NTNU

Norwegian University of Science and Technology

Data-driven sea state estimation from motion responses

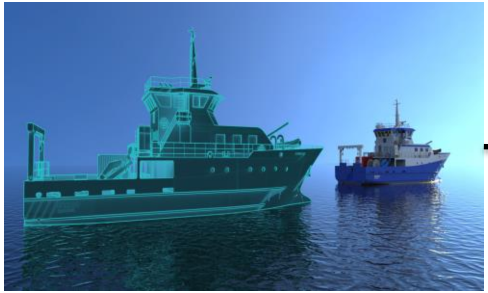
Peihua Han, Guoyuan Li, Stian Skjong, Baiheng Wu, Houxiang Zhang

Content

- Background
- Motivation
- Data collection and Methodology
- Experimental Results
- Conclusions

Background

- Digitalization and automation is a topic of interest in the maritime industry.



Digital twin

Remote control

Autonomous ship

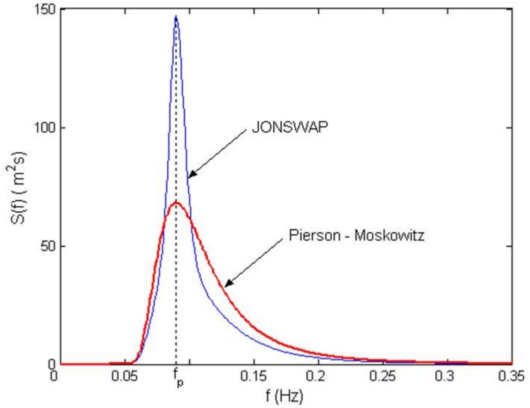
Understanding the current environmental conditions is important for safe operation and control of the vessel.

- Sea state condition is a key environmental condition.

Motivation

1. Represent ocean waves:

- Ocean waves are stochastic with time.
- The statistical properties can be evaluated.
- The statistical properties are not likely to change for hours (short-term statistics).



Sea state statistics: significant wave height, etc.

2. Tools to obtain the sea state information:

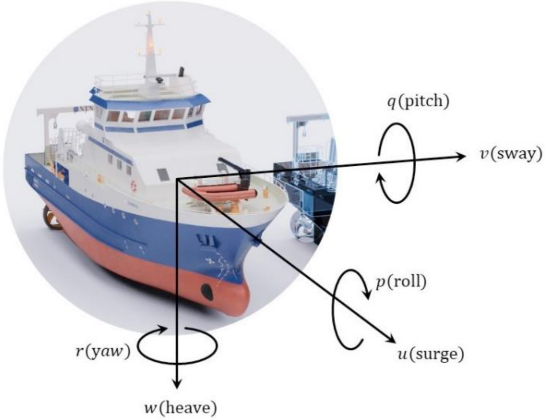
Wave buoy:

- Estimate wave from motions (roll, pitch, heave).
- Moored to the seabed (fixed to a specific location).
- Shallow water.

Wave radar:

- Image time series analysis.
- Not good at wave height.
- Only limited vessel have it.

Motivation



A ship can be considered as a large wave buoy and the motion responses (heave, pitch, roll) reflect the sea state condition.

Challenges:

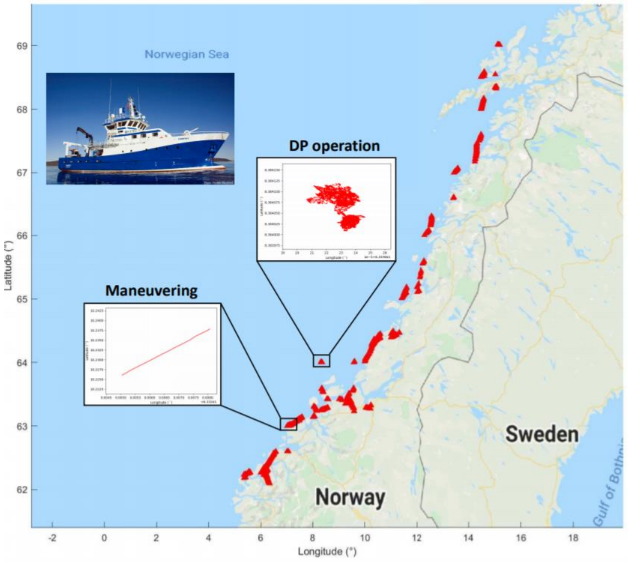
- The relationship between wave and ship motion is hard to obtain accurately.
- The moving of ship adds extra complexity.

Goals:

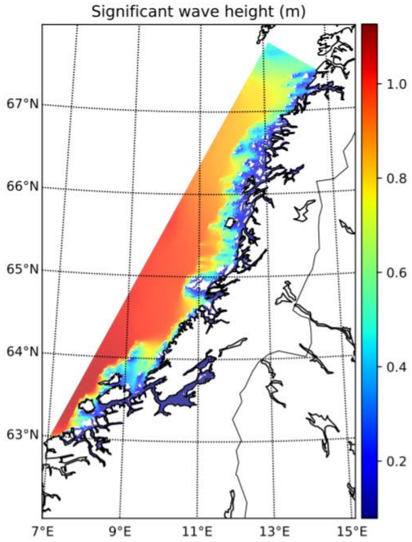
- Develop a model to estimate the on-site wave statistics using ship motion responses.
- Explicit feature representation.

Ship motion $\xrightarrow{\text{Mappings}}$ Sea state information

Data collection



Data match:
Nearest neighbor



Ship motion data:

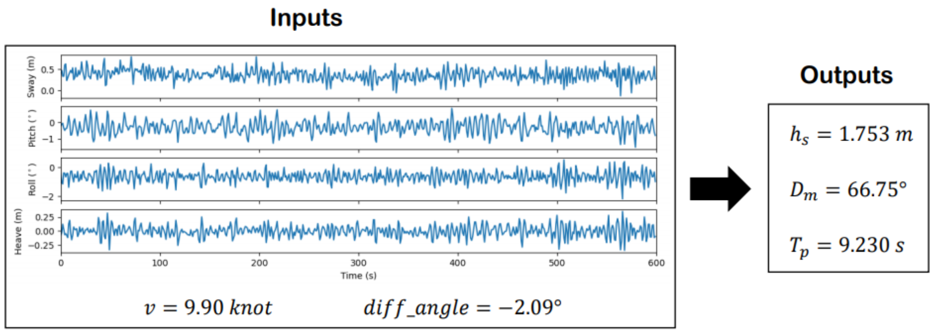
- R/V Gunnerus from 2017 to 2019.
- Stationary condition.
- *sway velocity*, *roll*, *pitch*, *heave*.
- Sampling frequency 1 Hz.

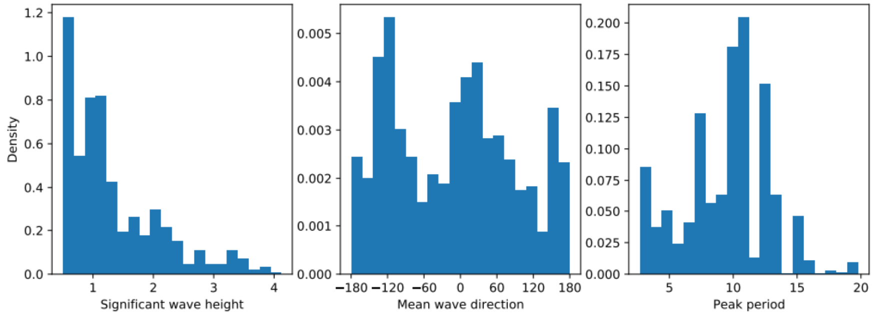
Wave data:

- Norwegian Meteorological Institute.
- *Significant wave height*, mean wave direction, *peak period*.

Data collection

- Data is cut into 10 minutes segment without overlapping.
- Significant wave height > 0.2m.

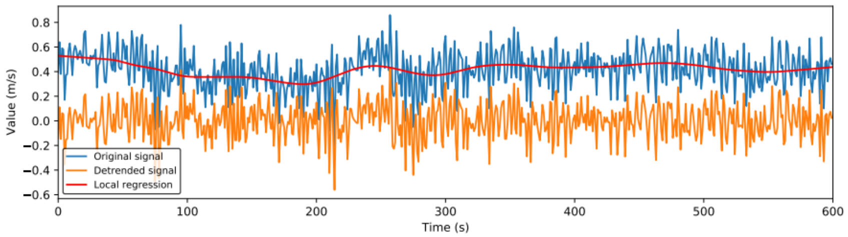




Data distribution

Methodology

1. Data pre-processing:



- Remove low-frequency motion.
- Remove measurement offset.

2. Feature extraction:

- **Domain-knowledge features:** forward speed (Doppler shift), course angle – heading angle.
- **Statistical features:** mean, variance, skew, kurtosis, etc.
- **Temporal features:** absolute energy, zero cross, autocorrelation, etc.
- **Spectral features:** use Welch method to transform the signal into frequency domain. Centroid, variation, fundamental frequency, etc.
- **Wavelet features:** wavelet transform is a time frequency analysis method which can 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.

Methodology

3. Feature selection:

Maximum relevance minimum redundancy (mRMR):

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

Where $I()$ is mutual information, S is the feature set, y is the target.

4. Model development:

- **K-nearest neighbour regression (kNN):** predict a testing point based on a fixed number k of its closest neighbors in the feature space.
- **Support vector regression (SVR):** same principle as the SVM (maximum-margin). RBF kernel is used in this paper.
- **Gradient boost decision tree (GDBT):** an ensemble model using gradient boost technique with decision trees as base learners.

Ensemble (voting): averaging the results from above three models.

Experimental results

The experimental results are evaluated with **5-fold cross-validation** using **mean absolute error (MAE)**.

1. Baseline comparison:

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	$\mathbf{1.528 \pm 0.084}$

Models:

- EN: Linear regression with elastic net regularization
- MLP: Multi-layer perceptron
- RF: Random forest
- kNN: k-nearest neighbor
- SVR: Support vector regression
- GDBT: Gradient boost decision tree
- SeaStateNet: Deep learning model proposed on ICRA2019

Experimental results

2. Ablation study:

Evaluation on the constructed features

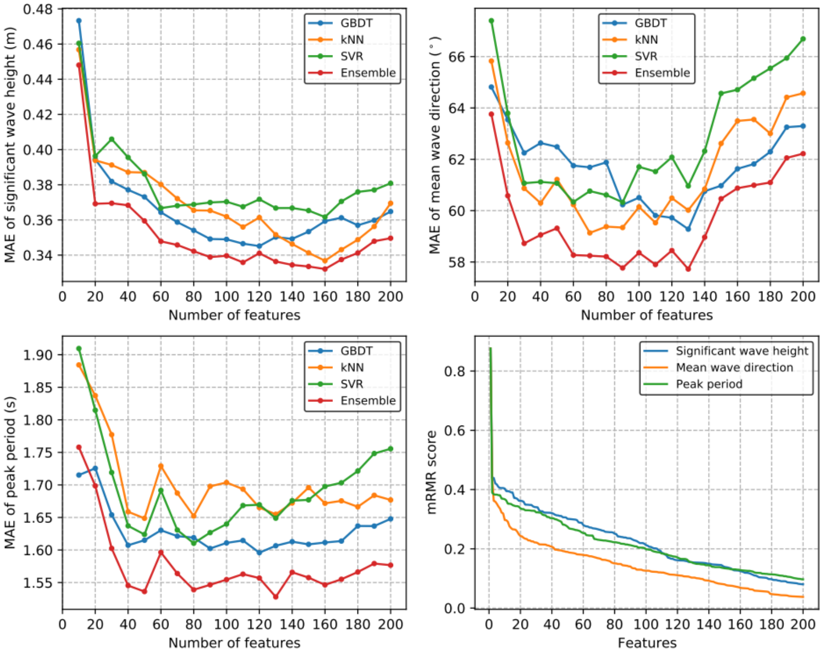
f_1: Domain-knowledge features
f_2: Statistical features
f_3: Temporal features
f_4: Spectral features
f_5: Wavelet features

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
$f_1+f_2+f_3+f_4+f_5$	$\mathbf{0.334 \pm 0.030}$	$\mathbf{57.72 \pm 1.30}$	$\mathbf{1.528 \pm 0.084}$

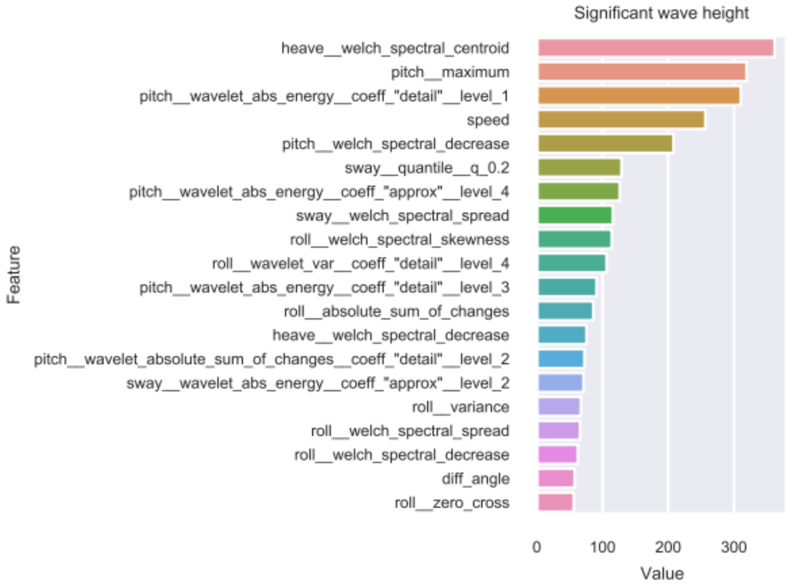
Evaluation on the number of used features



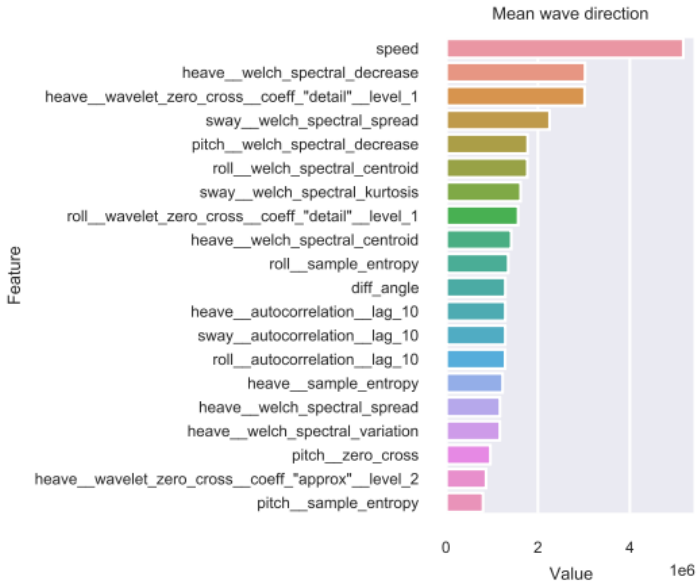
Experimental results

3. Feature importance:

Feature importance can be measured by the **total Gini gain** from the GDBT model.

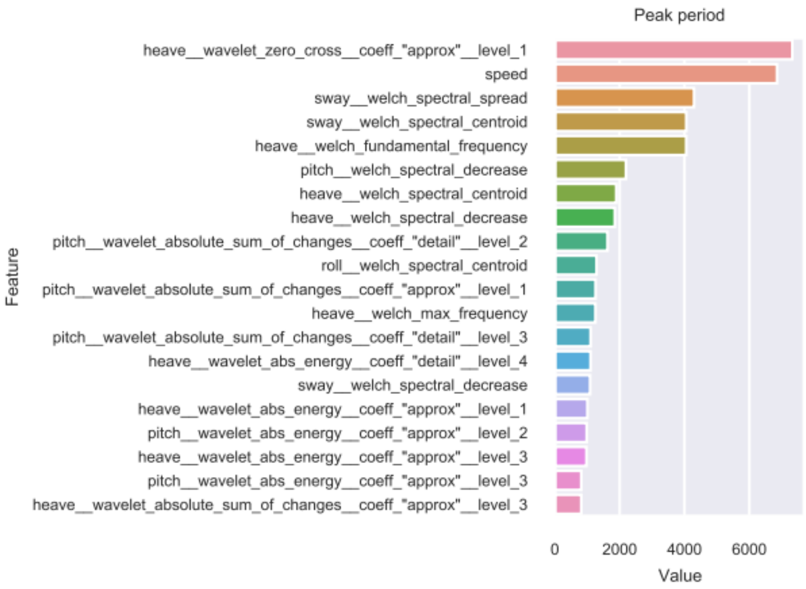


Feature type: Strength of the signal



Feature type: Shape of the spectrum

Experimental results



Feature type: wavelet features, spectral shape

Conclusions

- A machine learning model is developed for estimating the sea state information from measured ship motion responses.
- The developed model with hand-craft features archive similar and even slightly better performance than the deep learning model.
- The extracted features are explainable, which can be used to build a explainable model in the future.

Thanks