2021 IEEE International Conference on Robotics and Automation (ICRA 2021)
May 31 - June 4, 2021, Xi'an, China

Data-driven sea state estimation for vessels using multi-domain features from motion responses

Peihua Han[1], Guoyuan Li[1,*], Stian Skjong[2], Baiheng Wu[1], and Houxiang Zhang[1]

***Abstract*— Situation awareness is of great importance for autonomous ships. One key aspect is to estimate the sea state in a real-time manner. Considering the ship as a large wave buoy, the sea state can be estimated from motion responses without extra sensors installed. However, it is difficult to associate waves with ship motion through an explicit model since the hydrodynamic effect is hard to model. In this paper, a data-driven model is developed to estimate the sea state based on ship motion data. The ship motion response is analyzed through statistical, temporal, spectral, and wavelet analysis. Features from multi-domain are constructed and an ensemble machine learning model is established. Real-world data is collected from a research vessel operating on the west coast of Norway. Through the validation with the real-world data, the model shows promising performance in terms of significant wave height and peak period.**

I. INTRODUCTION

Remotely operated and autonomous ships are a topic of increasing interest among the maritime industry. These ships have the potential to reduce human-based errors while lower fuel consumption and extend the operational window [1]. The dependency on situation awareness is a common basis for autonomous systems. The system must be able to process the current environment and use it for safe and effective decisions. The sea state is crucial environmental information for the vessel operated on the sea. Real-time estimation of sea state are of key importance for autonomous ships.

In oceanography, a sea state refers to the general condition of the ocean with respect to wind waves and swell at a certain location and moment. A sea state is usually characterized by statistical parameters, e.g., significant wave height, average wave frequency, peak frequency [2]. The primary tools nowadays to collect accurate statistical wave data are floating wave buoys. However, wave buoys are not practical for a vessel in maneuvering since they are present at fixed locations. Other methods include meteorological satellite and wave radar. The meteorological satellite image quality may be affected if the cloud is low and it is often subjected to a time delay of several hours. Wave radar satisfies the need to estimate on-site sea state in real-time. But these systems are expensive to install, require frequent calibration [3], and only equipped to limited vessels.

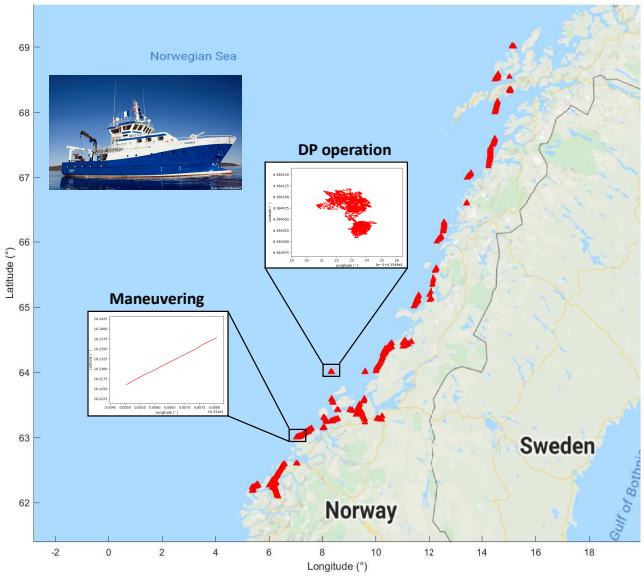

Fig. 1. Illustration of the data we collected from R/V Gunnerus operating on the west coast of Norway. R/V Gunerus is showed in the upper left corner. The red lines indicate the area where the ship has been. The data contains that the ship either in a DP operation or heading in a straight line with constant speed. The data is used to build data-driven model for real-time sea state estimation.

The majority of marine vessels are equipped with sensors that measure the ship motions in 6 degrees of freedom. A ship can be considered as a large wave buoy and the motion responses reflect the sea state condition. From this perspective, a ship is essentially equipped with an environmental condition estimation system [4]. Estimating the sea state based on the ship motion responses is of interest and has been investigated. Several challenges exist to estimating the sea state using motion data: (1) ocean waves are stochastic processes and wave energy is distributed over frequencies and directions; (2) the relationship between wave and ship motion is difficult to model; (3) the moving of the ship adds extra complexity. Previous works involve model-based methods that use response amplitude operators (RAOs) to relate the sea state to vessel responses. RAOs are complex-valued transfer functions that are calculated using strip theory and sometimes computational fluid dynamics. Ship responses are, in general, non-linearly related to the wave excitation. However, the transfer functions are linear and therefore only hold for light and moderate sea states [5]. Besides, the

This work was supported by a grant from the Research Council of Norway through the Knowledge-Building Project for industry "Digital Twins For Vessel Life Cycle Service" (project no. 280703)

* Corresponding author

[1]Peihua Han, Guoyuan Li, Baiheng Wu and Houxiang Zhang are with the Department of Ocean Operations and Civil Engineering, Norweigian University of Science and Technology, Aalesund 6009, Norway {peihua.han, guoyuan.li, baiheng.wu, hozh}@ntnu.no

[2]Stian Skjong is with SINTEF Ocean, Trondheim 7052, Norway Stian.Skjong@sintef.no

2021 IEEE International Conference on Robotics and Automation (ICRA) | 978-1-7281-9077-8/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICRA48506.2021.9561261

RAOs are difficult to estimate exactly and might need to be tuned with real-world data. On the other hand, data-driven approaches can be employed to learn the mapping from measured ship motion responses to actual sea state. The advantage of these approaches is that they are able to discover the pattern between ship motion and sea state based on historical experience.

While this task can naturally be posed as a supervised machine learning problem, how to extract salient features make this a challenging task. The extracted features not only have a significant influence on the model performance but also can enhance understandings of this problem. End-to-end deep learning methods can produce hierarchical representation but prevent interpretation of the generated features. However, features are essential for interpreting signals because it is not like images, which are straightforward to understand for humans. A certain degree of feature interpretability, e.g., how each feature contributes to the prediction, can provide post-hoc validation to examine if the model actually learns some useful rules. Besides, it could be beneficial for the user to trust the model, which is of key importance in the maritime industry.

In this paper, a data-driven model is proposed to estimate the sea state based on ship motion responses. Features are generated by means of statistical, temporal, spectral, and wavelet analysis. The features are then used to predict the sea state information with machine learning models. Real-world operation data is collected from a vessel as illustrated in Fig. 1. Validation is performed to show the effectiveness of the proposed method.

The remainder of this paper is organized as follows: an introduction to sea state estimation is given in Section II. Section III presents the design of network. The experiments are discussed in Section IV. Section V concludes the paper.

II. RELATED WORK

Estimating the sea state from measured ship motion responses has been studied by many researchers. Most of the research focuses on the field of frequency domain analysis. The ship motion response is first transformed into the frequency domain through fast Fourier transform or autocorrelation analysis. The RAOs are then used to relate the wave spectrum to the motion spectrum. The fundamental idea is to minimize the difference between the measured ship spectrum and the calculated ship spectrum [6]. If a wave spectra, e.g., JONSWAP, Bretschneider, is assumed, the wave parameters are obtained in the nonlinear optimization process [7], [8], [9]. Otherwise, a Bayesian approach can be applied [10], [11], in which the wave spectrum is represented in a discrete frequency-directional domain and the original least square problem is transformed into the maximization of posterior. The methods are initially developed for dynamically positioned (DP) vessel, Iseki and Ohtsu [10] extend this method to ship with forward speed by addressing the triple-valued function problem in the following seas.

Considering this method highly relies on spectral analysis which may influence the outcome, the estimation of sea state from ship motion response can be also addressed in the time domain. Pascoal and Soares [12] propose an estimation algorithm established based on Kalman filter. The wave components are introduced as state variables. The method is further extended to ship with forward speed and validation is performed through sea trial data [13]. A similar observer-based approach is developed by Belleter et al. [14], [15] to estimate the wave frequency using the measured roll or pitch angle. Nielsen et al. [16] propose a step-wise method that uses the frequency estimator above together with nonlinear least square for estimating wave amplitude and phase. Nevertheless, these two methods, either frequency or time domain, are dependent on RAOs to relate the wave to the ship motion. RAOs are simplified linear transfer function and hard to obtain exactly (tuning with real-world data is often needed). Errors may be produced and RAOs only hold for mild and moderate sea state [5].

Data-driven methods are alternative methods that learn the mapping between ship motion and sea state directly. The advantage of these methods is that it does not rely on an explicit model to link waves to ship motion. Tu et al. [17] extract time and frequency domain statistical information of the measured motion data and apply a three-layer classifier to classify the sea state. Arneson et al. [18] integrate the frequency components over the frequency range for the motion response spectrum as features and use them to train the machine learning models. A concern is how to extract useful features. Even though the end-to-end deep learning method also has been developed [19], [20], [21], the features generated automatically from this method are impossible to interpret. This paper focuses on generating interpretable features and applying them to develop a machine learning model for sea state estimation.

III. DATA-DRIVEN SEA STATE ESTIMATION

A. Overview

The overview of our proposed method is shown in Fig. 2. The measured motion data are time-series data. The measured motion data are analyzed by means of statistical analysis, temporal analysis, spectral transform, and wavelet transform. In this manner, a thorough analysis of the signals is conducted to capture more information. To avoid the curse of dimensionality, a filter-based feature selection technique is applied to limit the feature space. With the selected salient features, z-score normalization is first applied and then three different machine learning models are trained. The predictions from these three models are then ensembled using the voting method to provide the final prediction.

B. Signal detrending

In order to reduce the impact from ship actuation and eliminate possible offset introduced by the sensor measurement, local regression [22] is used to detrend the signal and assure the displacement has almost zero mean. At each point a weighted least-squares fit is applied to the local subset of the data, the signal trend is obtained after all the points in the data are fitted. Subsequently, the final displacement history

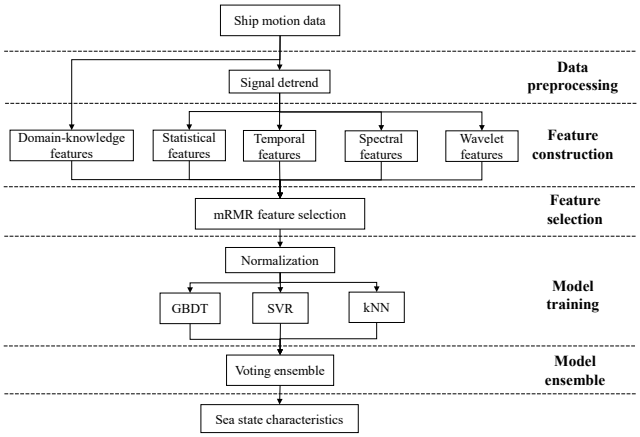

Fig. 2. A schematic illustration of the proposed model.

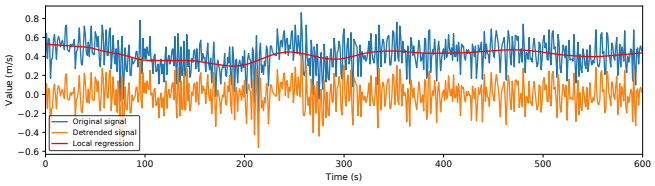

Fig. 3. Example of the detrending process.

is found as the difference between the original displacement and the signal trend. Fig. 3 shows an example from the signal detrending process.

C. Multi-domain feature construction

Considering a signal is a discrete time series data $(x_1, x_2, \ldots, x_n)$ with length n, four broad categories of features are constructed to describe sea state pattern.

1) Domain-knowledge features: Two basic features are extracted. *speed*: the forward speed is important for estimating wave characteristics due to the Doppler shift [5], *diff_angle*: the difference between course angle and heading angle during ship maneuvering.

2) Statistical features: Seven basic statistical features are extracted from each DOF measurement. Six standard features of the signal including *maximum*, *minimum*, *mean*, *variance*, *skew*, and *kurtosis* are considered. Additionally, the *q quantile* information of the signal is extracted, which is the value greater than q of the ordered values from the signal. The variable q is selected as 0.2, 0.4, 0.6, and 0.8.

3) Temporal features: Firstly five temporal features are considered, which include: *absolute sum of change* ($\sum_{i=1}^{n-1}|x_{i+1}-x_i|$), *absolute energy* ($\sum_{i=1}^{n}x_i^2$), *mean second derivative center* ($\frac{1}{2(n-2)}\sum_{i=1}^{n-2}\frac{1}{2}(x_{i+2}-2x_{i+1}+x_i)$), *zero cross* (the number of the signal crossings zero), *longest strike above mean* (the length of the longest consecutive subsequence in a signal that is largest than its mean).

Two advanced temporal features are also extracted:

Sample entropy: This feature uses $-log\frac{A}{B}$ to measure the complexity of a time series. Given a subsequence $seq_m(t) = [x_t, x_{t+1}, \ldots, x_{t+m}]$ with length m, B is the number of subsequence pairs have $d(seq_m(i), seq_m(j)) < r$ while A is the number of pairs that after being extended to length $m+1$ have $d(seq_{m+1}(i), seq_{m+1}(j)) < r$, where d is the Chebychev distance and r is a tolerance threshold.

Autocorrelation: This feature measures the similarity between observations as a function of the time lag between them. For a discrete process, the autocorrelation is obtained as $\hat{R}(k) = \frac{1}{(n-k)\sigma^2}\sum_{i=1}^{n-k}(x_i-\mu)(x_{i+k}-\mu)$, where μ and σ^2 are the mean and variance respectively. k denotes the time lag. Five different time lags 10, 20, 30, 40, 50 are used to extract this feature.

4) Welch spectral features: Welch method [23] is an approach to transform a signal from the time domain to the frequency domain and estimate the power of a signal at different frequencies. The method is based on the concept of using periodogram spectrum estimates by computing the fast Fourier transform (FFT). However, the periodogram is a high variance estimate due to the use of a finite data stream. Hamming window is used to smooth the spectrum. After the signal is transformed into frequency domain, four basic spectral features including *max power spectrum*, *fundamental frequency*, *max frequency*, and *median frequency* are extracted. Additionally, five features related to the shape of the spectrum [24] is also extracted: *centroid*, *variation*, *spread*, *skewness*, *kurtosis*.

5) Wavelet features: The wavelet transform is a time-frequency analysis method which selects the appropriate frequency band adaptively based on the characteristics of the signal. This enables wavelet transform to analyze a signal with multi-scales both in time and frequency domain. The wavelet transform is used to split a signal into different frequency sub-bands. The Daubechies wavelet of order 1 (db1) is selected as the basis function and the decomposition level is five. In total five approximation components and five detail components are obtained. For each components, the *mean*, *variance*, *median*, *skewness*, *kurtosis*, *absolute energy*, *absolute sum of changes*, and *zero cross* are extracted.

D. Minimum-redundancy maximum-relevance (mRMR) feature selection

In order to select salient features from the constructed multi-domain features, mRMR [25] feature selection framework is utilized. The mRMR criterion is a filter feature selection method which can effectively reduce the redundant features while keeping the relevant features for the model. The mRMR criterion can be expressed as:

$$f_{mRMR}(x_i) = I(y, x_i) - \frac{1}{|S|}\sum_{x \in S} I(x_s, x_i) \tag{1}$$

where the function $I(\cdot,\cdot)$ denotes the mutual information (MI). $|S|$ is the size of the feature set and $x_s \in S$ is one feature out of the feature set. The first term in Eq.(1) represents the relevant to the target y while the second term measures the redundancy. Since the MI is computationally expensive for continuous variables, the redundancy is replaced with correlation. The MI used to measure the relevance is normalized to $[0,1]$ to have a same range with the correlation:

$$f_{mRMR}(x_i) = I_{norm}(y,x_i) - \frac{1}{|S|}\sum_{x\in S}\rho(x_s,x_i) \tag{2}$$

where the function $I_{norm}(\cdot,\cdot)$ denotes the normalized mutual information and $\rho(\cdot,\cdot)$ denotes the Pearson correlation. The criterion is used to rank the features and provide a basis for feature selection of learning models.

E. Learning models

1) k-nearest neighbor (kNN): kNN is an intuitive and efficient method that has been used extensively in the machine learning literature. The basic idea of kNN is to predict a testing point based on a fixed number k of its closest neighbors in the feature space. These neighbors are chosen from a set of training points $((x_1,y_1),...,(x_n,y_n))$ whose targets are known. For a novel test point x, kNN regression computes the mean of the function values of its k-nearest neighbors. A distance-based version kNN is used here, where the k-nearest neighbors are weighted by the inverse of their distance:

$$f_{kNN}(x) = \frac{1}{k}\sum_{i\in N_k(x)}\frac{\frac{1}{d(x,x_i)}}{\sum_{i\in N_k(x)}\frac{1}{d(x,x_i)}}y_i \tag{3}$$

where set $N_k(x)$ containing the indices of the k-nearest neighbors of x. $d(x,x_i) = \sqrt{|x-x_i|^2}$ is Euclidean distance.

2) Support vector regression (SVR): SVR is an effective method for modeling and interpolating nonlinear functions. Due to the cost function that ignores any training data close to the model prediction, it depends only on a subset of the training data. The basic idea of SVR is to fit a function $f(x) = \langle w,x\rangle + b$ onto a training data set. The weights vector w,b can be obtained by solving the optimization problem:

$$\begin{aligned} \min_{w,b} \quad & \frac{1}{2}||x||^2 + C\sum_{i=1}^{n}(\xi_i - \xi_i^*) \\ \text{s.t.} \quad & -\varepsilon - \xi_i^* \le \langle w,x_i\rangle + b - y_i \le \varepsilon + \xi_i \\ & \xi_i, \xi_i^* \ge 0 \end{aligned} \tag{4}$$

where ξ and ξ^* are slack variables representing the deviation from a predefined gap with hyperparameter ε. The hyperparameter C denotes the strength of the regularization which is inversely proportional to C. Solving this problem requires the application of the Lagrangian multiplier technique, which by itself leads to a dual optimization problem:

$$\begin{aligned} \min_{\alpha,\beta} \quad & \frac{1}{2}\sum_{i,j=1}^{n}(\alpha_i - \beta_i)(\alpha_j - \beta_j)\kappa(x_i,x_j) \\ & + \varepsilon\sum_{i=1}^{n}(\alpha_i + \beta_i) - \sum_{i=1}^{n} y_i(\alpha_i + \beta_i) \\ \text{s.t.} \quad & 0 \le \alpha_i, \beta_i \le C \end{aligned} \tag{5}$$

where $\kappa(x_i,x_j)$ is a kernel function which is used to account for nonlinearities. In this paper, radial basis function (RBF) kernel is used, which is expressed as $\kappa(x_i,x_j) = exp(-\gamma||x_i - x_j||^2)$. γ is a hyperparameter in the RBF kernel. The details of SVR and the solving process for its dual optimization problem can be refer to [26].

3) Gradient boost decision tree (GBDT): GBDT is an ensemble model using gradient boost technique with decision trees as base learners. The prediction of the GBDT is the sum of M trees:

$$\hat{y}_i = \sum_{t=1}^{M} f_t(x_i) \tag{6}$$

In t iteration, a tree model $f_t(\cdot)$ is generated by minimizing the following function:

$$\begin{aligned} Obj^{(t)} &= \sum_{i=1}^{n} l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \Omega(f_t) \\ &\simeq \sum_{i=1}^{n}[g_i f_t(x_i) + \frac{1}{2}h_i f_t^2(x_i)] + \Omega(f_t) + constant \end{aligned} \tag{7}$$

where $l(\cdot,\cdot)$ denotes the loss function, mean square error is usually used for regression problem. $\Omega(\cdot)$ is a regularization term for decision tree. The objective function can be approximated by second-order Taylor expansion, where $g_i = \partial_{\hat{y}^{(t-1)}} l(y_i, \hat{y}^{(t-1)})$ and $h_i = \partial^2_{\hat{y}^{(t-1)}} l(y_i, \hat{y}^{(t-1)})$.

4) Model ensemble: Model ensemble is a commonly used method to improve model performance. The above three models are ensemble using voting method, which is simply averaging the predictions from the three models. Tree-based model (GBDT), kernel-based model (SVR) and distance-based model (kNN) are used here to ensure the diversity of sub-models, which also influences the ensemble performance.

IV. EXPERIMENT

A. Data collection

The ship operation data is collected from the Norwegian University of Science and Technology's research vessel, R/V Gunnerus. R/V Gunerus is a multi-purposed vessel with length $31.25m$ and draught $2.7m$. The data is collected from 2017 to 2019. The trajectory of the vessel is first filtered to obtain the data in nearly stationary operation conditions, where the course angle and forward speed do not change significantly. As shown in Fig. 1, the red line in the figure indicates the position of the ship when the data was collected and used in this paper. Four ship motion time-series responses are extracted: *sway velocity*, *roll*, *pitch*, and *heave*. The forward speed and the difference between the heading angle and course angle are extracted as domain-knowledge features. The sampling frequency for these sensors is 1 Hz. The data stream is then cut into 10 minutes segment without overlapping.

The sea state information is collected from the Norwegian Meteorological Institute. Three sea state characteristics are considered: significant wave height h_s, mean wave direction D_m, and peak period T_p. The three sea state characteristics are then used to label the ship motion data by nearest neighbor search in coordinate with latitude and longitude. Fig. 4 shows the distribution of the collected data.

A multi-target regression dataset is formed with forward speed, sway, roll, pitch, heave as input and significant wave

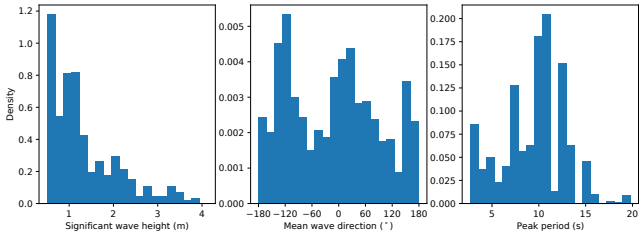

Fig. 4. Distribution of the collected sea state characteristics.

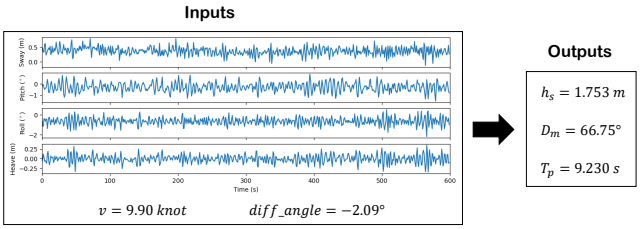

Fig. 5. Example of a random selected sample from the dataset.

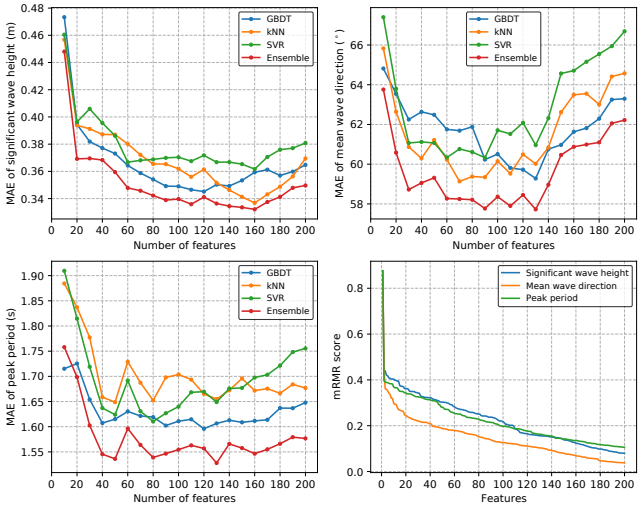

Fig. 6. The effect of the number of selected features.

height h_s, mean wave direction D_m, and peak period T_p as output. The forward speed is constant variable while the sway, roll, pitch, heave are time-series data. Fig. 5 presents a random sample drawn from the dataset.

B. Experimental setup

The mean absolute error (MAE) is used as the evaluation metric. We performed 5-fold cross-validation to avoid possible selection bias on splitting the dataset. It also helps to give an insight into how the model will generalize to the dataset. The hyperparameters of all the models are tuned using a random search strategy [27].

C. Performance comparison

To evaluate the performance of the proposed method, four baselines model are also implemented: (1) Random Guess: a simple model that makes the predictions by randomly drawing from the training data distribution. (2) Linear Regression with Elastic Net regularization (EN): a regularized linear regression method that linearly combines the l_1 and l_2 penalties, the hyperparameter for l_1 and l_2 are tuned. (3) Multilayer Perceptron (MLP): a class of feedforward artificial neural network, ReLU is used as the activation function and Adam is used as the optimizer. The learning rate and the weight for l_2 regularization are tuned. (4) Random Forest (RF): an ensemble model that uses the decision trees as base learners and bagging to improve the performance. The maximum depth, the minimum number of samples required to be at a leaf node, number of features to consider when looking for the best split are tuned.

In addition, an end-to-end model SeaStateNet [19] is implemented. This model consists of a convolutional neural network block, a long-short-term memory block, and an FFT block. The model uses the time series data as input directly and therefore does not rely on our constructed features. Since the model is originally designed for classification, the output nodes in the last layer are changed to three and the loss function is changed to mean square error in order to adapt to the dataset.

Table I reports the MAE of the models, evaluated through the 5-fold cross-validation. The results are presented in $mean \pm std$ format. For the feature-based approaches, GBDT consistently outperforms the other approaches. The high MAEs of the EN model suggest that the sea state characteristics are better captured using nonlinear relationships. The ensemble model consists of GBDT, kNN, and SVR outperforms any individual models. The end-to-end approach SeaStateNet outperform the ensemble model in D_m but have a slightly higher MAE in h_s and T_p. The reason might be that the mean wave direction is not so sensitive to the constructed features. Besides, the errors for h_s and T_p are in an acceptable range, while the error for D_m is relatively high even though it is clearly better than the random guess. The reason might be that most of the time the vessel is operating near the coast and the mean wave direction from the weather forecast system is not so accurate in this region.

TABLE I
THE MAE VALUES OF THE DIFFERENT METHODS

Model	Wave Characteristics		
	$h_s\,(m)$	$D_m\,(^\circ)$	$T_p\,(s)$
EN	0.484 ± 0.027	77.59 ± 3.32	2.032 ± 0.172
MLP	0.431 ± 0.045	71.84 ± 6.50	1.851 ± 0.119
RF	0.378 ± 0.024	64.34 ± 4.62	1.686 ± 0.116
kNN	0.359 ± 0.025	60.02 ± 3.58	1.655 ± 0.095
SVR	0.361 ± 0.024	60.96 ± 2.58	1.649 ± 0.100
GBDT	0.337 ± 0.027	59.28 ± 2.26	1.607 ± 0.096
SeaStateNet	0.348 ± 0.019	$\mathbf{53.82 \pm 3.09}$	1.659 ± 0.178
Our ensmble	$\mathbf{0.334 \pm 0.030}$	57.72 ± 1.30	$\mathbf{1.528 \pm 0.084}$

D. Evaluation on generated features

In order to verify the effect of the generated features, the four groups of features are added sequentially for performance validation. Table II summarizes the comparison

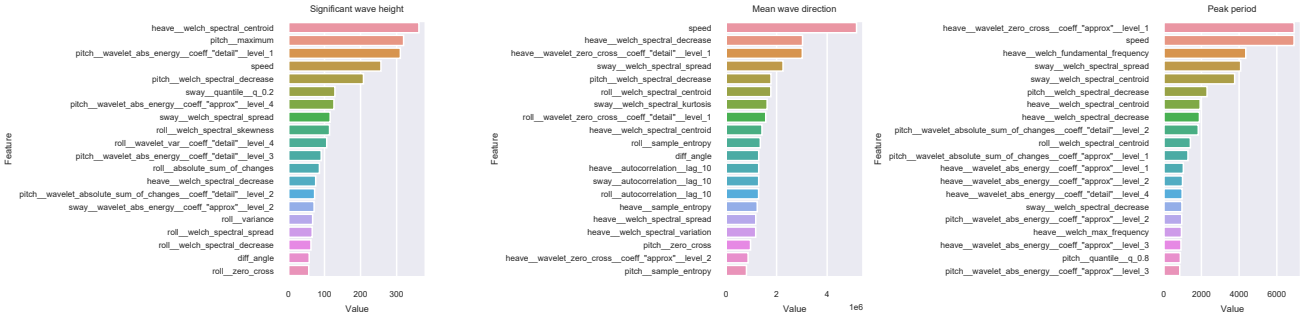

Fig. 7. Feature importance in GBDT measured by gain-based Gini importance.

results with the feature groups, where f_1, f_2, f_3, f_4 and f_5 denote domain-knowledge features, statistical features, temporal features, welch spectral features and wavelet features, respectively. The performance of the model in terms of h_s, D_m and T_p increases with an increased group of features. It shows that the features from the multi-domain are beneficial. With the addition of spectral features and wavelet features, the overall performance of the model on these three characteristics improves around 10%.

TABLE II

COMPARISON OF DIFFERENT FEATURES

Features	Wave Characteristics		
	$h_s\,(m)$	$D_m\,(^\circ)$	$T_p\,(s)$
f_1+f_2	0.431 ± 0.034	73.26 ± 4.43	1.956 ± 0.091
$f_1+f_2+f_3$	0.351 ± 0.027	61.96 ± 4.41	1.634 ± 0.130
$f_1+f_2+f_3+f_4$	0.338 ± 0.025	58.21 ± 2.30	1.546 ± 0.093
$f_1+f_2+f_3+f_4+f_5$	$\mathbf{0.334\pm0.030}$	$\mathbf{57.72\pm1.30}$	$\mathbf{1.528\pm0.084}$

Fig. 6 shows the MAE of h_s, D_m, and T_p versus the number of selected features using mRMR criterion. The results of GBDT, kNN, and SVR are also presented since they are the bases for the ensemble model. The subgraph in the bottom right shows the mRMR score with the features in descending order. The performance of the model first increases with the number of features and then the performance degrades. The effect of the features redundancy is much obvious at D_m. The optimal number of selected features for h_s, D_m and T_p is 160, 130 and 130, respectively.

Understanding how the features contribute to its prediction can be used to inspect the model with domain-knowledge as well as build trust for the users. The GDBT in our ensemble model provides a way to inspect the importance of features. Fig. 7 shows the top 20 most important features. The feature importance is measured by total Gini gains of splits that use the feature. The strength of the signal, represented by *abs_energy*, *absolute_sum_of_change*, is important for significant wave height. The mean wave direction gives much focus on the shape of the spectrum. As for the peak period, the wavelet features occupy an important part since it represents the signal in the different frequency ranges.

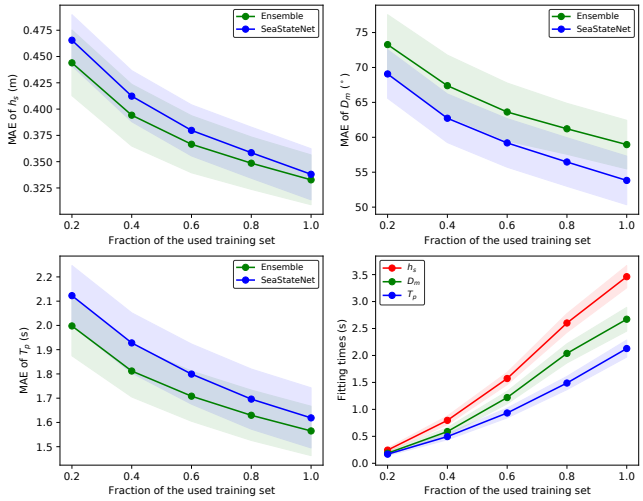

Fig. 8. The effect of the number of training samples.

E. Sensitivity of training data size

The sensitivity of the model with respect to the training data size is studied since it is difficult to collect a large dataset for different kinds of ships. Fig. 8 shows the validation error versus the fraction of the used training set. The subgraph in the bottom right shows the training time. The errors increase when fewer training samples are used. The end-to-end model is a little bit more sensitive to the training data size than our feature-based model.

V. CONCLUSIONS

This paper presents a data-driven model for performing real-time onboard sea state estimation for vessels. The model is based on extracting features with statistical, temporal, spectral, and wavelet analysis. Data collected from real-world scenarios show that the method can provide relatively accurate results in terms of significant wave height and peak period. A detailed analysis of the model is performed, which shows that the features from the multi-domain are beneficial. Future work should include more data far from shore to examine the model's ability in estimating wave direction. Also, a more detailed analysis of the extracted features should be conducted and the model uncertainty should be included for actual usage.

REFERENCES

[1] R. Jalonen, R. Tuominen, and M. Wahlström, "Remote and autonomous ships—the next steps: Safety and security in autonomous shipping—challenges for research and development," *Rolls-Royce, Buckingham Gate, London: The Advanced Autonomous Waterborne Applications (AAWA)*, pp. 56–73, 2016.

[2] T. I. Fossen, *Handbook of marine craft hydrodynamics and motion control*. John Wiley & Sons, 2011.

[3] D. C. Stredulinsky and E. M. Thornhill, "Ship motion and wave radar data fusion for shipboard wave measurement," *Journal of ship research*, vol. 55, no. 2, pp. 73–85, 2011.

[4] A. H. Brodtkorb, U. D. Nielsen, and A. J. Sørensen, "Sea state estimation using vessel response in dynamic positioning," *Applied Ocean Research*, vol. 70, pp. 76–86, 2018.

[5] U. D. Nielsen, "Estimation of directional wave spectra from measured ship responses," in *12th International Congress of the International Maritime Association of the Mediterranean: Maritime Transportation and Exploitation of Ocean and Coastal Resources*, 2005, pp. 1103–1112.

[6] U. D. Nielsen, "Estimations of on-site directional wave spectra from measured ship responses," *Marine Structures*, vol. 19, no. 1, pp. 33–69, 2006.

[7] J. Hua and M. Palmquist, "Wave estimation through ship motion measurement," 1994.

[8] E. A. Tannuri, J. V. Sparano, A. N. Simos, and J. J. Da Cruz, "Estimating directional wave spectrum based on stationary ship motion measurements," *Applied Ocean Research*, vol. 25, no. 5, pp. 243–261, 2003.

[9] U. D. Nielsen and D. C. Stredulinsky, "Onboard sea state estimation based on measured ship motions," in *12th International Ship Stability Workshop, Washington, DC, June*, 2011, pp. 12–15.

[10] T. Iseki and K. Ohtsu, "Bayesian estimation of directional wave spectra based on ship motions," *Control Engineering Practice*, vol. 8, no. 2, pp. 215–219, 2000.

[11] T. Iseki, D. Terada *et al.*, "Bayesian estimation of directional wave spectra for ship guidance system," *International Journal of Offshore and Polar Engineering*, vol. 12, no. 01, 2002.

[12] R. Pascoal and C. G. Soares, "Kalman filtering of vessel motions for ocean wave directional spectrum estimation," *Ocean Engineering*, vol. 36, no. 6-7, pp. 477–488, 2009.

[13] R. Pascoal, L. P. Perera, and C. G. Soares, "Estimation of directional sea spectra from ship motions in sea trials," *Ocean Engineering*, vol. 132, pp. 126–137, 2017.

[14] D. J. Belleter, D. A. Breu, T. I. Fossen, and H. Nijmeijer, "A globally k-exponentially stable nonlinear observer for the wave encounter frequency," *IFAC Proceedings Volumes*, vol. 46, no. 33, pp. 209–214, 2013.

[15] D. J. Belleter, R. Galeazzi, and T. I. Fossen, "Experimental verification of a global exponential stable nonlinear wave encounter frequency estimator," *Ocean Engineering*, vol. 97, pp. 48–56, 2015.

[16] U. D. Nielsen, R. Galeazzi, and A. H. Brodtkorb, "Evaluation of shipboard wave estimation techniques through model-scale experiments," in *OCEANS 2016-Shanghai*. IEEE, 2016, pp. 1–8.

[17] F. Tu, S. S. Ge, Y. S. Choo, and C. C. Hang, "Sea state identification based on vessel motion response learning via multi-layer classifiers," *Ocean Engineering*, vol. 147, pp. 318–332, 2018.

[18] I. B. Arneson, A. H. Brodtkorb, and A. J. Sørensen, "Sea state estimation using quadratic discriminant analysis and partial least squares regression," *IFAC-PapersOnLine*, vol. 52, no. 21, pp. 72–77, 2019.

[19] X. Cheng, G. Li, R. Skulstad, S. Chen, H. P. Hildre, and H. Zhang, "Modeling and analysis of motion data from dynamically positioned vessels for sea state estimation," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 6644–6650.

[20] B. Mak and B. Düz, "Ship as a wave buoy: Estimating relative wave direction from in-service ship motion measurements using machine learning," in *International Conference on Offshore Mechanics and Arctic Engineering*, vol. 58882. American Society of Mechanical Engineers, 2019, p. V009T13A043.

[21] X. Cheng, G. Li, A. L. Ellefsen, S. Chen, H. P. Hildre, and H. Zhang, "A novel densely connected convolutional neural network for sea state estimation using ship motion data," *IEEE Transactions on Instrumentation and Measurement*, 2020.

[22] W. S. Cleveland, "Robust locally weighted regression and smoothing scatterplots," *Journal of the American statistical association*, vol. 74, no. 368, pp. 829–836, 1979.

[23] P. Welch, "The use of fast fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms," *IEEE Transactions on audio and electroacoustics*, vol. 15, no. 2, pp. 70–73, 1967.

[24] G. Peeters, B. L. Giordano, P. Susini, N. Misdariis, and S. McAdams, "The timbre toolbox: Extracting audio descriptors from musical signals," *The Journal of the Acoustical Society of America*, vol. 130, no. 5, pp. 2902–2916, 2011.

[25] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 27, no. 8, pp. 1226–1238, 2005.

[26] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statistics and computing*, vol. 14, no. 3, pp. 199–222, 2004.

[27] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," *The Journal of Machine Learning Research*, vol. 13, no. 1, pp. 281–305, 2012.