are designated as black-box models (BBMs). However, as in the case of ship fatigue prediction, where ships are typically installed with sensors to measure strain/stress over relatively short periods (compared to lifecycle service), BBMs may yield significantly erroneous results for unseen scenarios. This limitation arises because their interpretation and extrapolation capabilities are poor. A third model category, known as the gray-box model (GBM), was delineated by Haranen et al. (2016). GBMs are developed by leveraging physical principles to guide data-driven methods with the aim of reducing the required volume of data while enhancing the accuracy of predictions and avoiding unreasonable results for unseen scenarios. While these methods have been applied to ship speed and motion predictions (Lang et al., 2024; Schirmann et al., 2023), their application in ship fatigue prediction remains unexplored. Notably, when an abundance of full-scale data is available, the predictive capability of BBMs can surpass that of GBMs.

To effectively monitor the fatigue damage of vessels throughout their operational lifecycle, it is imperative to establish a robust predictive model. This model should be adept at accurately forecasting fatigue damage based on data readily acquired in the future, such as ship motion and wave condition data collected under various navigational scenarios. Additionally, given the variability in the duration of structural stress measurement campaigns for different vessels, this model must maintain accuracy across scenarios ranging from limited data samples to those with abundant data. Furthermore, it should ensure that predictions remain reasonable under unseen conditions, thus preventing any unreasonable forecasts. This research proposes a pioneering physics-guided metamodel to predict fatigue damage induced by vertical bending in the midsection of a 2800 TEU container ship. Mao et al. (2010) have demonstrated that fatigue damage in the deck region between midships is governed by vertical bending across all operational profiles of the heading angle. The proposed metamodel synthesizes a BBM predicated on ship heave and pitch motion, with a GBM guided by numerical analysis. This fusion yields a model with enhanced robustness that is capable of effective predictions based on datasets of varying sizes, ranging from a few months to over two years. It capitalizes on the strength of the GBM to avoid improbable predictions in unseen scenarios while also retaining the BBM's ability to accurately capture more profound and intricate patterns and relationships when substantial full-scale data are available. The proposed method is demonstrated using a data set collected from the case study container vessel during three years of actual sailing. The following content is structured as: Section 2 presents the physical ship fatigue estimation methods. Section 3 introduces the method and architecture of the proposed physics-guided metamodel. The data acquisition and processing for model establishment are presented in Section 4. Section 5 compares the proposed metamodel to the conventional BBM and physics-guided GBM. Conclusions are drawn in Section 6.

2. Physics-based ship fatigue calculation approach

2.1. Ship fatigue and the S-N method

2.1.1. Measured stress responses

Ship fatigue is commonly assumed to be a high-cycle accumulation process, often addressed through the S-N approach. Fatigue damage can be quantified using the linear Palmgren-Miner's rule, which relies on designated S-N curves (IACS, 2006; DNV, 2018). This method is mathematically represented in Eq. (1):

$$d_{(t)}^{RFC} = \sum_i \frac{n_i S_i^m}{\alpha}, \tag{1}$$

where α and m are parameters of the S-N curve, t is the duration of a sea state. The quantity of stress cycle ranges S_i, represented by n_i, is approximated using the rainflow counting provided that stress data is accessible. The rainflow cycle was first defined by Matsuishi and Endo (1968). An alternative, non-recursive definition, provided by Rychlik (1987), was used in this study.

2.1.2. Numerical estimated stress responses

Acquiring ship stress signals for such analytical purposes is predominantly constrained. This is because a relatively small proportion of vessels are equipped with strain/stress measurement sensors, most typically installed during specific measurement campaigns. Furthermore, the duration of data collection by these sensors seldom extends

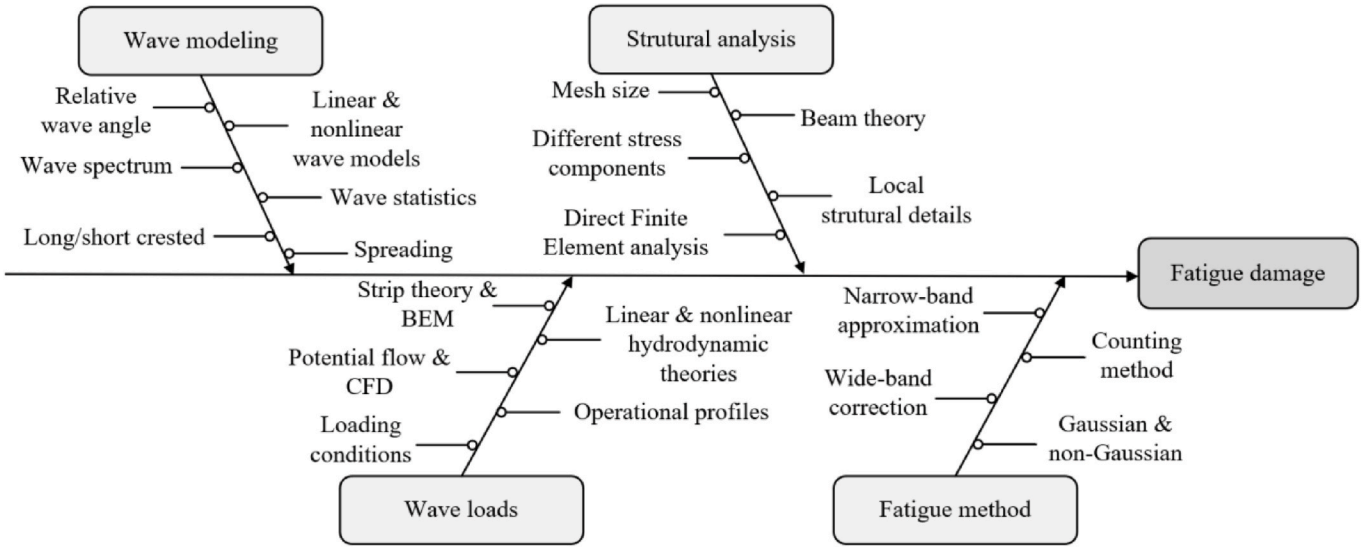

Fig. 1. Typical procedure for direct ship fatigue prediction as per the classification guidelines and inherent uncertainties in each assessment step.

beyond a few years, thus not encompassing the entire lifecycle of the vessel. Given that cyclic stresses induced by waves are a primary cause of ship fatigue, an alternative methodology for fatigue calculation can be employed when measured stress data are unavailable. In this approach, stress responses are numerically calculated in the frequency domain using Response Amplitude Operators (RAOs) in conjunction with wave conditions (Lang et al., 2021). Then, the fatigue damage can be estimated from spectral moments and the S-N curve, as depicted in Fig. 2.

2.2. Narrow-band fatigue approximation

RAOs are commonly derived through the assumption of a ship's rigid body movements for the analysis of wave loading, with beam theory being employed for stress analysis. The estimation of wave loads on ships is carried out using potential theory to predict ship fatigue, and is illustrated through a damped spring-mass system, as depicted in Eq. (2):

$$(\boldsymbol{M}+\boldsymbol{A})\ddot{\boldsymbol{Z}}+\boldsymbol{B}\dot{\boldsymbol{Z}}+\boldsymbol{C}\boldsymbol{Z}=\boldsymbol{F}_e, \tag{2}$$

where $\boldsymbol{Z}$ is the location in a ship's local coordinate system, $\dot{\boldsymbol{Z}}$ and $\ddot{\boldsymbol{Z}}$ represent the 6DoF ship motions' velocities and accelerations, respectively. $\boldsymbol{M}$ is the ship's mass matrix, and $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{C}$ denote the added mass, radiation damping coefficients, and hydrostatic stiffness.

Through the application of numerical analysis to solve Eq. (2), the RAOs for the vertical bending moment $M_v(\omega|V,\theta)$ within the frequency domain are derived. The RAOs are computed across different wave frequencies, denoted as ω, for a specific ship speed through water V and relative wave heading θ (0° represents the head sea), as follows:

$$H_v(\omega|V,\theta)=\frac{M_v(\omega|V,\theta)}{I_v}\Delta z, \tag{3}$$

where I_v defines the area moment of inertia and Δz is the distance between the detail and the neutral axis. The RAOs can be utilized in the estimation of the short-term response spectra pertaining to the vertical bending moment. For the purposes of fatigue prediction, long-term wave conditions are considered to consist of a series of short-term stationary sea states, each potentially lasting from 30-min to 6-h. These sea states are typically characterized by parameters such as significant wave height (H_s), wave period (T_z), and wave spectrum. Various wave spectra are utilized to model these conditions. The International Ship and Offshore Structures Congress (ISSC) recommended ISSC wave spectrum (Tucker, 1991) is adopted in this study and described in the following:

$$S(\omega|H_s,T_z)=\frac{4\pi^3H_s^2}{T_z^4\omega^5}\exp\left[-\frac{1}{\pi}\left(\frac{\omega T_z}{2\pi}\right)^{-4}\right]. \tag{4}$$

The ship stress response spectra are ultimately determined through the calculation of RAOs' squares in conjunction with the wave spectrum. This methodology facilitates the calculation of the stress spectral moments, denoted as λ_n for $n = 0, 1, 2, \ldots$, as follows:

$$\lambda_n(V,\theta,H_s,T_z)=\int_0^\infty\left|\omega+\frac{\omega^2V\cos\theta}{g}\right|^nH_v^2(\omega|V,\theta)S(\omega|H_s,T_z)d\omega. \tag{5}$$

If the stress signals for each sea state are approximated as a narrow-band process, their stress ranges adhere to the Rayleigh distribution. The so-called narrow-band approximation can further approximate fatigue damage accumulated during this period as follows:

$$d^{NB}=\frac{tf_z}{\alpha}2^{\frac{3m}{2}}\Gamma\left(1+\frac{m}{2}\right)\lambda_0^{m/2}, \text{ and } f_z=\frac{1}{2\pi}\sqrt{\frac{\lambda_2}{\lambda_0}}, \tag{6}$$

where the frequency of zero-crossing (referred to as f_z) and the gamma function denoted as $\Gamma()$ are the key components in the narrow-band approximation discussed above, the spectral moments are the sole variables in this approximation, and their estimation can be achieved using Eq. (5).

2.3. Bandwidth correction

The stress responses, however, are rarely strictly confined processes within a narrow-band, as shown in Fig. 3. The broadness of a stress signal can be defined as the ratio of the average period between peaks and the zero up-crossings frequency and peak frequency as $\varepsilon=\sqrt{1-\lambda_2^2/\lambda_0\lambda_4}$. If $\varepsilon = 0$, it indicates a narrow-band process, while $\varepsilon = 1$ means an extremely broad-band process. Narrow-band approximation in Eq. (6) is an upper bound limit of fatigue damage caused by a Gaussian process.

While stress responses may not adhere to Gaussian processes when ships navigate particularly harsh sea states, the assumption of a Gaussian response remains a widely adopted approximation for ship fatigue prediction. To enhance the accuracy of fatigue damage estimation under this assumption in Eq. (6), various correction methods have been developed and applied. This research employs a correction approach based on spectral moments as proposed by Wirsching and Light (1980) to assess its accuracy in fatigue damage estimation and identify valuable features for gray-box modeling. This bandwidth correction method uses only the bandwidth parameter ε_2, but in terms of the S-N curve slope parameter m:

$$d^{WL}=\left(a(m)+\left(1-a(m)\left(1-\sqrt{1-\varepsilon_2^2}\right)^{b(m)}\right)\right)d^{NB}, \tag{7}$$

where $a(m) = 0.926 - 0.033m$ and $b(m) = 1.587m - 2.323$.

3. Methodology

As previously introduced, physics-based WBMs for ship fatigue calculation are subject to various inherent uncertainties. This study

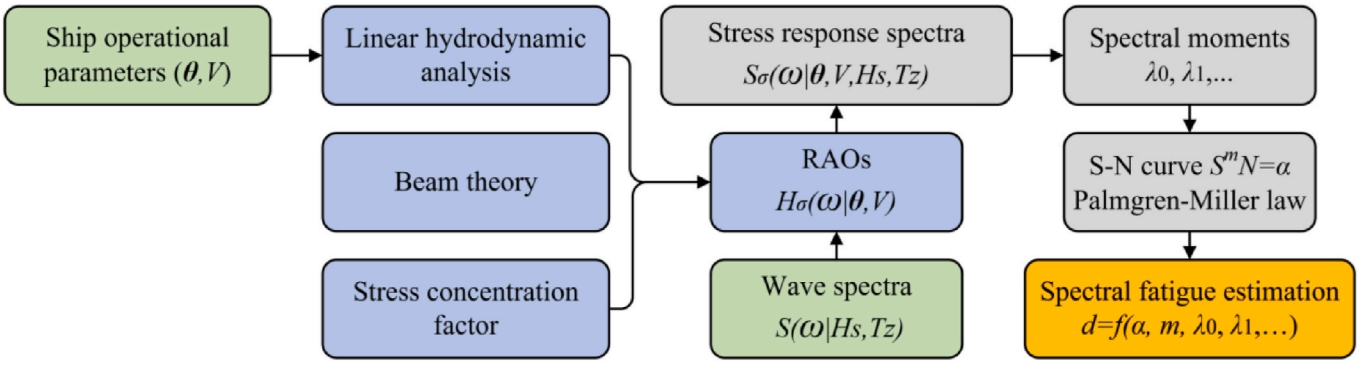

Fig. 2. The procedure for conducting direct ship fatigue calculations involves utilizing numerically estimated stress responses.

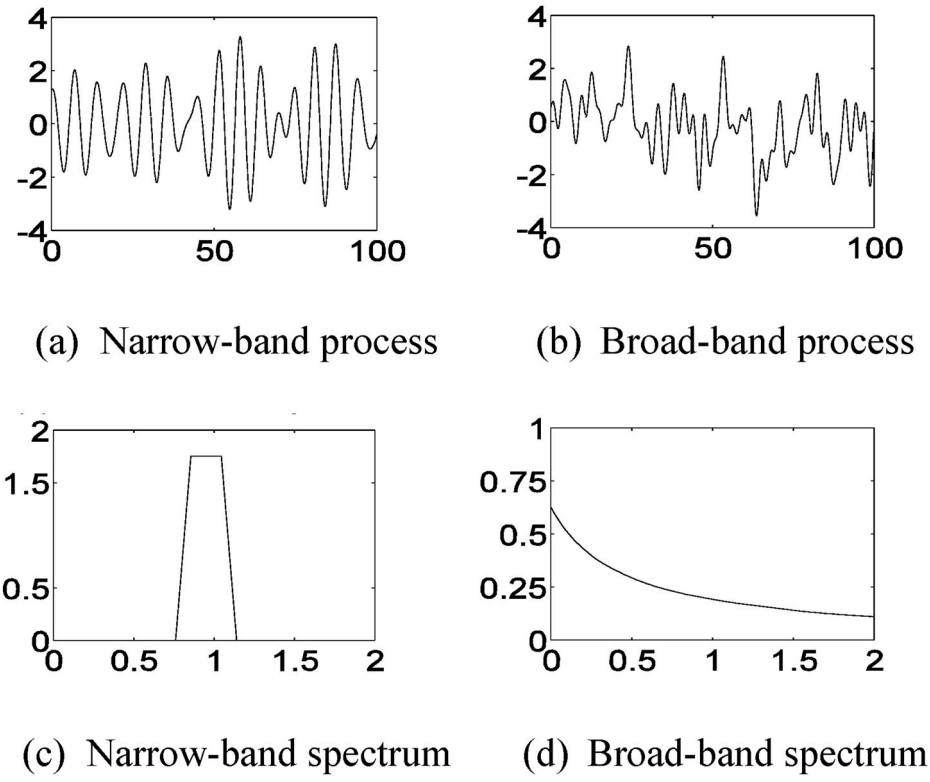

Fig. 3. Examples of narrow-band and broad-band processes.

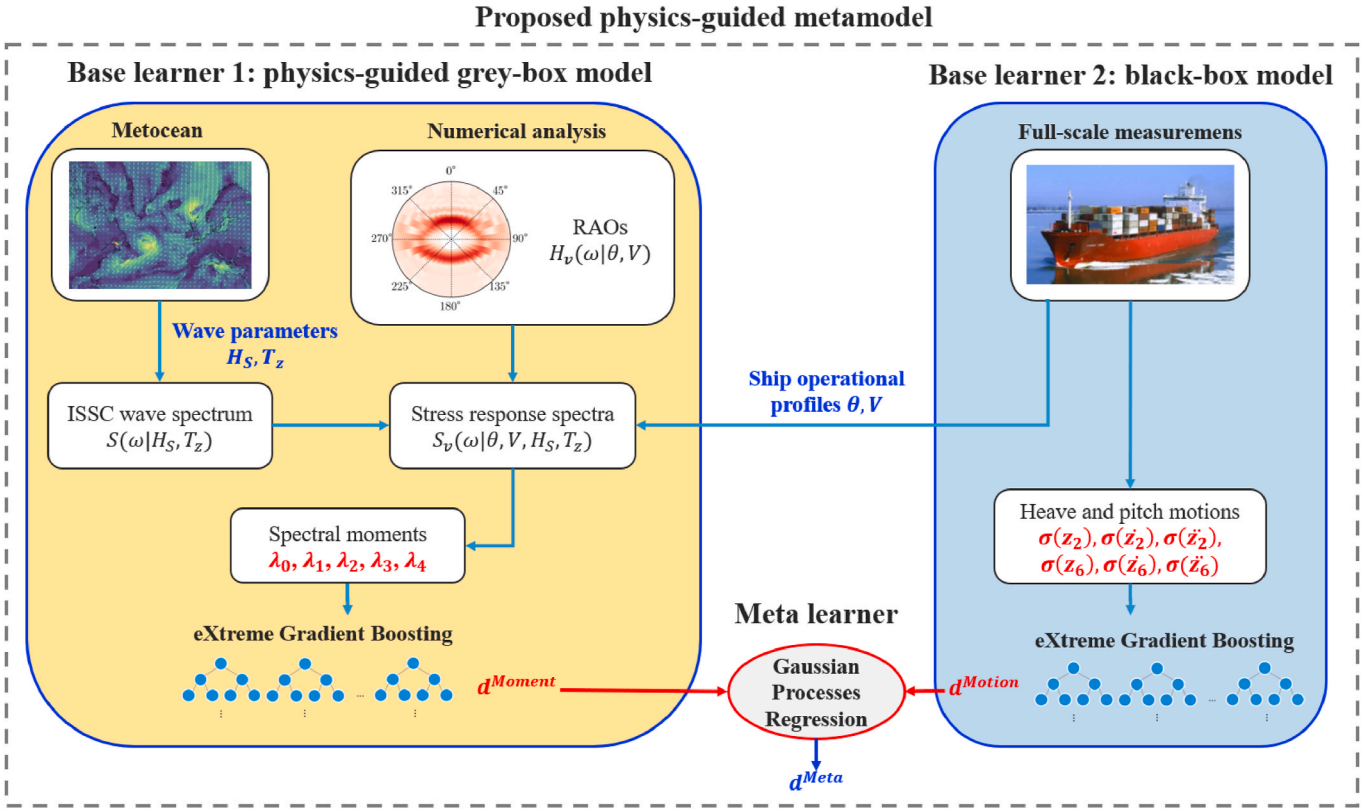

Fig. 4. Parallel modeling architecture of the proposed physics-guided metamodel for vertical bending-induced fatigue damage prediction.

seeks to combine a purely data-driven BBM with a numerical analysis-guided GBM, thereby establishing a metamodel that enhances the accuracy of predicting fatigue damage induced by vertical bending under various volumes of available data. In this section, we first delineate the architecture of the modeling framework. This is followed by an explanation of the machine learning algorithms implemented within the study, concluding with an in-depth discussion of the model's establishment process.

3.1. Modeling architecture

The proposed physics-guided metamodel employs a parallel modeling architecture, as illustrated in Fig. 4. This architecture configures two distinct base learner models to operate concurrently. The first base learner in this architecture is a physics-guided GBM that utilizes numerical analysis based on 2-D potential theory to obtain the RAOs $H_v(\omega|V,\theta)$ that characterize the case study vessel's response to vertical bending under different operational profiles. This is followed by a synthesis of the encountered metocean data, specifically significant wave height H_s, wave period T_z, and the vessel's operational data (including V, θ). This integration follows Eq. (5) and yields the stress response spectra, from which the spectral moments—denoted as $[\lambda_0, \lambda_1, \lambda_2, \lambda_3, \lambda_4]$—are derived by Eq. (6). These spectral moments $[\lambda_0, \lambda_1, \lambda_2, \lambda_3, \lambda_4]$ serve as inputs to XGBoost to approximate the rainflow counting fatigue damage d^{RFC}, with the forecasted outcome referred to as d^{Moment}.

The second base learner is a BBM that estimates fatigue damage by leveraging ship heave and pitch motion measurements. The selected input features include the standard deviations (σ) of heave (z_2), heave velocity ($\dot{z}_2$), and heave acceleration ($\ddot{z}_2$; denoted as $\sigma(z_2), \sigma(\dot{z}_2), \sigma(\ddot{z}_2)$), as well as the standard deviations of pitch (z_6), pitch velocity ($\dot{z}_6$), and pitch acceleration ($\ddot{z}_6$; represented as $\sigma(z_6), \sigma(\dot{z}_6), \sigma(\ddot{z}_6)$). These parameters were chosen due to their high feature importance for fatigue damage induced by vertical bending, as identified in previous research by Lang et al. (2023). Subsequently, $[\sigma(z_2),\sigma(\dot{z}_2),\sigma(\ddot{z}_2),\sigma(z_6),\sigma(\dot{z}_6),\sigma(\ddot{z}_6)]$ are input into the eXtreme Gradient Boosting (XGBoost) algorithm to estimate the rainflow counting fatigue damage d^{RFC}, with the resultant prediction denoted as d^{Motion}. Finally, the d^{Motion} value predicted by the BBM and the d^{Moment} value predicted by the GBM are fed into a Gaussian processes regression (GPR), which functions as the meta learner within the proposed metamodel. The GPR is tasked with predicting fatigue damage d^{RFC} again, and the predictive result is denoted as d^{Meta}. Because the second base learner is based on actual measured heave and pitch motions, it can capture nonlinearities present in the measured stress responses that arise from both wave and structural effects. Consequently, this compensates for the uncertainties that arise due to assumptions and simplifications (such as reliance on linear theory, wave spectrum difference, and the exclusion of wind and current loads) employed in the physical modeling process of the first base learner, the physics-guided GBM.

3.2. Machine learning algorithms

In the proposed metamodel, the black-box and gray-box models serve as the base learners, utilizing XGBoost to learn from original data and generate the foundational predictions d^{Motion} and d^{Moment}. The choice of XGBoost as the base learner is driven by its superior performance and efficiency in ship performance modeling, as demonstrated in prior work by Lang et al. (2022). Additionally, this selection maintains consistency with the authors' previous research on ship motion-based fatigue prediction using BBM (Lang et al., 2023). The meta learner (i.e., the GPR model) then synthesizes and refines these predictions. The XGBoost algorithm is a machine learning technique that uses numerous simple evaluators to create a more accurate prediction model through a process known as gradient tree-boosting. The main aim of XGBoost is to find out how different training features, represented as $\underline{\mathbf{X}}$, relate to the target variable, which in this study is the rainflow counting fatigue damage d^{RFC}. If the model includes K trees, the prediction for the i -th sample, $\widehat{d_i^{RFC}}$, is calculated by summing the outputs from all K trees as:

$$\widehat{d_i^{RFC}} = \sum_{k=1}^{K} f_k(\underline{\mathbf{X}}_i), i = 1, \ldots, n, \tag{8}$$

where n is the training samples' quantity, f_k represents the k-th tree, and the weight on the k-th tree's q-th leaf is denoted as $f_k(\underline{\mathbf{X}}_i) = \omega_{q(\underline{\mathbf{X}}_i)}$.

Traditional objective functions are limited in measuring a model's performance and do not account for its computational efficiency. In contrast, the objective function employed by XGBoost incorporates model complexity as a means to evaluate efficiency:

$$Obj = \sum_{i=1}^{n} l\left(d_i^{RFC}, \widehat{d_i^{RFC}}\right) + \sum_{k=1}^{K} \Omega(f_k), \tag{9}$$

where the first component of the objective function is the traditional loss function, measuring the difference between the measurement, d_i^{RFC}, and the prediction, $\widehat{d_i^{RFC}}$. The second component evaluates the model's complexity, focusing on the structure of the decision trees. This algorithm employs an additive training approach, meaning that each new tree is incorporated based on the outcomes of the preceding iteration. Therefore, during the γ-th iteration, the loss function is evaluated as $l\left(d_i^{RFC(\gamma)}, \widehat{d_i^{RFC}}^{(\gamma-1)} + f_\gamma(\underline{\mathbf{X}}_i)\right)$, where f_γ represents the tree added in the γ-th iteration. To simplify and optimize this iterative process, the algorithm uses a Taylor expansion of the loss function around the previous iteration's predictions as:

$$Obj^{(\gamma)} = \sum_{i=1}^{n} \left[g_i f_\gamma(\underline{\mathbf{X}}_i) + \frac{1}{2} h_i f_\gamma(\underline{\mathbf{X}}_i)^2\right] + \Omega\left(f_\gamma\right), \tag{10}$$

where g_i and h_i represent the 1st and 2nd derivatives of the loss function $l\left(d_i^{RFC(\gamma)}, \widehat{d_i^{RFC}}^{(\gamma-1)}\right)$ to $\widehat{d_i^{RFC}}^{(\gamma-1)}$, respectively. If a tree in the XGBoost model has δ leaves, each leaf can be identified by an index ε, then the weight assigned to each leaf is denoted as ω_ε. For a set of instances $I_\varepsilon = \{i|q(\underline{\mathbf{X}}_i) = \varepsilon\}$ that fall into leaf ε, the optimal weight ω_ε^* that minimizes the objective function $Obj^{(\gamma)}$ can be calculated as:

$$\omega_\varepsilon^* = -\frac{\sum_{i\in I_\varepsilon} g_i}{\sum_{i\in I_\varepsilon} h_i + \beta}, \tag{11}$$

where β serves as the regularization parameter that manages overfitting. Additional details regarding the hyperparameters are provided by Chen and Guestrin (2016).

For the ship motion-based BBM base learner F_{Motion}, its input feature is $\underline{\mathbf{X}}_{Motion} = [\sigma(z_2), \sigma(\dot{z}_2), \sigma(\ddot{z}_2), \sigma(z_6), \sigma(\dot{z}_6), \sigma(\ddot{z}_6)]$ and its output is d^{Motion}. For the numerical analysis-guided GBM base learner F_{Moment}, its input feature is $\underline{\mathbf{X}}_{Moment} = [\lambda_0,\lambda_1,\lambda_2,\lambda_3,\lambda_4]$, while its output is d^{Moment}. The meta learner GPR synthesizes the predictions obtained from two distinct base learners. GPR is a nonparametric modeling technique that incorporates a Gaussian process (GP) before conducting regression analysis, as Rasmussen (2004) described. This approach incorporates the regression residual and a prior GP determined through Bayesian inference. The GPR employs a mean function and a kernel or covariance function to provide the most probable prediction and a measure of uncertainty. The kernel function enables the GPR to understand and adapt to the structure within the base learners' output. For the new samples' predictions using the proposed metamodel, the GPR meta learner takes the BBM and GBM predictions ($\boldsymbol{d}_*^{Motion}$ and $\boldsymbol{d}_*^{Moment}$, respectively) for new samples and generates the final prediction $\boldsymbol{d}^{Meta}$, as follows:

$$\boldsymbol{d}^{Meta} = GP\left(U\left(\left[\boldsymbol{d}_*^{Motion}, \boldsymbol{d}_*^{Moment}\right]\right), RBF\left(\left[\boldsymbol{d}_*^{Motion}, \boldsymbol{d}_*^{Moment}\right], \left[\boldsymbol{d}_{Trained}^{Motion}, \boldsymbol{d}_{Trained}^{Moment}\right]\right)\right), \tag{12}$$

where GP represents the Gaussian process that encapsulates the distribution over the possible functions that fit the data. $RBF\left(\left[\boldsymbol{d}_*^{Motion}, \boldsymbol{d}_*^{Moment}\right], \left[\boldsymbol{d}_{Trained}^{Motion}, \boldsymbol{d}_{Trained}^{Moment}\right]\right)$ is the kernel function that evaluates the similarity between the new predictions from the base learners and the training data. The radial basis function (RBF) is employed here. $U\left(\left[\boldsymbol{d}_*^{Motion}, \boldsymbol{d}_*^{Moment}\right]\right)$ is the mean function, where $\boldsymbol{d}_*^{Motion} =$

$F_{Motion}\left(\underline{\mathbf{X}}_{Motion,*}\right)$ and $\boldsymbol{d}_{*}^{Moment} = F_{Moment}\left(\underline{\mathbf{X}}_{Moment,*}\right)$.

3.3. Model establishment

Fig. 5 outlines the workflow for establishing the metamodel. The dataset is initially pre-processed to prepare it for analysis. An independent set of voyages, designated as the unseen test set, is set aside from this pre-processed dataset. The remainder of the dataset is divided into a training set, which is used to build and refine the metamodel.

There are some hyperparameters in the XGBoost algorithm and GPR. Notably, a change in hyperparameter values can affect the performance of the constructed model. Since the hyperparameters interact, the optimum combination cannot be obtained by adjusting a specific hyperparameter individually. Table 1 outlines the hyperparameters under consideration and their tuning ranges for the XGBoost modeling. For GPR, two hyperparameters—i.e., *alpha* (scale mixture parameters) and *length_scale* (the length scale of the RBF kernel)—are tuned in the range of [0.01,100] and [1,15], respectively.

Bayesian optimization is employed in this study to identify the optimal hyperparameters for machine learning models. Bayesian optimization utilizes the Bayesian theorem to adapt hyperparameter data and employs surrogate models to ascertain their optimal values. This method, contrasting with traditional grid and random searches, leverages prior iteration results to tune superior hyperparameter configurations, effectively balancing search efficiency and performance while mitigating the risk of local optima.

This research adopts a 10-fold cross-validation to prevent overfitting and validate the reliability of the hyperparameter settings. The dataset is segmented into ten sequential batches, maintaining their original order without shuffling. The model is trained on nine of these batches, while the tenth batch evaluates its performance. The performance of the 10-fold cross-validation is quantified by the average of the evaluation metrics across all splits, with root mean square error (*RMSE*) selected as the primary metric for assessing cross-validation in this analysis.

4. Case study and full-scale measurement

4.1. Data acquisition

This study utilizes data collected from the hull monitoring system to predict vertical bending-induced fatigue damage in a 2800 TEU container vessel operating in the North Atlantic. The vessel in the case study has its main characteristics detailed in Table 2.

Table 1
Hyperparameters used in XGBoost modeling and their respective tuning ranges.

Parameter	Description	Tuning domain
Eta	Step size (learning rate)	[0.01, 1]
max_depth	Maximum depth of a tree	[3, 10]
n_estimators	Number of trees	[100, 5000]
Gamma	Minimum loss reduction required to make a split	[0, 5]
reg_alpha	L1 regularization term	[0, 100]
reg_lambda	L2 regularization term	[0, 100]
min_child_weight	Minimum sum of instance weight required in a child	[0, 10]
colsample_bytree	Subsample ratio	[0.5, 1]

The hull monitoring system on the vessel was installed in accordance with the DNV hull monitoring regulations (DNV, 2005). This system captures real-time data, including GPS coordinate, ship speed over ground and heading, and 6DoF ship motions. Four strain sensors are strategically placed at the case study container ship, two of which are positioned at the midsection, one on the port and the other on the starboard side of the stiffener web, to measure longitudinal strain. The strain gauges used are model WFLA 6–11, manufactured by Tokyo Measuring Instruments Laboratory Co., Ltd. (TML). Each gauge has a length of 6 mm, and a width of 2.2 mm. Fig. 6 shows the amidship strain sensor on the starboard side. This container ship was equipped with strain sensors because cracks were detected during its operation (before 8 years of trade). Consequently, DNV installed strain sensors near the

Table 2
Main characteristics of the 2800TEU container vessel operating in the North Atlantic.

Parameter	Symbol	Magnitude
Max. TEU	–	2800
Length between perpendicular	L_{pp}	232 [m]
Molded breadth	B	32.2 [m]
Molded depth	D	19.0 [m]
Design draft	T	10.78 [m]
Block coefficient	C_B	0.685
Deadweight	DWT	40900 [tons]
Service speed	$V_{service}$	21.3 [knots]
Building year	–	1998

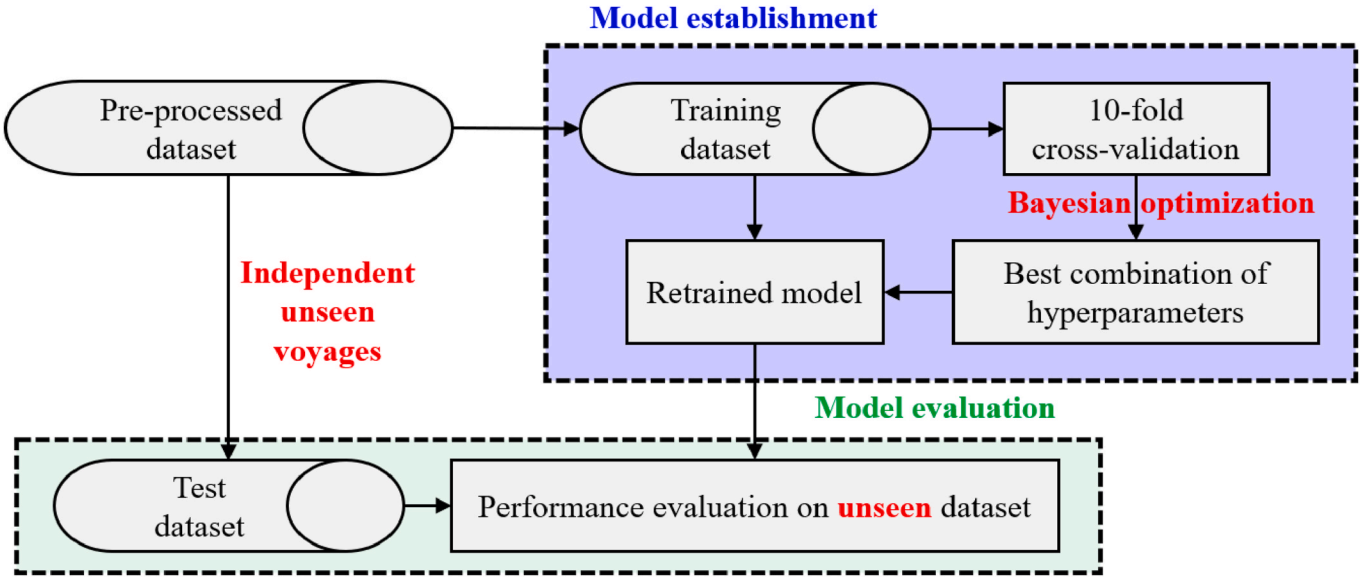

Fig. 5. Workflow for the establishment of the proposed metamodel.

crack site to monitor the strain, and the fatigue damage (stress) at the hot spot is obtained by applying an assigned stress concentration factor (SCF) of 2 recommended by DNV for this vessel (Mao, 2010). For other vessels, it may be necessary to conduct model experiments or FEM analysis to determine different SCFs, depending on the sensor's location and the ship's structure. For the fatigue analysis, only the stress induced by vertical bending at the midsection is considered, averaging the stress measurements from the strain sensors on both the port and starboard sides.

The considered container vessel is engaged in trade routes between Western Europe and Quebec, Canada. From the full-scale measurements, forty-eight routes spanning from 2007–09 to 2010–02 were chosen for inclusion in the case study. Voyages that lacked more than 50% of the required measurements were excluded from this study. The data collection frequency for strain/stress and ship motions is 25 Hz, and 1 Hz for other operational variables.

4.2. Data pre-processing and analysis

Fig. 7 illustrates all forty-eight case studies, categorized into winter and summer voyages. During winter, the container ship often adopts alternative routes that significantly deviate from the shortest voyages to avoid harsh wave conditions, as depicted in Fig. 7(a). The emphasis of this study is on the open sea; hence, data collected near coastal and shallow-water areas have been omitted. This is achieved by establishing two geographical boundaries at 55°W and 5°W longitudes (marked by red lines in Fig. 6) also help exclude anomalous strain/stress measurements from ports and coastal regions.

Fig. 8 showcases two examples of response stress (25 Hz) from voyages that commenced on 2008-03-01 and 2008-04-24. The blue frame highlights the span selected by the spatial boundaries, with the mean stress level normalized to 0. This setup ensures that the analysis includes only the strain/stress induced in open sea navigation, excluding strain/stress with less fluctuation during anchoring or when the vessel is near berths.

This study assumes stationary sea states have a fixed 30-min duration, following the assumption in prior work by Mao et al. (2010). Post data synchronization and the application of geographic filters for open sea conditions, the final dataset size for fatigue machine learning modeling comprises more than 10000 samples.

For the i-th 30-min sea state, accumulated fatigue damage, d_i^{RFC}, is calculated using the rainflow counting method. Fig. 9 showcases the normalized stress spectra ($\lambda_0 = 1$), employing the fast Fourier transform to convert the response into the power spectral density from the stress signals of the case study vessel's voyage commenced on 2008-01-06. Following the application of geographic boundaries, the voyage, which includes 277 sea states, is analyzed. The blue frame in Fig. 9(a) specifically marks the 150th sea state.

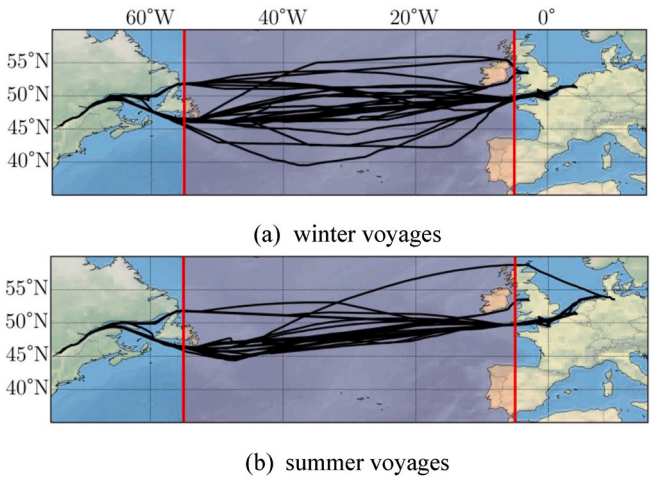

Fig. 7. The 2800TEU container ship's case study voyages are depicted for (a) winter and (b) summer navigation. The black lines illustrate the GPS positions recorded during these voyages, and the red frame highlights the routes chosen for analysis, filtered by the spatial boundaries set between 55°W and 5° W longitudes.

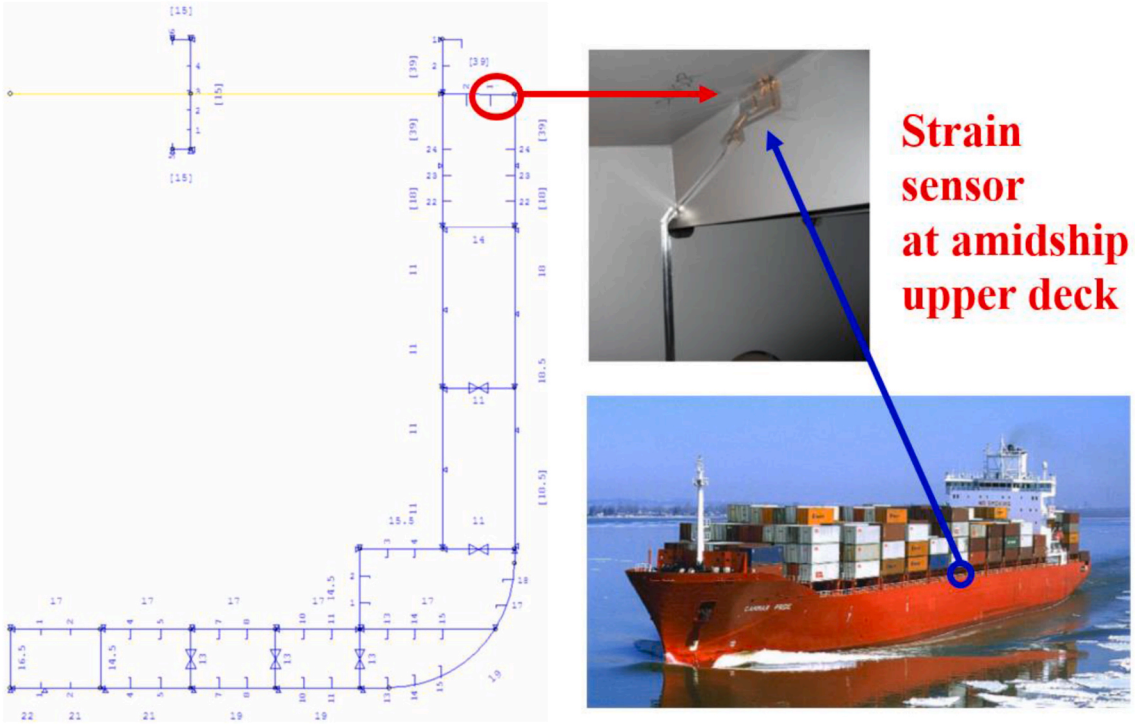

Fig. 6. The 2800TEU container ship utilized in this study, along with the arrangement of the strain sensor placement at the midsection, indicating the specific measurement point on the starboard upper deck.

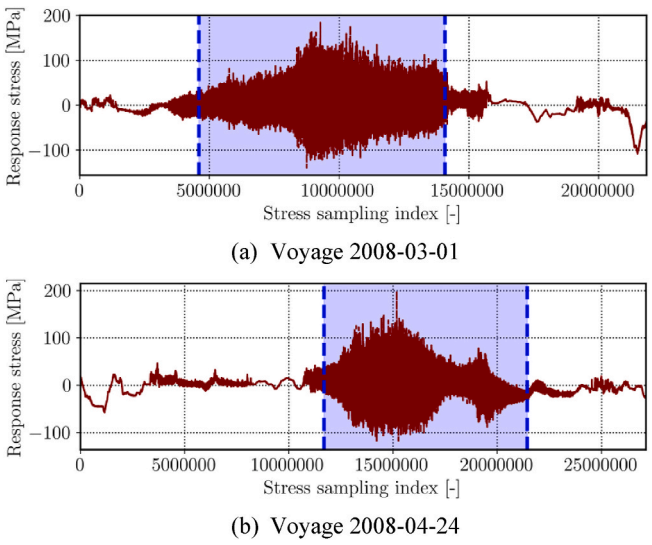

Fig. 8. Time series data depicting stress measured at midship during example voyages departing on (a) 2008-03-01 and (b) 2008-04-24. The blue frame delineates the stress span selected for the study, filtered according to the voyage's spatial boundaries. The mean stress level is normalized to 0.

In the analysis presented in Fig. 9(b), it is possible to observe the distinct multi-peak nature of the spectra across the 277 spectra that were examined. The initial crest in the spectra arose due to the wave frequency (WF) ship responses, while the subsequent crest was a result of whipping/springing. The third peak, generated by high-order ship responses, had a minor impact on fatigue forecasting. The spectral method applied in this study, when calculating RAOs, considers only wave-induced hydrodynamic loads (through numerical analysis); consequently, vibrations (such as whipping and springing) and high-order ship responses are not accounted for. Therefore, this paper focuses on fatigue damage induced by WF loads, which are extracted from the overall response within a frequency band of 0–3 rad/s, to ensure a fair comparison between the proposed metamodel and the conventional spectral method.

The metocean conditions encountered, including the mean wave direction (D_{wave}), significant wave height (H_s), mean wave period (T_z), were sourced hourly from the ERA5 reanalysis dataset, which offers a spatial resolution of 0.5° by 0.5° (Copernicus, 2019). Since the case study container ship onboard measurements do not include metocean data, employing spatial-temporal interpolation of the hindcast data is reasonable and aligns with practices in state-of-the-art research. Additionally, the current velocities ($U_{current}$ and $V_{current}$) were acquired from the Copernicus Marine Service, featuring a geographic resolution of 0.083° by 0.083° and a temporal resolution of 24 h (CMEMS, 2021). During the measurement campaign for the case study container ship, speed over ground was the only measured speed variable. Current velocities are therefore utilized to calculate the speed through water V (required for the spectral moments' calculation in Eq. (5)) based on the speed over ground, according to ISO (2015) guidelines.

The frequency distribution histograms of the final pre-processed data for the encountered significant wave height H_s, relative wave heading θ, and ship speed through water V are presented in Fig. 10. The statistical distribution of wave height is divided into winter and summer voyages (see Fig. 10(a)). As previously mentioned, the North Atlantic has the harshest navigation environment in winter. Even if this case study container vessel was equipped with an old conventional weather routing system to guide the ship to avoid severe environmental conditions, the waves encountered during winter sailings are significantly higher. In summer, ships rarely encountered waves with H_s values greater than 5 m. However, in winter, at least 25% of the samples had wave heights greater than 5 m, with the highest waves encountered being close to 10 m. Fig. 10(b) presents the distribution of relative wave angles in the case of westbound (Europe to North America) and eastbound (North America to Europe) sailing. In Fig. 10(b), heading angle $\theta = 0°$ means the head sea operation, while $\theta = 180°$ stands for the following sea. When sailing eastbound, the ship is mainly subjected to the following wave. When sailing westbound, the relative angle is concentrated at $0° \sim 50°$, and the waves come from the bow side of the ship. When sailing westbound in winter, the container vessel's speed through water V is much smaller than when sailing eastbound in summer due to larger waves and more head sea conditions, as shown in Fig. 10(c).

The encountered metocean environment and operational profiles have a wide distribution. Therefore, the dataset in the current investigation should be able to represent most of the operating conditions for the ship's entire service life.

4.3. Fatigue damage calculated by rainflow counting and spectral methods

The fatigue damage for each 30-min sea state across the forty-eight case study routes was calculated using the rainflow counting method, applied to stress signals induced by WF loads. For the case study ship, the Ib S-N curve characterized by $\alpha = 10^{12.76}$ and $m = 3$ is employed as per the Palmgren-Miner rule in Eq. (1) (DNV, 2005). Fig. 11 displays the fatigue damage d^{RFC} for each sea state in chronological order of the

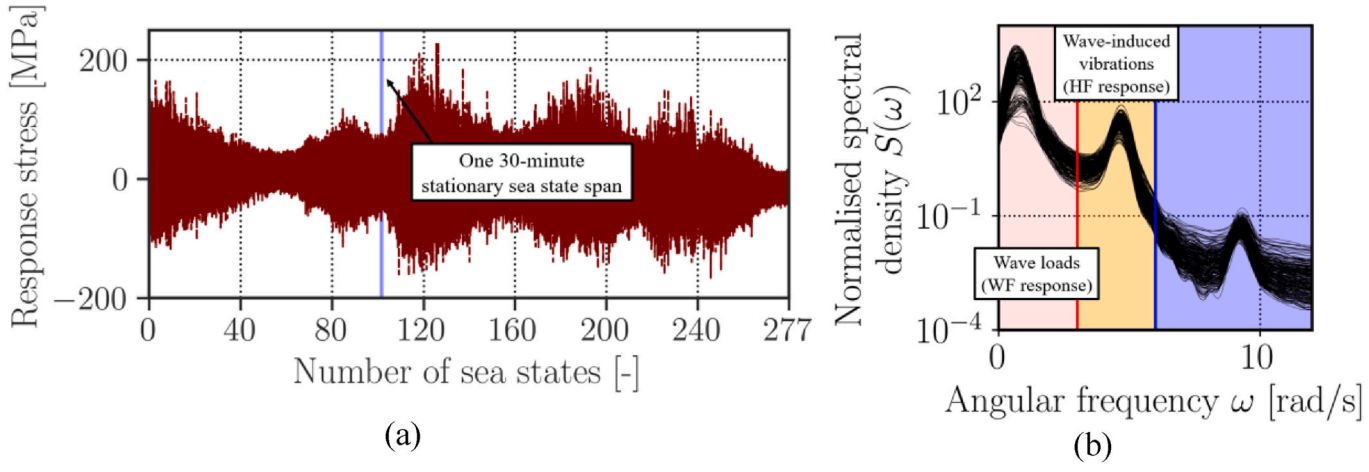

Fig. 9. (a) The first westbound winter voyage departing 2008-01-06 includes stress data for fatigue analysis filtered by geographic boundaries, encompassing 277 sea states. (b) From this voyage, 277 normalized spectra are derived from stress measurements taken during every 30-min sea state.

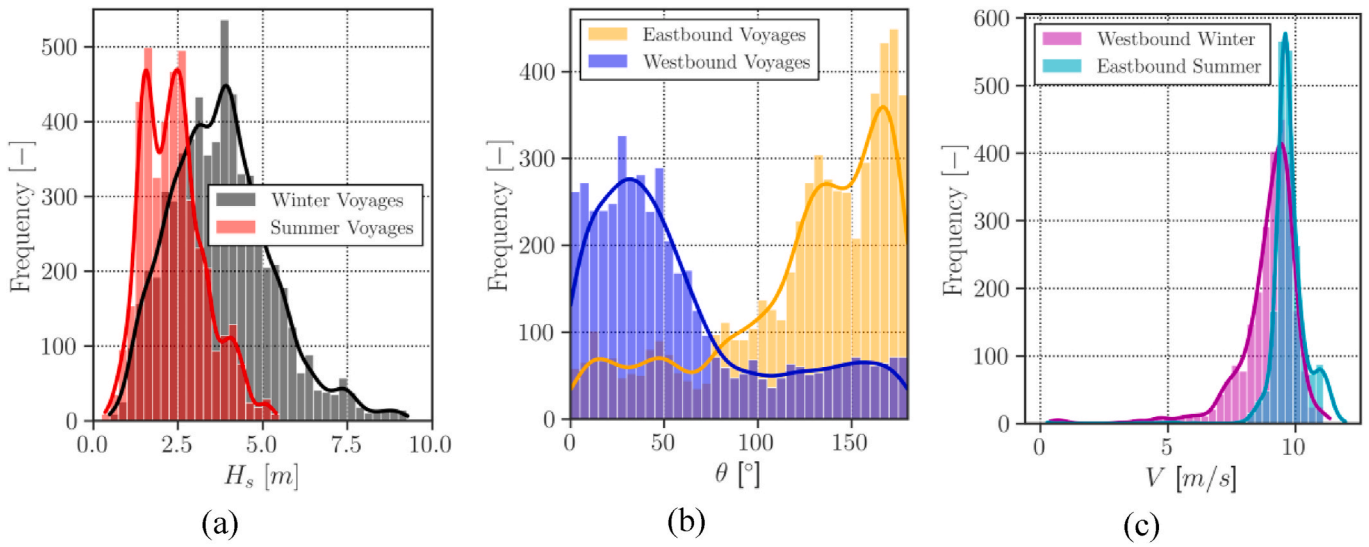

Fig. 10. Frequency distribution histograms of (a) encountered significant wave height H_s for winter and summer voyages, (b) relative wave angle θ for eastbound and westbound voyages, and (c) estimated ship speed through water V for westbound winter sailing and eastbound summer sailing, based on the processed data.

voyages, spanning from 2007–09 to 2010–02. Additionally, Fig. 12 illustrates the significant wave heights H_s encountered along the sailing route for each sea state. Orange frames in the figure denote westbound voyages, while blue frames indicate eastbound voyages.

As depicted in Fig. 12, the significant wave height H_s , as systematically quantified in Fig. 10(a), exhibits a clear seasonal pattern, with wave conditions becoming markedly more severe during winter. This seasonal variation in wave intensity significantly contributes to increased fatigue damage during winter voyages when compared to those undertaken in the summer, even when following the same sailing direction. Furthermore, the majority of the most substantial fatigue damage predominantly occurred during westbound voyages in winter. This is largely attributable to the relative wave angles for westbound navigation being comparatively closer to head sea conditions, as illustrated in Fig. 10(b).

For the calculation of ship fatigue damage based on numerically estimated stress responses, the RAOs for the vertical bending moments at the midship section of the upper deck are derived through a hydrodynamic analysis. This analysis uses the 2-dimensional potential theory-based Waveship software, considering the vessel's fully loaded condition (Mao et al., 2009, 2010). Fig. 13 displays the RAOs in polar heatmaps for three different speeds: $V = 1$ m/s (left), $V = 7$ m/s (middle), and $V = 12$ m/s (right).

In these figures, the angle represents the relative wave direction θ, and the radius indicates the wave frequency ω in rad/s ([0.2895, 1.63]). The heatmap colors denote the RAO values, with darker colors indicating higher values of $H_v(\omega|V, \theta)$, thus reflecting a greater stress response under the same significant wave height H_s and wave period T_z conditions. As shown in Fig. 13, the RAOs demonstrate sensitivity to changes in θ and V.

Based on the obtained RAOs, the fatigue damage for each stationary sea state has been predicted using the physical spectral methods introduced in Sections 2.2 and 2.3 for all 48 case study voyages. The narrow-band and bandwidth correction methods estimated fatigue damage and are denoted as d^{NB} and d^{WT}, respectively. Fig. 14 presents the discrepancies between the predictions of these physical models and the fatigue damage derived from rainflow counting for each sea state, represented as $d^{NB}-d^{RFC}$ and $d^{WL}-d^{RFC}$. Indeed, presenting dimensionless relative error would better characterize the deviations. However, for very small waves ($H_s < 1$ meter), d^{RFC} would also be small, resulting in a large relative error. It causes the data for larger wave heights to appear densely plotted and overlapping at the lower part of the figure, thus failing to accurately present the changes in deviations. Therefore, this study determines to present the fatigue damage prediction differences to ensure clarity and precision in illustrating the results. As shown in Fig. 14(a), the prediction error of the narrow-band method is relatively minor when H_s is less than 2 m. However, as H_s increases, the difference $d^{NB}-d^{RFC}$ shows a significant rise, indicating a nonlinear increase with larger waves. Fig. 14(b) illustrates the prediction of the Wirsching and Light method with bandwidth correction. Notably, the overestimation observed in the narrow-band method is mitigated by applying bandwidth correction. However, a significant nonlinear increasing trend in $d^{WL}-d^{RFC}$ with the growth of H_s remains evident. The prediction errors from both methods are attributed to uncertainties involved in the physical modeling (see Fig. 1). Specifically, the inability of linear

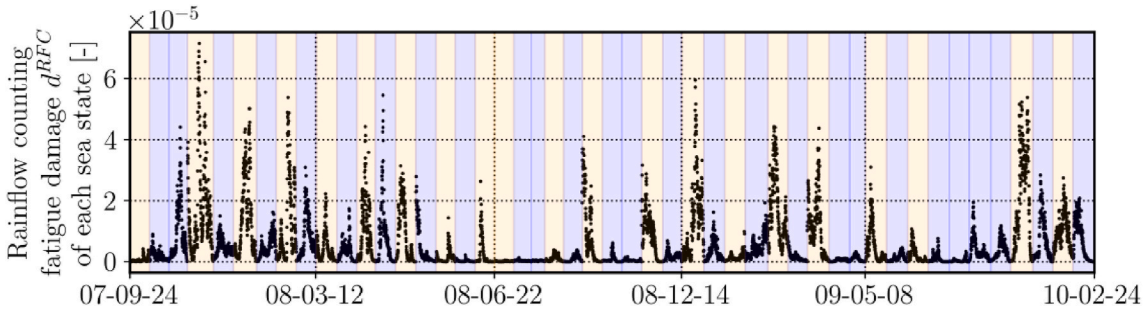

Fig. 11. Fatigue damage for every 30-min sea state across the forty-eight voyages, conducted from 2007–09 to 2010–02, was calculated using the rainflow counting method.

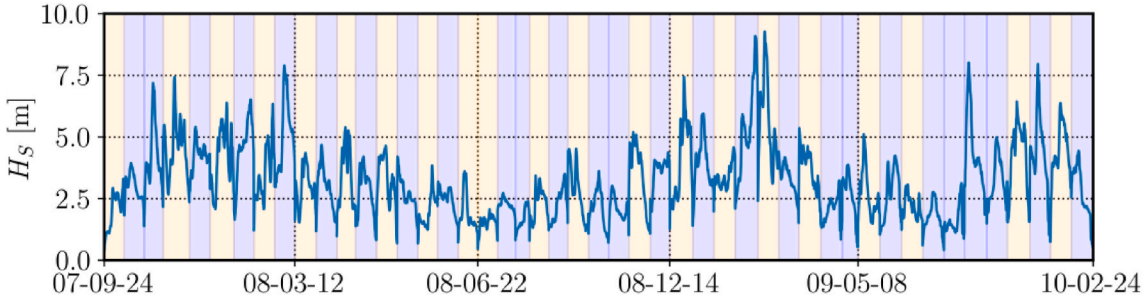

Fig. 12. Encountered significant wave height H_s of every 30-min sea state along the forty-eight voyages from 2007–09 to 2010–02.

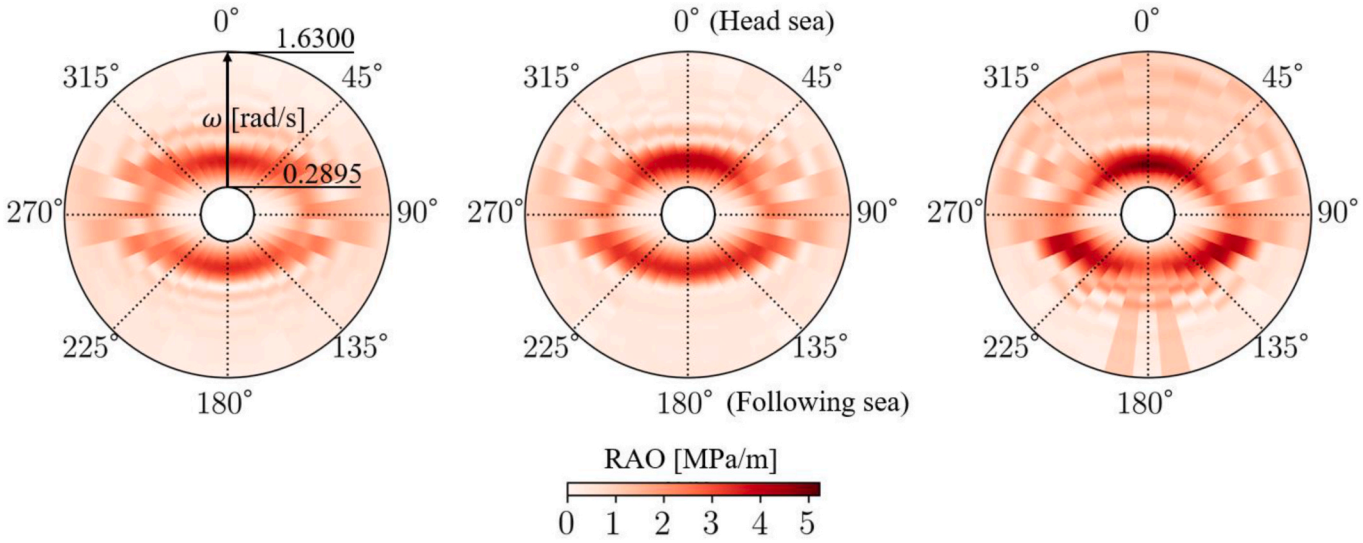

Fig. 13. Polar heatmaps of the RAOs ($H_v(\omega|V,\theta)$) for the longitudinal stiffener at the amidship deck of the case study container vessel are shown for speeds through water of $V = 1$ m/s (left), $V = 7$ m/s (middle), and $V = 12$ m/s (right).

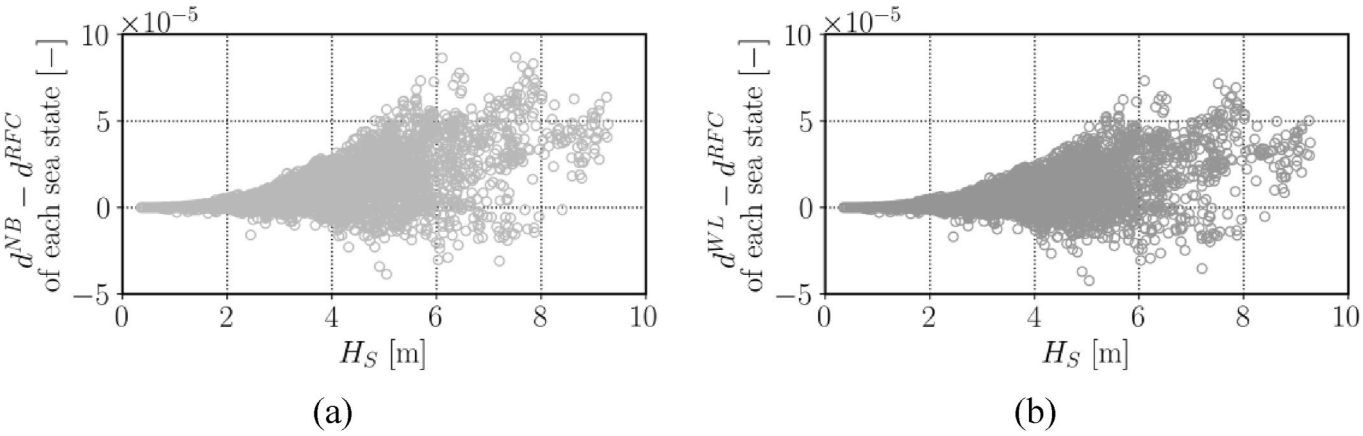

Fig. 14. Scatter plots of fatigue damage prediction differences with significant wave height H_S increases: (a) $d^{NB}-d^{RFC}$ and (b) $d^{WL}-d^{RFC}$.

hydrodynamic analysis and beam theory to accurately account for the nonlinear effects of increasing significant wave height on RAOs leads to these discrepancies. This overestimation is consistent with findings from previous studies by Mao et al. (2010) and Lang et al. (2023), which highlight physical models' limitations in capturing the complex dynamics of ship responses in high waves.

5. Results and discussion

This section compares the pure data-driven BBM based on ship motion with the physics-guided GBM informed by numerical analysis, as well as the metamodel proposed in Section 3. The physics-guided GBM (referred to as F_{Moment}, with predictions denoted as d^{Moment}) and the ship motion-based BBM (referred to as F_{Motion}, with predictions denoted as d^{Motion}) correspond to base learner 1 and base learner 2 in Section 3.1 and Fig. 4, respectively. The input features for these models are

numerical calculations of spectral moments $[\lambda_0, \lambda_1, \lambda_2, \lambda_3, \lambda_4]$ and standard deviations of variables related to heave and pitch motions, specifically $[\sigma(z_2), \sigma(\dot{z}_2), \sigma(\ddot{z}_2), \sigma(z_6), \sigma(\dot{z}_6), \sigma(\ddot{z}_6)]$. These two models are compared to the metamodel to test each base learner's capabilities, pros, and cons. Another physics-guided GBM is included in this comparison, referred to as $F_{Combine}$ (with predictions denoted as $d^{Combine}$). It remains an XGBoost-based model but combines the input features from both base learners, comprising $[\sigma(z_2),\sigma(\dot{z}_2),\sigma(\ddot{z}_2),\sigma(z_6),\sigma(\dot{z}_6),\sigma(\ddot{z}_6),\lambda_0,\lambda_1,\lambda_2,\lambda_3,\lambda_4]$. The comparison between the combined method and the meta method aims to determine whether a traditional machine learning model, using only the input features considered by the meta method, can achieve similar predictions. This comparison is crucial for validating the effectiveness and significance of the proposed metamodel.

The comparisons are conducted under three different data volume cases, as illustrated in Fig. 15: using 25% of the data for training to predict the remaining 75%; using 50% for training to predict the remaining 50%; using 75% for training to predict the remaining 25%. This analysis aims to determine the efficacy of each model under varying volumes of actual measured stress signal data. Such evaluations are critical in practical navigation, where ships monitor fatigue for future voyages—including potentially unseen voyages throughout their life cycle—following a short measurement campaign.

5.1. Black-box model

Firstly, for the pure data-driven BBM, the differences between the machine learning model prediction and the actual fatigue damage (i.e., $d^{Motion} - d^{RFC}$ for each sea state) are presented in Fig. 16 for three comparison cases, with the *RMSE* also noted. As shown in Fig. 16(a), when only 12 routes are used as training data, the BBM based on ship motion significantly overestimates fatigue damage in large waves ($H_s >$ 6 meters). This is because, in the training dataset consisting of 12 routes, the maximum significant wave height encountered by the case study container vessel is approximately 7.5 m, with most of the other training data being below 6 m. Therefore, it fails to accurately predict the more severe sea states of unseen voyages, as shown in Fig. 12. However, as the training data increases to 50%, the overestimation noticeably decreases (see Fig. 16(b)). When 75% of the data is used for training to predict the remaining 25%, the model's predictions become more stable, the *RMSE* significantly reduces, and the prediction differences mainly concentrate near zero even though the remaining 25% of the voyages include harsh wave conditions with H_s over 8 m.

In addition to comparing predictive capabilities across each sea state, monitoring fatigue damage accumulation over individual voyages is equally critical. For the three comparison cases, Cases 1–3, 36, 24, and 12 unseen voyages were used for evaluation. Fig. 17 presents the accumulated difference in fatigue damage between the BBM predictions and the rainflow counting calculation—denoted as $\Delta\left(d^{Motion} - d^{RFC}\right)$—along each unseen voyage. The x-axis represents the number of sea states of each voyage, while the color of the lines indicates the value of encountered significant wave height.

As illustrated in Fig. 17(a), in Case 1, for unseen harsh wave conditions, the BBM accumulates an error of up to 0.0035 on one route where the case study vessel encountered waves with H_s exceeding 9 m. Some of the other unseen routes also exhibited accumulated errors exceeding 0.001. However, with an increase in training data, the accumulated errors decrease in Case 2 (Fig. 17(b)), consistent with the trends observed in Fig. 16. For Case 3, regardless of whether the encountered wave heights exceed 8 m or not, the accumulated $\Delta\left(d^{Motion} - d^{RFC}\right)$ remains within a range of approximately ±0.0005.

5.2. Physics-guided gray-box model

For the physics-guided GBM, two models were compared: one using only spectral moments as input features and another considering both spectral moments and ship motions. Their predictions are denoted as d^{Moment} and $d^{Combine}$, respectively. The prediction errors of these models compared to the rainflow counting fatigue damage (i.e., $d^{Moment} - d^{RFC}$ and $d^{Combine} - d^{RFC}$ for each sea state across three different cases) are presented in Fig. 18.

For the spectral moments-guided GBM (see Fig. 18(a)), when the training data volume is only 25%, there is no significant overestimation—similar to the BBM results. However, noticeable prediction errors begin to appear on either side of zero when H_s exceeds 4 m; however, there is no trend of errors increasing with larger H_s. In contrast, for Cases 2 and 3, when the training data increases to 50% (Fig. 18(c)) and 75% (Fig. 18(d)), the improvement in prediction accuracy is not significant. Unlike the BBM, where an increase to 50% in training data slightly improves accuracy and a further increase to 75% significantly enhances it due to learning from sufficient data, the spectral moments-guided GBM does not show marked improvement in predictive capability with increased data volume, even though it has lower *RMSE* values than the BBM for Cases 1 and 2.

$F_{Combine}$, which considers spectral moments and ship motions (as depicted in Fig. 18(b) and (d)), exhibits a balanced approach. Its predictive capability shows some overestimation with small data volumes, though this overestimation is not as pronounced as the motion-based BBM because the spectral moments serve as guided additional features. Furthermore, as the data volume increases to 75% in Fig. 18(f), $F_{Combine}$ outperforms the pure spectral moments-based GBM, with $d^{Combine} - d^{RFC}$ more closely clustered around zero; however, its *RMSE* remains lower than that of the BBM.

The following analysis concerns the accumulated fatigue damage over individual voyages for the two GBMs. Fig. 19 presents the accumulated differences in fatigue damage between the predictions from the spectral moments-based GBM and the rainflow counting calculations, denoted as $\Delta\left(d^{Moment} - d^{RFC}\right)$. The results confirm the observation in Fig. 18, showing that the maximum accumulated error remains within ±0.001, without distinct overestimation regardless of the data volume. However, the accumulated errors do not decrease with increased training data. Unlike the BBM results in Fig. 17, where several voyages exhibited significant overestimation and most had more minor accumulated errors, the spectral moments-based GBM displays diverging accumulated errors.

The accumulated error for the GBM that incorporates both spectral moments and ship motions, $\Delta\left(d^{Combine} - d^{RFC}\right)$, is presented in Fig. 20. As demonstrated in Fig. 18, the inclusion of additional ship motions as

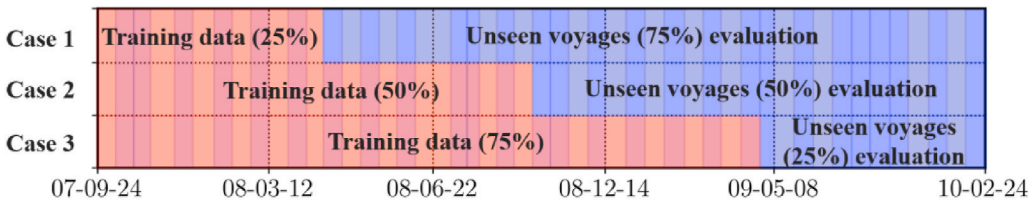

Fig. 15. Three comparison cases in this study: Case 1 utilizes the first 25% of data, Case 2 uses 50%, and Case 3 uses 75% as training data to monitor vertical bending-induced fatigue damage in the remaining unseen voyages.

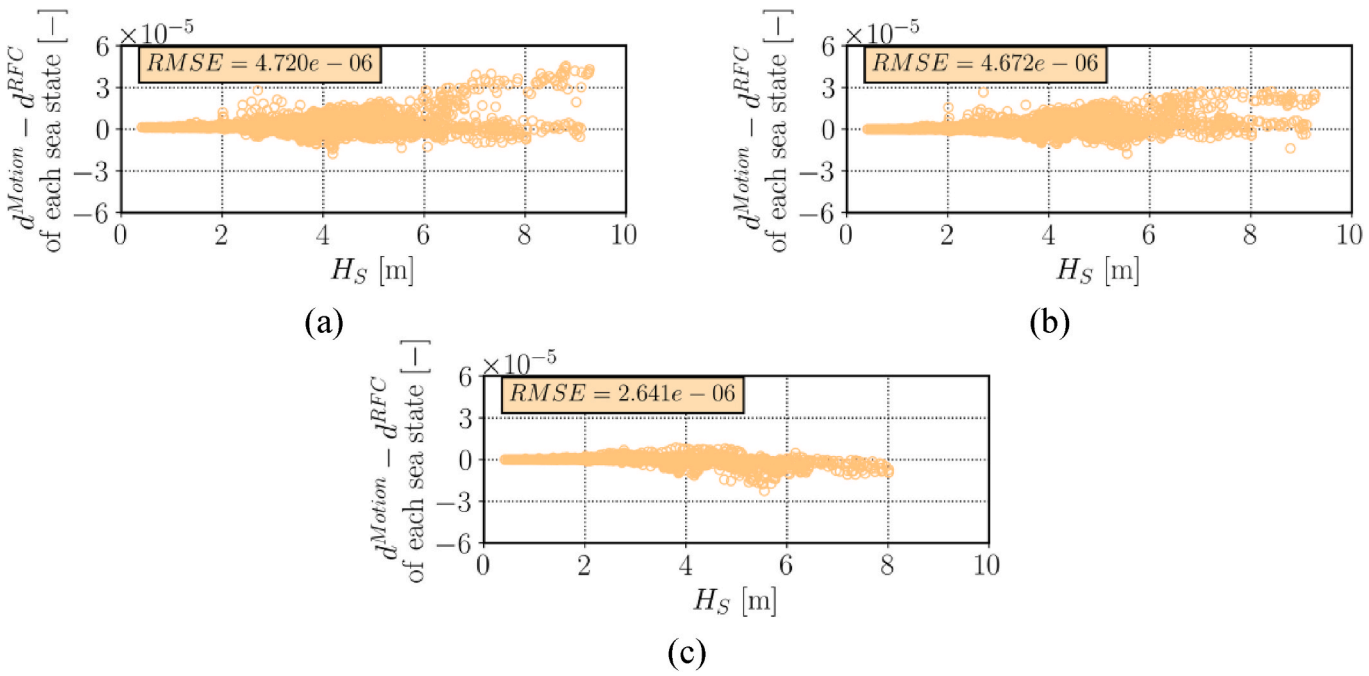

Fig. 16. Prediction error from the pure data-driven BBM based on ship motion versus the rainflow counting fatigue damage $d^{Motion} - d^{RFC}$ across each sea state for three cases: (a) Case 1, (b) Case 2, and (c) Case 3.

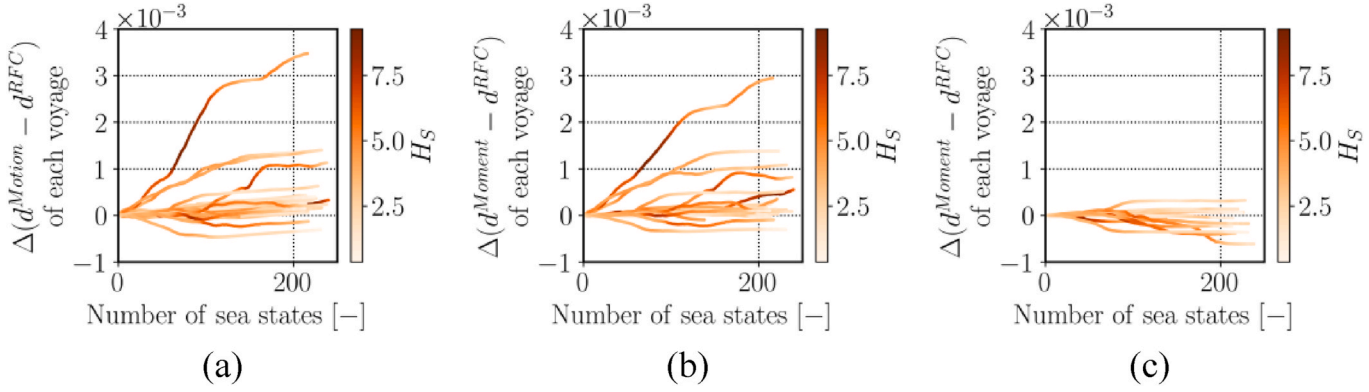

Fig. 17. Accumulated prediction error from the pure data-driven BBM based on ship motion versus the rainflow counting fatigue damage $\Delta\left(d^{Motion} - d^{RFC}\right)$ across each unseen voyage for three cases: (a) Case 1, (b) Case 2, and (c) Case 3.

input features leads to overestimation under harsh wave conditions. However, as the data volume increases, the predictive capability of this model improves, aligning with the observed reduction in prediction error. This trend indicates that while initial models may overestimate, their accuracy increases with additional data, as per the BBM prediction results.

The comparison between the pure data-driven BBM and the physics-guided GBMs reveals that spectral moments obtained from numerical analysis can rapidly capture the main variations in the data when the data volume is limited. This allows the model to rapidly achieve high accuracy with minimal overestimation. However, as the data volume increases, the improvement in model performance shows diminishing returns because these features cannot capture more complex relationships that may exist within the data. This may be explained by the inherent uncertainties in calculating spectral moments during numerical analysis, which limit their ability to capture complex patterns in the data. As the data volume increases, the representativeness and diversity of the data also increase. When used as input features, ship motions initially do not exhibit their full predictive potential due to insufficient data. As the data volume increases, these features can reveal more profound and complex patterns and relationships within the data, leading to significant improvements in model performance and eventually surpassing the performance of models based on spectral moments alone.

However, the GBM $F_{Combine}$, which simply combines spectral moments and ship motions as input features, serves as a compromise between the two models. With a small data volume, there is significant overestimation, though not as extensive as with the BBM. With a large data volume, the predictive capability improves significantly but does not reach the level of accuracy achieved by the BBM. This result illustrates that simply combining features can mitigate some of the issues observed in single-feature models but does not always produce optimal results—especially in complex predictive scenarios.

5.3. Physics-guided metamodel

The prediction errors between the proposed physics-guided metamodel and the rainflow counting fatigue damage, $d^{Meta} - d^{RFC}$, are pre-

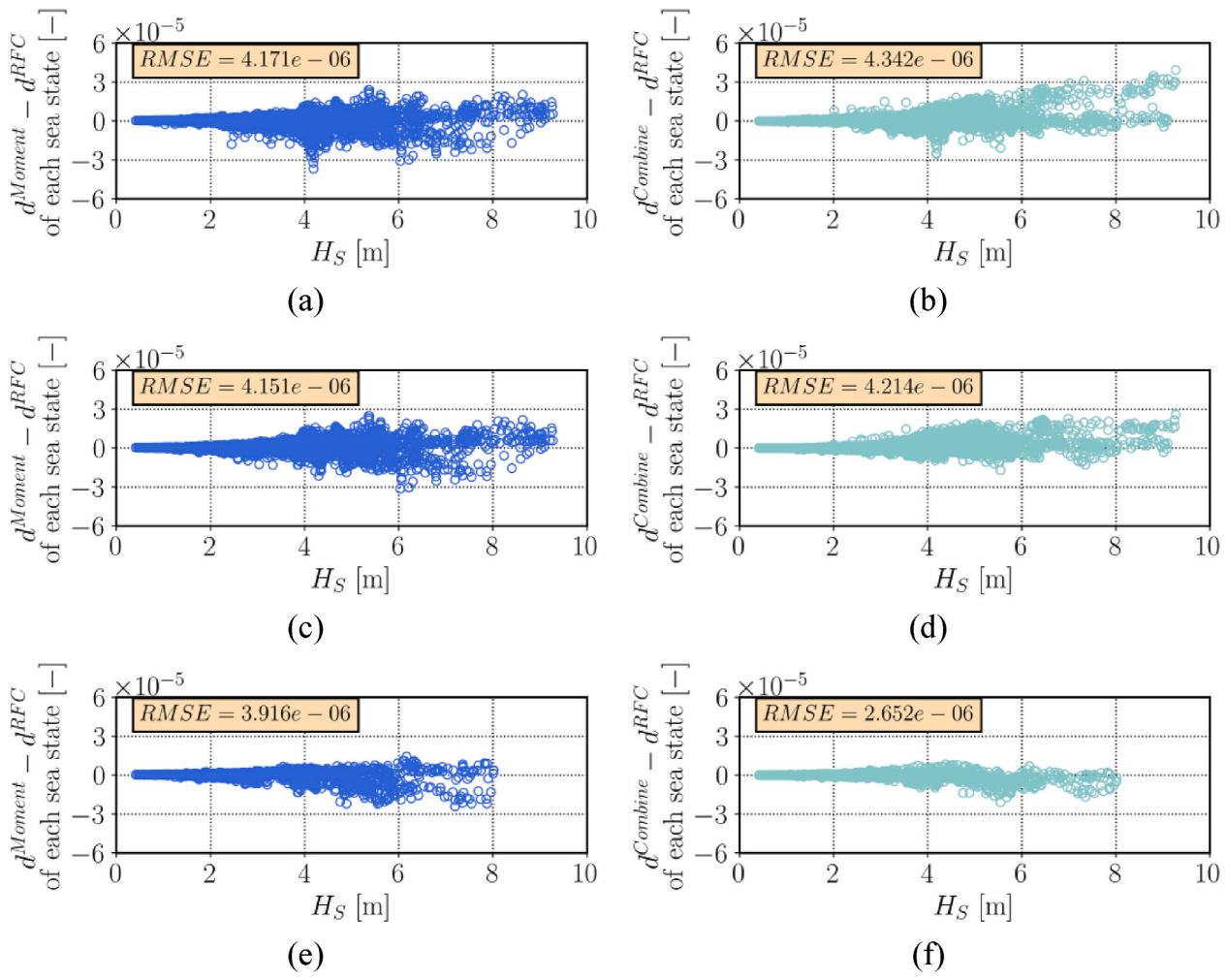

Fig. 18. Prediction error from the physics-guided GBMs versus the rainfall counting fatigue damage across each sea state for three cases: (a) $d^{Moment} - d^{RFC}$ of Case 1; (b) $d^{Combine} - d^{RFC}$ of Case 1; (c) $d^{Moment} - d^{RFC}$ of Case 2; (d) $d^{Combine} - d^{RFC}$ of Case 2; (e) $d^{Moment} - d^{RFC}$ of Case 3; (f) $d^{Combine} - d^{RFC}$ of Case 3.

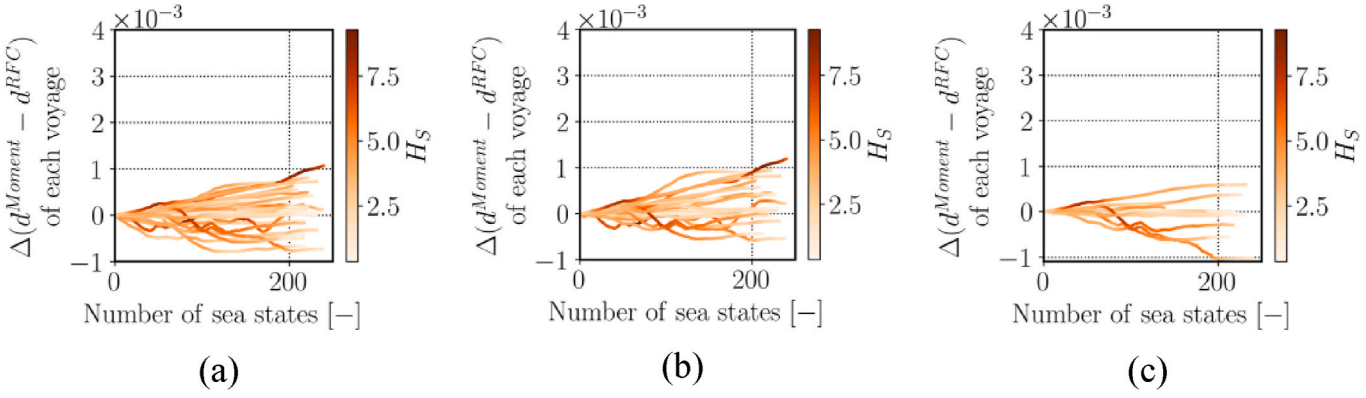

Fig. 19. Accumulated prediction error from the physics-guided GBM based on spectral moments versus the rainfall counting fatigue damage $\Delta\left(d^{Moment} - d^{RFC}\right)$ across each unseen voyage for three cases: (a) Case 1, (b) Case 2, and (c) Case 3.

sented in Fig. 21 for all three evaluation cases. Overall, the metamodel achieved the best predictive results when compared to the BBM and GBMs, with the lowest *RMSE* across the three different data volume cases.

Although only 25 and 50% of the data were used for training in Cases 1 (Fig. 21(a)) and 2 (Fig. 21(b)), respectively, the metamodel did not exhibit significant overestimation for large waves. Moreover, as the data volume increased to 75% (see Fig. 21(c)), the predictive capability of the metamodel improved rapidly, surpassing that of the pure data-driven BBM. The proposed model effectively learns from the strengths of both the BBM and the spectral moments-guided GBM. It rapidly improves the prediction accuracy in scenarios with limited data, thereby avoiding unreasonable predictions. With larger data volumes, the model can learn deeper and more complex patterns within the data, thus demonstrating

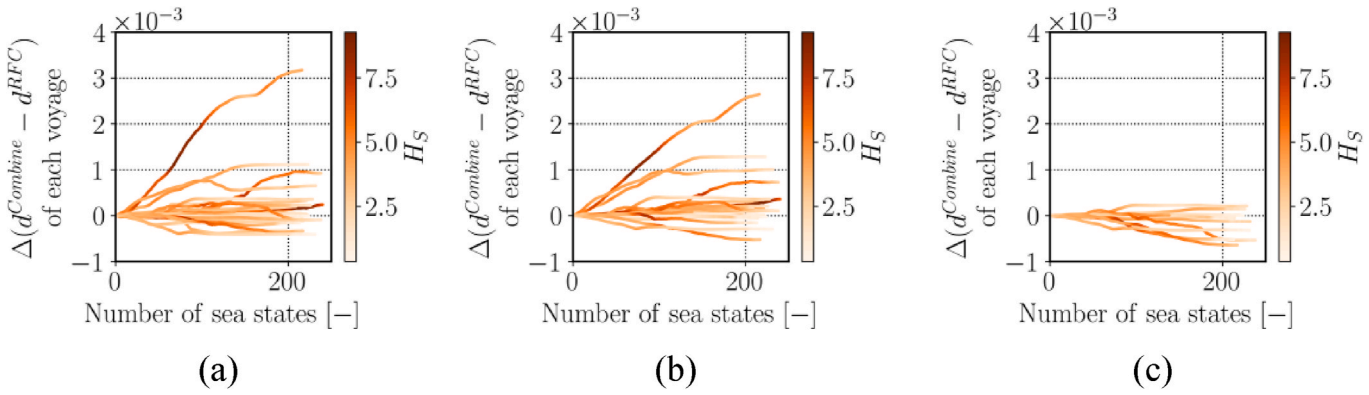

Fig. 20. Accumulated prediction error from the physics-guided GBM based on spectral moments and ship motions versus the rainfall counting fatigue damage $\Delta\left(d^{Combine} - d^{RFC}\right)$ across each unseen voyage for three cases: (a) Case 1, (b) Case 2, and (c) Case 3.

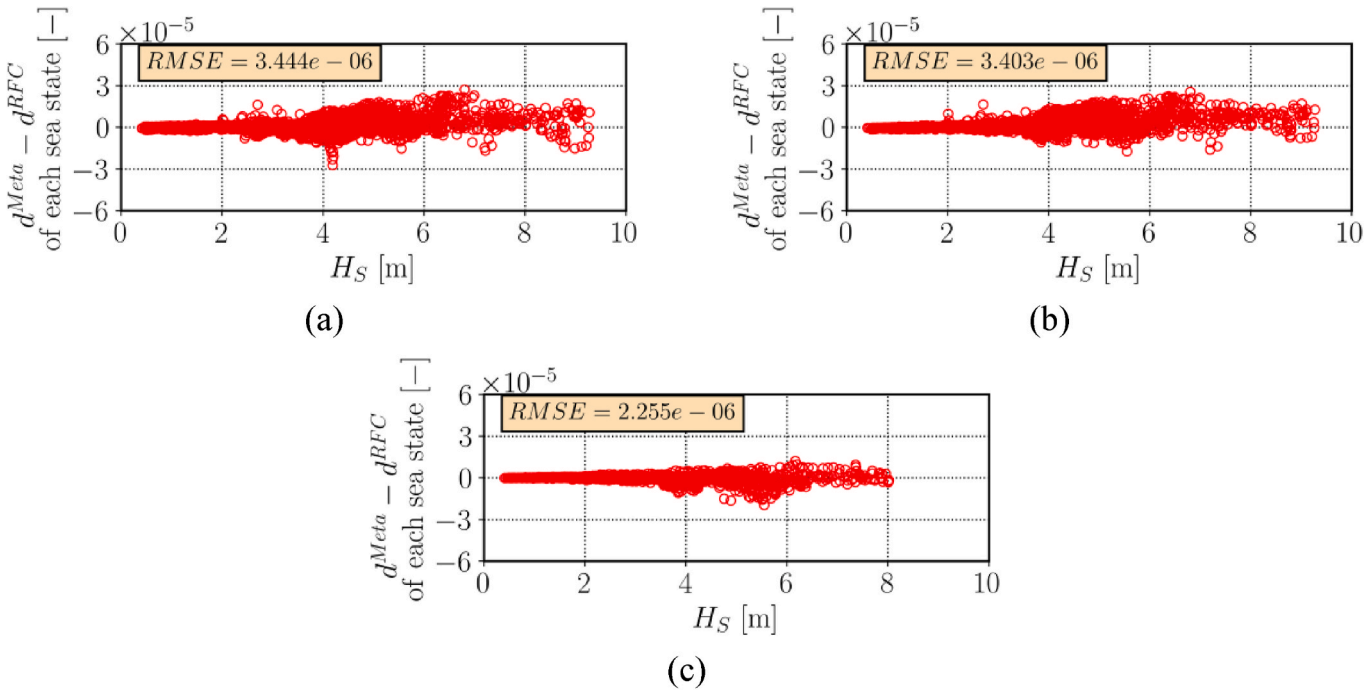

Fig. 21. Prediction error from the physics-guided metamodel versus the rainfall counting fatigue damage $d^{Meta} - d^{RFC}$ across each sea state for three cases: (a) Case 1, (b) Case 2, and (c) Case 3.

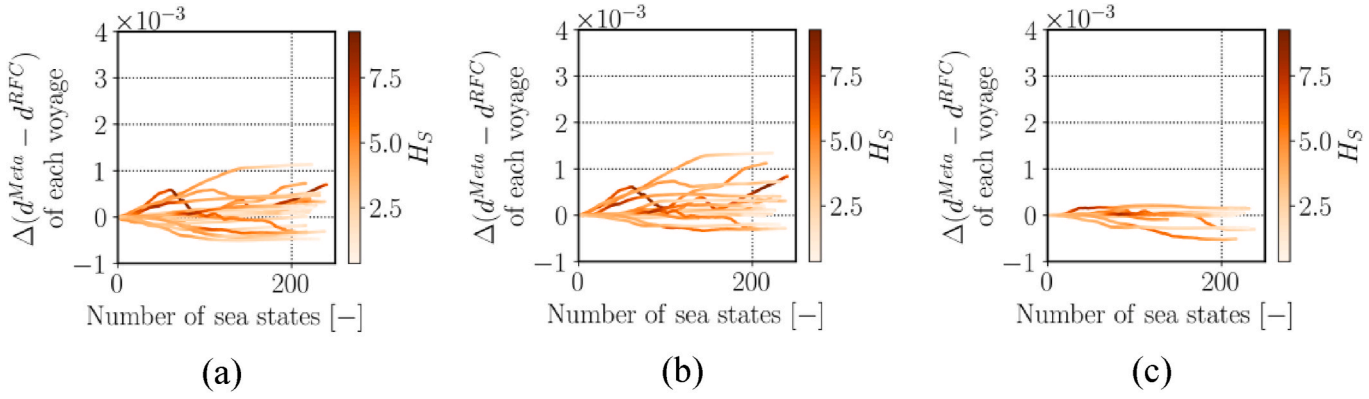

Fig. 22. Accumulated prediction error from the physics-guided metamodel versus the rainfall counting fatigue damage $\Delta\left(d^{Meta} - d^{RFC}\right)$ across each unseen voyage for three cases: (a) Case 1, (b) Case 2, and (c) Case 3.

its robustness and adaptability across different training scenarios.

The accumulated prediction error $\Delta\left(d^{Meta} - d^{RFC}\right)$ across each unseen voyage for the three cases is illustrated in Fig. 22. As depicted in Fig. 22(a), when the data volume is small, the predictive capability of the metamodel is similar to that of F^{Moment}, with no excessively large accumulated errors. Moreover, the distribution range of accumulated errors is even smaller; except for one voyage where the accumulated error is about 0.001, most are within ±0.0005. This demonstrates the metamodel's precision, even with limited data. As the training data increases to 50%, the results remain consistent (see Fig. 22(b)), with no significant overestimation like that observed with F^{Motion} and $F^{Combine}$.

When the data volume is increased to 75%, the distribution of accumulated errors becomes very narrow (see Fig. 22(c)), with the errors of nine out of twelve unseen voyages being close to zero. This indicates a substantial improvement in model reliability and effectiveness, highlighting the metamodel's ability to learn and adapt effectively to diverse and complex datasets. The performance of this model showcases the model's advanced capability in integrating the strengths of both the BBM and GBM approaches while minimizing their individual weaknesses, leading to improved accuracy in fatigue damage predictions across tested scenarios.

As predicted by four models, the accumulated fatigue damage for all evaluation voyages in three cases was also compared with the rainflow counting fatigue damage in Fig. 23. For Case 1, with limited training data, the accumulated fatigue damage predictions for remaining unseen voyages are presented in Fig. 23(a). The pure data-driven BBM shows the largest error due to overestimation in some voyages caused by insufficient data volume, reaching up to 45% error when compared to d^{RFC}. For the GBM based on spectral moments and ship motions, $F^{Combine}$, its relatively large overestimation also results in a significant cumulative error of approximately 19% across 36 unseen voyages (see Fig. 18); however, this is somewhat lower when compared to the BBM. The spectral moments-based GBM (see Fig. 18) has its prediction errors distributed around zero, resulting in a much lower accumulated error of approximately 14%. The proposed metamodel achieves the lowest level of accumulated error (10%). In scenarios with limited data, it is recommended to use the proposed metamodel or the physics-guided GBM for model building. This approach is advocated because, with

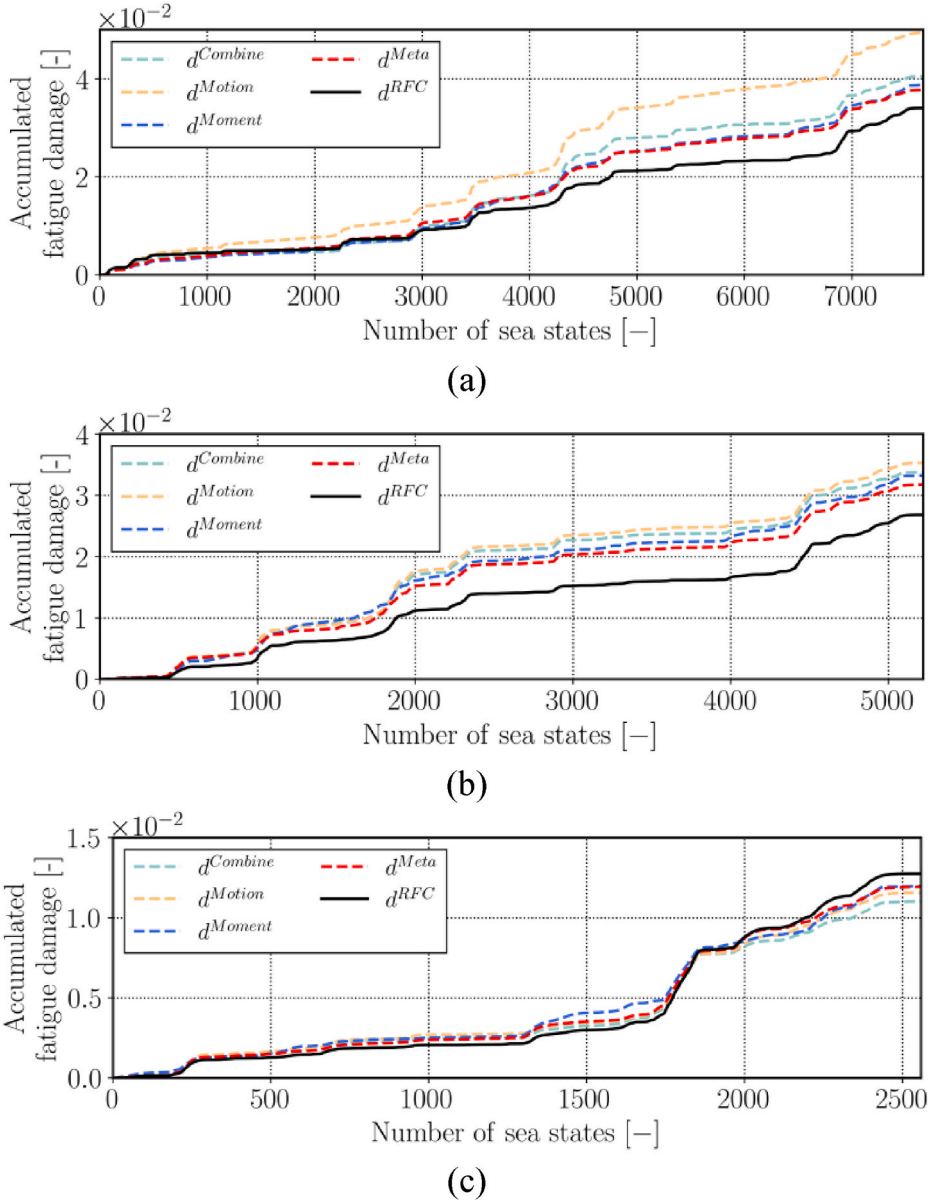

Fig. 23. Accumulated fatigue damage across (a) 36 unseen voyages for Case 1, (b) 24 unseen voyages for Case 2, and (c) 12 unseen voyages for Case 3.

insufficient training data, relying solely on ship motion can lead to unrealistic predictions in future unseen scenarios.

Similar predictive outcomes are observed in Case 2, as shown in Fig. 23(b). Although the BBM still has the largest accumulated error, its overestimation decreases to 32% with additional training data. For $F^{Combine}$, F^{Moment}, and F^{Meta}, while the *RMSE* of fatigue prediction is smaller in Case 2 than in Case 1, their accumulated errors increase slightly to 26, 24, and 18%, respectively. However, the metamodel's prediction d^{Meta} accumulation remains the closest to the accumulated rainflow counting fatigue damage among the four models. In Case 3, where the training data volume reaches 75%, all models' fatigue accumulation predictions closely match the accumulated rainflow counting fatigue damage. The BBM presents an underestimation of approximately 10% and a 13% underestimation of $F^{Combine}$. Although both the metamodel and the spectral moments-based GBM have accumulated errors of 6%, the spectral moments-based GBM exhibits a significant overestimation at around the 1500th sea state, followed by a decrease to underestimation, thereby showing instability. In contrast, the metamodel's predictions are more stable and superior.

6. Conclusions

This study proposes a physics-guided metamodel to predict the vertical bending-induced fatigue damage of a North Atlantic-sailing 2800TEU container vessel by utilizing full-scale measurement data from the onboard hull monitoring system. The metamodel integrates two base learners. The first is a physics-guided GBM developed using XGBoost, which employs spectral moments derived from numerical analysis as inputs. The second is a BBM developed using XGBoost based on ship heave and pitch motion measurements. A meta learner, i.e., GPR, subsequently processes the predictions from both the BBM and GBM, to predict the final fatigue damage. This metamodel has been evaluated against the BBM and GBMs for predicting unseen voyages, using varying volumes of training data. The key conclusions of this study can be summarized as follows.

- The physics-guided GBM captures primary data variations under limited data volumes but struggles with complex relationships as data increases, thus showing diminishing returns in performance improvements.
- The BBM fails to unleash its full predictive potential with limited data, often generating unreasonable predictions for unseen scenarios. However, it excels at uncovering complex patterns with more data, which significantly enhances model performance.
- The innovative metamodel synergistically combines the strengths of both gray-box and black-box models, effectively reducing their individual limitations and enhancing accuracy in fatigue damage predictions across various data volume scenarios.
- This metamodel consistently delivers optimal and stable predictive performance in scenarios with limited data availability and sufficient training data, improving predictions of fatigue damage accumulation by up to 35% when compared to conventional BBMs.

This study also has some assumptions and limitations. To ensure a fair comparison between the conventional spectral method and the proposed method, only fatigue damage induced by wave frequency loads has been considered, achieved through filtering of the stress time series. Due to the absence of onboard metocean measurements, our research employs spatial-temporal interpolation of hindcast data, a method widely used in state-of-the-art research, to match the wave parameters encountered by the case study ship. This work contributes to advancing data-driven methodologies for monitoring ship fatigue damage in open sea navigation. It provides the maritime transportation industry with innovative tools and methodologies to account for ship fatigue accumulation and, when appropriate, to perform predictive maintenance before structural failure.

CRediT authorship contribution statement

Xiao Lang: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Mingyang Zhang:** Writing – review & editing, Investigation, Formal analysis, Conceptualization. **Chi Zhang:** Validation, Investigation, Formal analysis, Conceptualization. **Jonas W. Ringsberg:** Writing – review & editing, Supervision, Methodology, Investigation. **Wengang Mao:** Writing – review & editing, Supervision, Methodology, Funding acquisition, Data curation, Conceptualization.

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.

Acknowledgments

The authors acknowledge the financial support from the Swedish Transport Administration project TRV2019/119989 and the Vinnova project 2021–02768. We are also grateful to the support from the Swedish Foundation for International Cooperation in Research and Higher Education (CH2016-6673), and ship owners for providing model tests results and related full-scale measurements data.

References

Bao, H., Wu, S., Wu, Z., Kang, G., Peng, X., Withers, P.J., 2021. A machine-learning fatigue life prediction approach of additively manufactured metals. Eng. Fract. Mech. 242, 107508 https://doi.org/10.1016/j.engfracmech.2020.107508.

Chen, T., Guestrin, C., 2016. XGBoost. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, New York, NY, USA, pp. 785–794. https://doi.org/10.1145/2939672.2939785.

CMEMS, 2021. Copernicus marine service: global ocean physics reanalysis. URL: https://marine.copernicus.eu/.

Copernicus, 2019. Copernicus Climate Change Service (C3S): ERA5 Fifth generation of ECMWF atmospheric reanalyses of the global climate. URL:https://cds.climate.copernicus.eu/cdsapp#!/home. .

Dirlik, T., 1985. Application of Computers in Fatigue Analysis. Department of Engineering. University of Warwick, Conventry, UK. Doctoral (PhD) Thesis.

DNV, 2005. Fatigue Assessment of Ship Structure, Classification Notes No. 30.7. Høvik, Norway.

DNV, G.L., 2018. DNVGL-CG-0129: Fatigue Assessment of Ship Structure. Det Norske Veritas.

Dong, Y., Garbatov, Y., Guedes Soares, C., 2022. Review on uncertainties in fatigue loads and fatigue life of ships and offshore structures. Ocean Eng. 264, 112514 https://doi.org/10.1016/j.oceaneng.2022.112514.

Friedman, J., Hastie, T., Tibshirani, R., 2000. Additive logistic regression: a statistical view of boosting (With discussion and a rejoinder by the authors). Ann. Stat. 28, 337–407. https://doi.org/10.1214/aos/1016218223.

Gaidai, O., Storhaug, G., Naess, A., Ye, R., Cheng, Y., Xu, X., 2020. Efficient fatigue assessment of ship structural details. Ships Offshore Struct. 15, 503–510. https://doi.org/10.1080/17445302.2019.1661623.

Haranen, M., Pakkanen, P., Kariranta, R., Salo, J., 2016. White, grey and black-box modelling in ship performance evaluation. Proceedings of the 1st Hull Performance & Insight Conference, pp. 115–127.

He, L., Wang, Z., Akebono, H., Sugeta, A., 2021. Machine learning-based predictions of fatigue life and fatigue limit for steels. J. Mater. Sci. Technol. 90, 9–19. https://doi.org/10.1016/j.jmst.2021.02.021.

IACS, 2006. In: Guideline for Rule Development - Ship Structure, No. 106, International Association of Classification Societies.

ISO, 2015. In: Ships and Marine Technology - Guidelines for the Assessment of Speed and Power Performance by Analysis of Speed Trial Data 15016.

ISSC, 2018. In: Kaminski, M.K., Rigo, P. (Eds.), Proceedings of the 20th International Ship and Offshore Structures Congress, Committee III.2: Fatigue and Fracture. The IOS Press.

Lang, X., Wang, H., Mao, W., Osawa, N., 2021. Impact of ship operations aided by voyage optimization on a ship's fatigue assessment. J. Mar. Sci. Technol. 26, 750–771. https://doi.org/10.1007/s00773-020-00769-8.

Lang, X., Wu, D., Tian, W., Zhang, C., Ringsberg, J.W., Mao, W., 2023. Fatigue assessment comparison between a ship motion-based data-driven model and a direct fatigue calculation method. J. Mar. Sci. Eng. 11, 2269. https://doi.org/10.3390/jmse11122269.

Lang, X., Wu, D., Mao, W., 2024. Physics-informed machine learning models for ship speed prediction. Expert Syst. Appl. 238, 121877 https://doi.org/10.1016/j.eswa.2023.121877.

Li, Z., Mao, W., Ringsberg, J.W., Johnson, E., Storhaug, G., 2014. A comparative study of fatigue assessments of container ship structures using various direct calculation approaches. Ocean. Eng. 82, 65–74. https://doi.org/10.1016/j.oceaneng.2014.02.022.

Mao, W., 2010. Fatigue Assessment and Extreme Prediction of Ship Structures. Doctoral (PhD) Thesis, Department of Mathematical Sciences. Chalmers University of Technology, Gothenburg, Sweden.

Mao, W., 2014. Development of a spectral method and a statistical wave model for crack propagation prediction in ship structures. J. Ship Res. 58, 106–116. https://doi.org/10.5957/jsr.2014.58.2.106.

Mao, W., Ringsberg, J., Rychlik, I., Storhaug, G., 2009. Comparison between a fatigue model for voyage planning and measurements of a container vessel. In: Proceedings of International Conference on Ocean Offshore and Arctic Engineering. OMAE), Honolulu, Hawaii. May 31 - June 5, 2009.

Mao, W., Ringsberg, J.W., Rychlik, I., Storhaug, G., 2010. Development of a fatigue model useful in ship routing design. J. Ship Res. 54, 281–293. https://doi.org/10.5957/jsr.2010.54.4.281.

Masoudi Nejad, R., Sina, N., Ghahremani Moghadam, D., Branco, R., Macek, W., Berto, F., 2022. Artificial neural network based fatigue life assessment of friction stir welding AA2024-T351 aluminum alloy and multi-objective optimization of welding parameters. Int. J. Fatig. 160, 106840 https://doi.org/10.1016/j.ijfatigue.2022.106840.

Matsuishi, M., Endo, T., 1968. Fatigue of metals subjected to varying stress. In: Proceedings of the Kyushu Branch of Japan Society of Mechanics Engineering, pp. 37–40. Fukuoka, Japan.

Miner, M.A., 1945. Cumulative damage in fatigue. J. Appl. Mech. 12, A159–A164. https://doi.org/10.1115/1.4009458.

Olsen, A.S., Schrøter, C., Jensen, J.J., 2006. Wave height distribution observed by ships in the North Atlantic. Ships Offshore Struct. 1, 1–12. https://doi.org/10.1533/saos.2005.0009.

Palmgren, A.G., 1924. Life length of roller bearings or durability of ball bearings. Zeitschrift des Vereines Dtsch. Ingenieure 14, 339–341.

Rasmussen, C.E., 2004. Gaussian processes in machine learning. In: Bousquet, O., von Luxburg, U., Rätsch, G. (Eds.), Advanced Lectures on Machine Learning, ML 2003, Lecture Notes in Computer Science, vol. 3176. Springer, Berlin, Heidelberg.

Rychlik, I., 1987. A new definition of the rainflow cycle counting method. Int. J. Fatig. 9, 119–121. https://doi.org/10.1016/0142-1123(87)90054-5.

Schirmann, M.L., Gose, J.W., Collette, M.D., 2023. A comparison of physics-informed data-driven modeling architectures for ship motion predictions. Ocean Eng. 286, 115608 https://doi.org/10.1016/j.oceaneng.2023.115608.

Storhaug, G., Moe, E., Piedras Lopes, T.A., 2007. Whipping measurements onboard a midsize container vessel operating in the North Atlantic. Marintec China Proceedings (RINA, CMP and SNAME), Proceedings of International Symposium on Ship Design and Construction, pp. 55–70.

Thompson, I., 2018. Fatigue damage variation within a class of naval ships. Ocean Eng. 165, 123–130. https://doi.org/10.1016/j.oceaneng.2018.07.036.

Tovo, R., 2002. Cycle distribution and fatigue damage under broad-band random loading. Int. J. Fatig. 24, 1137–1147. https://doi.org/10.1016/S0142-1123(02)00032-4.

Tucker, M.J., 1991. Waves in Ocean Engineering. Ellis Horwood Ltd., Chichester.

Winterstein, S.R., 1985. Non-normal responses and fatigue damage. J. Eng. Mech. 111, 1291–1295. https://doi.org/10.1061/(ASCE)0733-9399(1985)111:10(1291.

Winterstein, S.R., 1988. Nonlinear vibration models for extremes and fatigue. J. Eng. Mech. 114, 1772–1790. https://doi.org/10.1061/(ASCE)0733-9399(1988)114:10(1772).

Wirsching, P.H., Light, M.C., 1980. Fatigue under wide band random stresses. J. Struct. Div. 106, 1593–1607. https://doi.org/10.1061/JSDEAG.0005477.

Yamamoto, N., 2017. Fatigue evaluation of ship structures considering change in mean stress condition. Weld. World 61, 987–995. https://doi.org/10.1007/s40194-017-0461-x.

Yan, W., Deng, L., Zhang, F., Li, T., Li, S., 2019. Probabilistic machine learning approach to bridge fatigue failure analysis due to vehicular overloading. Eng. Struct. 193, 91–99. https://doi.org/10.1016/j.engstruct.2019.05.028.

Yang, P., Li, J., Zhang, W., Wu, D., Gu, X., Ma, Q., 2021. Analysis on statistical uncertainties of wave loads and structural fatigue reliability for a semi-submersible platform. Ocean Eng. 237, 109609 https://doi.org/10.1016/j.oceaneng.2021.109609.

Yosri, A., Leheta, H., Saad-Eldeen, S., Zayed, A., 2022. Accumulated fatigue damage assessment of side structural details in a double hull tanker based on spectral fatigue analysis approach. Ocean Eng. 251, 111069 https://doi.org/10.1016/j.oceaneng.2022.111069.

Zhao, W., Baker, M.J., 1990. A new stress-range distribution model for fatigue analysis under wave loading. Proceedings of International Conference on Environmental Forces of Offshore Structures and Their Prediction. Kluwer Academic Publishers.