Estimation of Stress on Ship Structures Using Full-Scale Measurement Data and Machine Learning

Hideki MIYAJIMA*

1. INTRODUCTION

In the full-scale measurement project, various data such as navigation data (main engine speed, ship speed, heading, etc.), weather data (wind, waves, etc.), and ship motion data (acceleration, stresses on ship structure, etc.) are obtained and accumulated to understand the state of the ship during navigation. These data are used to assess structural strength, estimate life by fatigue strength assessment, and provide feedback for design [1-3]. From the viewpoint of ensuring the safety of ships, it is important to understand the history of the stresses generated on ship structures.

One of the problems of stress measurement in full-scale is that installation and maintenance of sensors are costly. Since it is difficult to measure all the measurement points that are in demand, it is desirable to have a method to grasp the stress of the whole ship with fewer measurement points. As an approach to understand the stress history, which is different from the full-scale measurement, research on “load and structural consistent analysis” has been conducted to estimate the stress generated on ship structure by incorporating structural analysis. However, since there is no established method for stress estimation, there is room to consider new approaches.

If the estimation of stress generated on ship structure is considered as a regression problem, an approach using machine learning, which has been developed in recent years, is considered to be effective. Since machine learning can make estimations considering various factors related to the problem, the stresses generated on the ship structure can be estimated by using the stress-related data obtained from full-scale measurements.

The data obtained from full-scale measurements include data that are affected by the natural environment such as wind, waves, and currents. It is difficult to grasp the weather and ocean conditions accurately, and the full-scale measurement data contains many uncertain measurement values. In the field of machine learning, Natural Gradient Boosting (NGBoost) [4], a method for estimating probability distributions, has been proposed as an effective method for making numerical estimations based on such data. By using NGBoost, it is expected to make reasonable numerical estimations considering probability distributions. Therefore, in order to confirm the feasibility of stress estimation using a new approach, research on stress estimation on ship structures using full-scale measurement data and NGBoost has been conducted. In this paper, the contents of our research is introduced.

2. OVERVIEW OF FULL-SCALE MEASUREMENT

In this study, from the viewpoint of the measurement items and the number of data, the data for about two years obtained in the full-scale measurement project on the 8,600 TEU container ship is used. Table 1 shows the main particulars of the ship and Table 2 shows the measurement items of the ship.

The Sensors to measure acceleration and stress are installed on the ship. The locations of the sensors are shown in Fig. 1. The Optical Strand Monitoring System (OSMOS) sensors, which uses optical fiber to measure the strain of structural members, was used to measure the stress. And OSMOS sensors were installed in 12 locations, four in each of the three cross sections of the ship. Three-axis (x, y, z) accelerometers were used to measure acceleration, and were installed in three locations: fore part, midship part, and aft part. ERA-5 wave hindcast data provided by the European Centre for Medium-Range Weather Forecasts (ECMWF) is used for the wave data. Note that the ship's regular route was changed during the measurement period.

* Research Institute, ClassNK

Table 1　Main particulars of the ship

Length overall (L_{OA})	Abt. 334.5 m
Breadth	45.6 m
Depth	24.4 m
Design draft	14.0 m
Gross tonnage	Abt. 97,000 GT

Table 2　Measurement items

Data	Contents
Navigation	Ship's speed (through water, over the ground), Course over ground, Main engine speed, Power of main engine
Weather	Wind direction, Wind speed
Acceleration	3-axis (x, y, z) for fore, midship and aft part
Stress	Hull girder stress
Wave	Wave height, Wave direction, Wave period, Directional width of wave, Kurtosis of wave, Relative width of wave frequency spectrum, Water depth

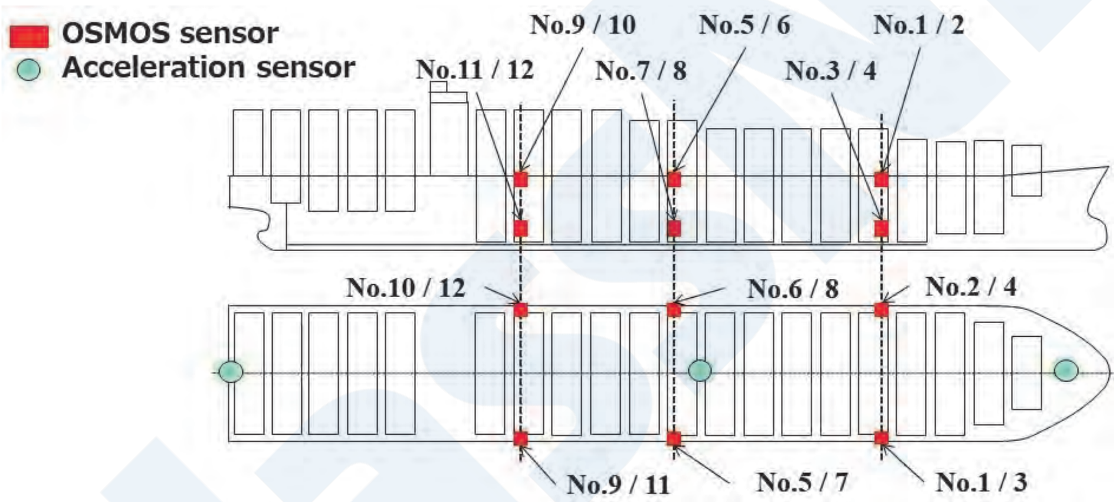


Figure 1　The positions of OSMOS sensors and acceleration sensors

3.　STRESS ESTIMATION METHOD

3.1　Estimation Target

It is important to estimate the stress generated on the upper structure because container ships have large openings that cause high stress in the ship structure. In addition, the midship part of the ship is important because the stress is relatively high among the measurement points. Therefore, in this study, the stress measured at the midship part and on the port side of the deck (No. 6 in Fig. 1) is used as the estimation target.

3.2　NGBoost

In this study, considering the uncertainty of the full-scale measurement data, NGBoost, which is a method that can estimate the probability distribution, is used to estimate the stress generated on the ship structure. NGBoost is a regression model adapted gradient boosting for estimation of probability distributions. Gradient boosting is a type of ensemble learning method that creates one learner by combining multiple weak learners with low estimation accuracy. A feature of NGBoost is that it uses natural gradients [5] to improve the learning efficiency of the weak learner in order to estimate the multiple parameters of probability distributions simultaneously.

In this study, the log-likelihood $L(\theta, y)$ shown in Equation (1) is used as the loss function which is optimized in the training phase of NGBoost. The natural gradient $\tilde{\nabla}L(\theta, y)$ is shown in Equation (2). Here, the parameters of the probability distribution are θ, the correct answer label is y, and the probability distribution P, and $I_L(\theta)$ is the Fisher Information matrix. The normal distribution is used as the probability distribution, and the decision tree is used as the weak learner.

$$L(\theta, y) = -logP_\theta(y) \quad (1)$$

$$\tilde{\nabla}L(\theta, y) \propto -I_L(\theta)^{-1}L(\theta, y) \quad (2)$$

4. STATISTICAL PROCESSING OF FULL-SCALE MEASUREMENT DATA

In this study, using the full-scale measurement data processed as hourly statistics, the data set for stress estimation was created. The processing of each data and the data set created are described in this chapter.

4.1 Navigational Data, Weather Data, Wave Data

Navigational data and weather data were processed as hourly averages. For wave data, hourly wave hindcast data were used. The data representing the angle of wave direction, wind direction, etc., takes the value of 360 degrees clockwise with the bow direction as 0 degrees. Therefore, the amount of change in angle from the bow direction was added as the variables.

4.2 Stress Data

The stresses in the ship structure can be separated into two major components by frequency analysis: wave response component and elastic response component, with peaks around 0.1 Hz and 0.5 Hz, respectively. In this study, the wave response component from 1/60 to 0.3 Hz, the elastic response component from 0.3 Hz to 1.0 Hz, and the wave and elastic response component from 1/60 to 1.0 Hz were separated from the hourly stress time series data. Then the stress range per wave was calculated by the zero-up crossing method, and the standard deviation was calculated.

4.3 Acceleration Data

Frequency analysis showed peaks at around 0.1Hz and 0.6~0.8Hz. These components are wave response component and mainly elastic response component, respectively [2]. In this study, the wave response component from 1/60 to 0.3 Hz, the elastic response component from 0.3 Hz to 1.0 Hz, and the wave and elastic response component from 1/60 to 1.0 Hz were separated to include the peaks, and the maximum values and standard deviations for each hour were calculated and used as variables. The maximum value and standard deviation for each hour were used as the features.

4.4 Data Set for Stress Estimation

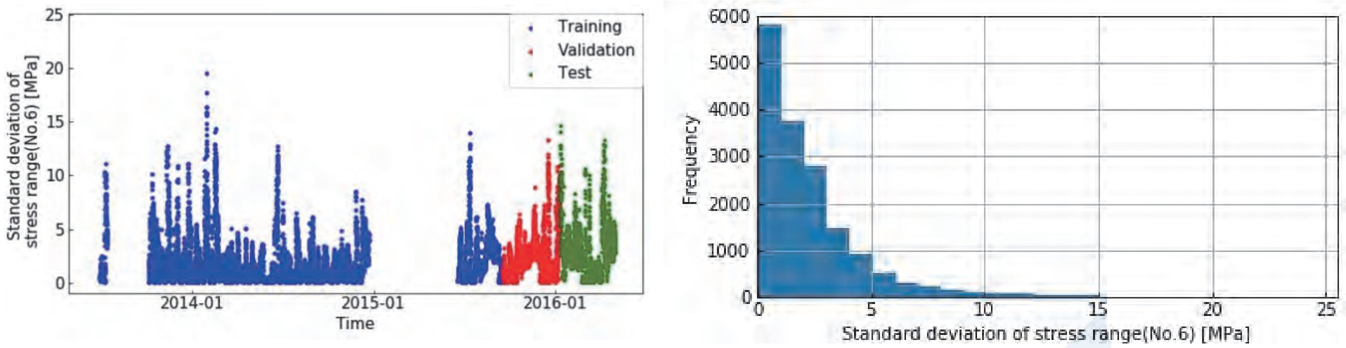
After the above process, the data set consisting of one objective variable (the measured value of stress at No. 6) and 108 explanatory variables was created. Table 3 shows the number of data points, and Fig. 2 shows the time series plot and histogram of the objective variable.

From Fig. 2 (b), it was confirmed that there were few data where high stress was measured. In the case of machine learning using such imbalanced data, there is a concern that the accuracy of estimation model will decrease in the areas where the number of data is small. Therefore, in order to compensate for the bias of the imbalance data, oversampling using SMOTE [6], a method of oversampling that increases the data with a small number of samples, is conducted.

SMOTE is a method of increasing data using the k-nearest neighbor method, which interpolates new data using specific data belonging to a minority group and randomly selected data from k of its neighbors. In this study, a threshold was set for the stress as the objective variable, and the data were labeled as above and below the threshold, and oversampling was performed so that the majority and minority groups had the same number of data.

4.5 Stress Estimation Model

In this study, two estimation models were developed, one using all explanatory variables (Case 1) and the other using only navigation, weather, and wave data (Case 2), as shown in Table 4. Note that, in Case 2, oversampling was not performed because it tended to reduce the estimation accuracy. The number of explanatory variables in Case 2 is smaller than in Case 1. Therefore, the pattern of data included in the training data decreased, and the number of similar data increased, which may have caused this problem.



(a) Time series plot (b) Histogram

Figure 2 Time series plot and histogram of the stress at No. 6

Table 3 Number of data in the data set

	Number of data
Training data	11,435
Validation data	2,614
Test data	2,288
Total	16,337

Table 4 Explanatory variables to use

Data	Case 1	Case 2
Navigation	○	○
Weather	○	○
Acceleration	○	○
Stress	○	—
Wave	○	—

5. ESTIMATION RESULTS

5.1 Comparison of Estimation Results in Case 1 and Case 2

Figure 3 and Fig. 4 show the results of stress estimation for the test data in Case 1 and Case 2, respectively. The mean of the normal distribution is plotted with the x-axis as the measured value and the y-axis as the estimated value. When the measured and estimated values of stress are equal, they are plotted on a reference line drawn diagonally. The 2σ confidence interval of the estimated normal distribution is also shown. Table 5 shows the mean squared error and correlation coefficient of the measured and estimated values, and the mean value of the estimated standard deviation.

For Case 1, Fig. 3 shows that the plots of the estimation results are distributed along the reference line, and the value of the correlation coefficient is about 0.99. Therefore, the measured values and estimated values are considered to be in good agreement with each other.

For Case 2, Fig. 4 shows that the plots are distributed near the reference line. On the other hand, from Table 5, it is confirmed that the value of the correlation coefficient decreased, and the value of the mean squared error and standard deviation increased in Case 2 compared to Case 1, which means that the estimation accuracy decreased. In Case2, the estimated value tended to be lower than the measured value in the high stress areas. From the viewpoint of the strength of the ship structure, it is not desirable for the estimated values to underestimate the stress, so it is also important to improve the estimation accuracy in the high stress areas.

The reason for the lower estimation accuracy in the high stress areas of Case 2 may be due to the lack of data measuring the high stress. The frequency of measuring high stress in full-scale measurements is low, and the weather and sea conditions that ships encounter vary depending on the route. In order to improve the estimation accuracy of high stress areas using only navigation, weather and wave data, it may be effective to collect measurement data for a longer period of time, and to create estimation model for each route.

5.2 Calculating the Importance of Explanatory Variables

When using decision tree-based machine learning methods, the importance of each explanatory variable to the estimation results of the created estimation model can be calculated. Since the stress estimation model using NGBoost outputs the mean and standard deviation, which are parameters of the normal distribution, as estimation results, the importance can be calculated for each of them. As part of the importance of the explanatory variables, Table 6 and Table 7 show the top five in Case 1 and Case 2, respectively.

In Case 1, the importance of stresses at No. 1, No. 5, No. 9, and No. 10, which were measured near the objective variable (stress at No. 6) or at the upper part of the ship, were high. On the other hand, the importance of the stresses at No.3, No.4, and No.7, which are located in the lower part of the ship and bow side, was lower than other measurement points. In this study, the measurement point No. 6 was targeted. The accumulation of knowledge on other measurement points and on ships for estimation is expected to lead to the clarification of measurement points necessary for stress estimation and to the accurate estimation of stress at multiple points with a small number of measurement points.

In Case 2, the importance of wave height, wave period, and wave direction related to wave load increased. In particular, the wave height has a linear relationship with the stress at the measurement location No. 6, as shown in Fig. 5, and is considered to contribute significantly to the estimation.

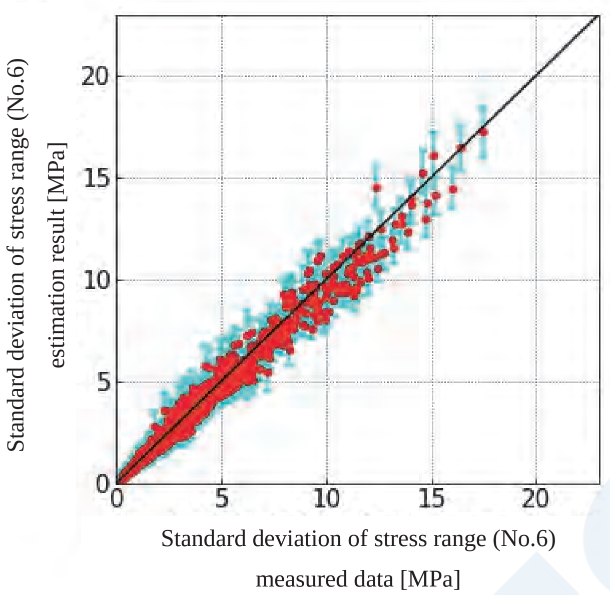


Figure 3　Measured value and estimated value: Case 1

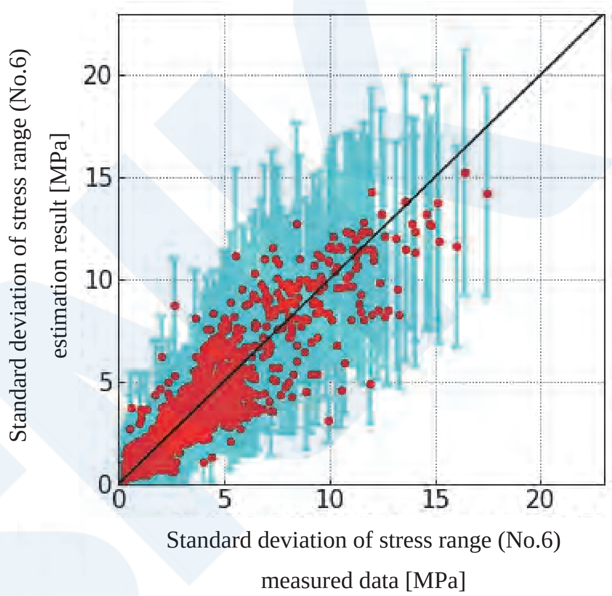


Figure 4　Measured value and estimated value: Case 2

Table 5　Summary of estimation results

	Case 1	Case 2
Mean squared error	0.13	1.10
Correlation coefficient	0.99	0.91
Standard deviation (mean)	0.19	0.74

Table 6　Importance of explanatory variables: Case 1

	Mean	Standard deviation
1	Stress (No.5) wave and elastic response component	Stress (No.5) wave response component
2	Stress (No.1) wave and elastic response component	Stress (No.5) wave and elastic response component
3	Stress (No.9) wave and elastic response component	Stress (No.10) wave and elastic response component
4	Stress (No.10) wave and elastic response component	Stress (No.12) wave and elastic response component
5	Stress (No.10) wave response component	Stress (No.9) wave and elastic response component

Table 7 Importance of explanatory variables: Case 2

	Mean	Standard deviation
1	Wave height	Wave height
2	Wave period	Wind speed
3	Wind speed	Wave direction
4	Ship’s speed through water	Main engine speed
5	Power of main engine	Ship’s speed through water

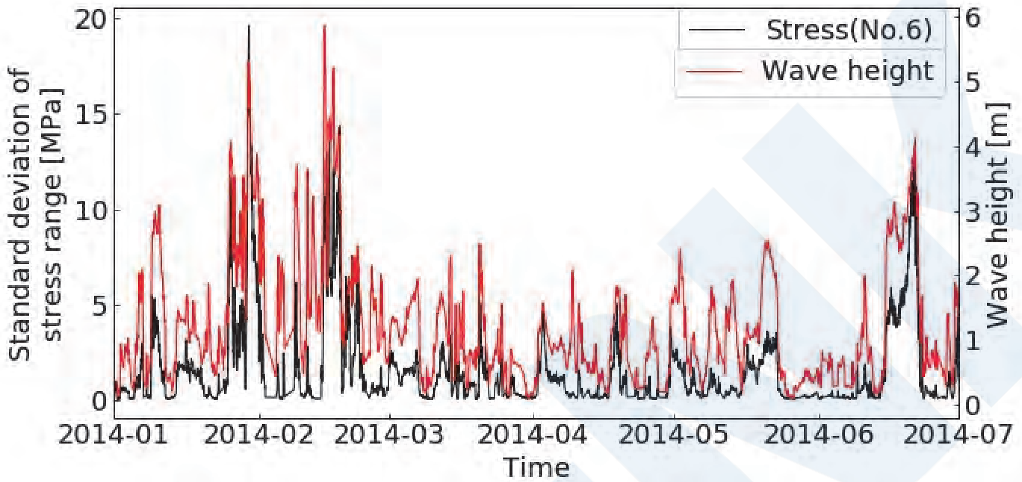


Figure 5 Time series data of the stress at No. 6 and wave height

6. CONCLUSION

In this paper, research on estimation of stress generated on ship structures using full-scale measurement data and machine learning.

Since this study was conducted for one specific ship, it is necessary to confirm the effectiveness and versatility of this method through comparison and verification with the stress estimation results when this method is applied to other ships. In addition, through such efforts, it is expected to obtain knowledge about the number of data required to secure a certain estimation accuracy and the explanatory variables that are effective in improving the estimation accuracy.

In this study, stress estimation was conducted for the point where stress was measured in full-scale measurement, using the measured values as the correct data. On the other hand, there are many areas where there is a need to estimate stress generated on the ship structure, and it is difficult to install sensors in all points and obtain the measured values. Therefore, as a future work, it will be worthwhile to establish a method to estimate the stress history accurately without relying on direct measurement.

ACKNOWLEDGMENT

The author wishes to express his profound gratitude to Mr. Ihara of Miotsukushi Analytics Co., Ltd. for his cooperation in this research.

REFERENCES

1) Toshiki Kawai, Nao Ou, Yasumi Kawamura, Tetsuo Okada, Taiga Mitsuyuki, Xi Chen : A Study on Sea State Estimation Using Monitoring Data of Large Container Ships－Estimation of Directional Wave Spectrum by Machine Learning－, Journal of the Society of Naval Architects of Japan, Vol. 28, 2019 (in Japanese)

2) Katsutoshi Takeda, Tsutomu Fukui, Toshiyuki Matsumoto : A Study on Hull Girder Response Evaluation for a Large Container Ship by Full-Scale Measurements and Direct Load Structure Analysis, Journal of the Society of Naval Architects of Japan, Vol. 29, 2019 (in Japanese)
3) Masahiko Fujikubo : R&D of Digital Twin for Ship Structures- Outcome of Phase-1 Project and its Future Extension-, Journal of the Society of Naval Architects of Japan, Vol. 30, 2020 (in Japanese)
4) Duan, T., Avati, A., Ding, D. Y., Basu, S., Ng, A. Y., Schuler, A.: NGBoost: Natural Gradient Boosting for Probabilistic Prediction, https://arxiv.org/abs/1910.03225, 2019
5) Amari, S.: Natural Gradient Works Efficiently in Learning, Neural Computation Vol. 10, 1998
6) Chawla, N., Bowyer, K., Hall, L., Kegelmeyer, W.: SMOTE: Synthetic Minority Oversampling Technique, Journal of Artificial Intelligence Research 16, 2002