Analysis and Modeling of Sensor Data for Ship Motion Prediction

Guoyuan Li* and Houxiang Zhang *Faculty of Maritime Technology and Operations*
Norwegian University of Science and Technology
Postboks 1517, N-6025, Aalesund, Norway
{guoyuan.li, hozh}@ntnu.no
Bikram Kawan, Hao Wang, Ottar L. Osen and Arne Styve *Faculty of Engineering and Natural Sciences*
Norwegian University of Science and Technology
Postboks 1517, N-6025, Aalesund, Norway
bikram.kawan@stud.ntnu.no, {hawa, ottar.l.osen, asty}@ntnu.no

Abstract

This paper presents an analyzing and modeling scheme that can dig into the ship sensor data to generate models for ship motion prediction. As the raw sensor data contains information that is noisy and discontinuous, three data cleaning methods used for noise reduction, data continuing and resampling are developed. According to specific ship applications, various ship motion constraints are designed to filter and establish meaningful data sets which will be used in the modeling step. Correlation analysis within the data sets is performed to figure out how significant the sensor data contributes to the predictive target. An flexible neural network based modeling mechanism is implemented, in which the user can design the model based on the correlation analysis. By training and testing the neural network using the generated data sets, the ship motion predictive model is obtained. Through a case study of fine maneuvering, the scheme is verified efficient in analyzing and modeling sensor data for ship motion prediction.

Index Terms

Data analysis, Neural network, Modeling, Ship motion prediction.

I. Introduction

A ship in an open sea is a very complex dynamic system, which is affected by its control systems such as propulsion and steering systems, and external perturbations produced by wind, waves, and sea currents [1][2]. To better perceive the ship behavior, usually there are many sensors installed in the ship in various positions. Some data from sensors are used in real time for maneuvering and related actions and some sensors are placed in sensitive area like propeller blade to collect the data for future purpose such as system diagnosis. The data on the vessel will be stored for years in huge size, forming a data set in terms of big data [3]. In addition, there are emerging demands from ship owners and companies to random access time series of the huge data set and generate ship motion models for certain specific requirements of marine operations. Therefore, how to effectively dig into the data set and find out valid ship motion models is challenging but of great importance.

* Corresponding author

Data analysis and modeling has been studied for decades. Many reliable technologies have been developed in the last few decades, ranged from Kalman filtering, Markov dynamic models, to non-parametric statistical methods [4]. In many applications, such as time series data analysis, it is possible to estimate models through historical data from a given state to predict its future behavior. Models are thus expressed as transfer functions or in terms of state-space parameters [5]. A lot of inductive reasoning techniques and algorithms such as fuzzy logic, Bayesian network and regression analysis can be used for creating predictive models. The results have been successfully applied to different areas including weather forecasting [6], mobile user movement prediction [7] and electric load prediction [8].

For ship maneuvering, health, safety, environment, security and cost will be given high priority during maritime operations. Ship data analysis and modeling are therefore essential for ship maneuvering, especially for the emergence of new demands in offshore operations. In the literature, relevant data analysis and modeling techniques have been developed. On the one hand, attempts have been made to estimate and identify the dynamic parameters for ship maneuvering to complete the control model [9]-[11]. On the other hand, efforts have been done to accurately estimate and predict the ship motion in ocean navigation, such as trajectory estimation [12][13], rolling motion prediction [14] and thruster forces prediction [15]. Even though these techniques are available for ship motion prediction, more concerns about data analysis and modeling in a big data manner are urgently needed for today's maritime applications.

Our on-going project aims to combine data mining techniques with modeling methods to design and implement a ship motion prediction framework for different complicated maneuvering cases. The framework will extract and purify vast amounts of sensor data to model and predict ship motion, including position, speed, course, pitch and rolling. By collecting and mapping diverse sources of sensor data and further modeling and visualizing the data of interest, the framework will provide users with an intuitive way to identify useful associations between data and shows how sensitive the sensor data is to affect the accuracy of these predictive models. In addition, the framework also aims to be flexible enough to be able to plug-in other user customized predictive models for results comparison. In this paper, we will only focus on the ship position prediction and highlight the ease-of-use of the system from raw sensor data to final predictive model verification.

The rest of the paper is organized as follows. Section II shortly introduces the overall structure of the ship motion prediction system. In Section III, all relevant process units including data cleaning, data analysis and modeling and data visualization are described in detail. Section IV presents case studies and evaluation results. Conclusion and future work are shown in section V.

II. Design of Ship Motion Prediction Structure

This section briefly describes the design of the ship motion prediction system based on real sensor data for analyzing and modeling the position of the ship. Due to complex dynamic nature of sea, the ship behavior is very hard to predict. Our partner in Norway therefore started to collect on-board ship sensor data long time ago and intended to create robust predictive models to improve ship maneuvering technologies.

What we used in this project is three years of raw data from disparate source of sensors on one vessel. Four types of data modules, with a sampling frequency from up to 4000Hz to 1Hz, will be utilized for ship motion prediction. The high sampling frequency data modules are mostly used for propulsion system analysis, e.g., the vibration, the torque, the bending moment. Whereas the low sampling frequency data modules are the collection of ship-environment status, as shown in Table I. More specifically, M1 is the ship's extrinsic representation caused by the intrinsic control parameters like shaft rotational rate and pitch angle from M2. It is noted here that for ship trajectory prediction purpose, only the low sampling

TABLE I

SPECIFICATION OF LOW SAMPLING FREQUENCY DATA MODULES

Module	Frequency	Parameter	Unit
M1	1Hz	Speed	[m/s]
		Position	[m, m]
		Heading	[deg]
		Roll	[deg]
		Pitch	[deg]
		Yaw rate	[deg/s]
		Roll rate	[deg/s]
		Pitch rate	[deg/s]
M2	1.65Hz	Rotational speed	[RPM]
		Drive of motor	[W]
		Propeller pitch	[deg]

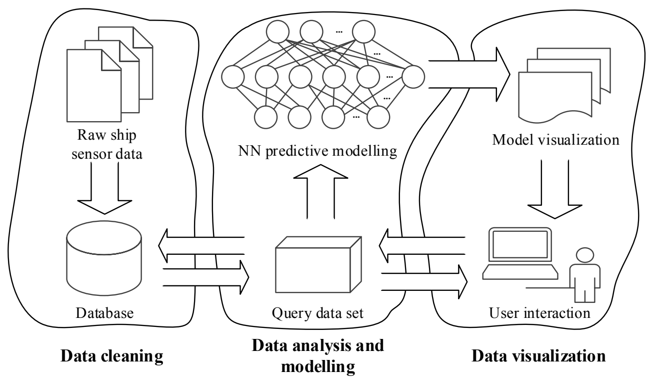

Fig. 1. System structure for ship motion prediction.

frequency data modules are used. The high sampling frequency data modules are only referred for data cleaning thereafter in this paper.

Fig. 1 illustrates the overall system structure for ship motion prediction. The design follows most requirements from our industrial partner. It consists of three components:

- Data cleaning
- Data analysis and modeling
- And data visualization

Data cleaning for redundant sensor data representation, different sampling frequency, as well as sensor noises needs to be done before the transmission of the data set into a database. Through the Graphic User Interface (GUI), users can make a query to select the interested time series of data set. The next step is to analyze and determine proper types of measured data according to the specific application. Some artificial intelligence methods, e.g. Neural Network (NN), are used to train and test predictive models. An animation for reproduction of the ship movements simultaneously with predicting plot is realized in the model visualization component, which helps to check the accuracy of the predictive model. In addition, users can interact with the predictive model by setting related parameters, e.g., the learning rate and the time period involved in prediction, to observe and evaluate its effect on the prediction result.

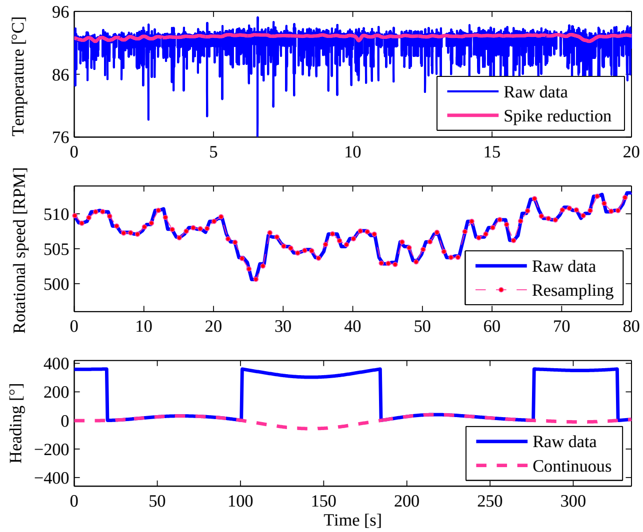

Fig. 2. Three types of data cleaning.

III. Implementation of Ship Motion Prediction System

A. Data Cleaning

Considering the raw data may contains noisy, discontinuous and redundant information, it is necessary to clean the data so that its affection on further analysis and modeling can be minimized.

Noise reduction is the first step in data cleaning process. Noise sources can be either internal or external. For internal noise source, the noise can be removed through statistical estimation. But for external noise sources, sensor noise due to gradients, nonhomogeneous mediums and other influences is unavoidable in most cases. In our case, the major sensor noise is external and mainly from high frequency data modules. An example is the temperature measurement in the engine room. Because of external noise like cables and the coupling of electric and magnetic fields, the measured temperature is full of spikes. A natural way to eliminate them is to use median filtering technique [16]. After three iterations of median filtering, the spikes are successfully removed and the series of temperature data becomes smooth, as seen in the top panel in Fig. 2.

Another necessary data cleaning process is resampling. Because both the data modules in Table I with sampling frequency of 1.65Hz and 1Hz respectively are involved for ship motion prediction, it is beneficial to synchronize them before analysis and modeling. The middle panel in Fig. 2 gives an example to resample the shaft rotational speed from 1.65Hz to 1Hz. To preserve the power of signal and remove aliasing effect, the downsampling is performed in three steps:

(1) Make interpolation to increase the sample frequency to M times of the desired frequency;
(2) Reduce the high-frequency signal component with a lowpass filter to limit aliasing effect;
(3) Downsample the filtered signal by M.

The last type of data cleaning is related to the continuity of the raw data. As mentioned in II, the raw data is a collection of different sensors on the vessel for three years. Nevertheless, some of the sensors are not always active due to maintenance and other safety reasons, which consequently leads to the raw data not continuous. Therefore, we have to split the raw data into continuous segments to facilitate further analysis. In addition, raw data may contain jumping phenomenon due to data definition. The bottom panel of Fig. 2 is an example for heading recording. Because the definition of heading is within

$[0°, 360°]$, it is inevitable to occur jumping phenomenon when the heading changes near the border. To remove this type of discontinuity, the heading angle is defined unbounded and a variable for the number of laps is introduced:

Pseudo code

```
procedure
    Given the jumping threshold Θ
    lap_num ← 0
    last_heading ← 0
    i ← 0
loop:
    heading(i) ← heading(i) − 360 ∗ last_heading
    if heading(i) − last_heading > Θ then
        lap_num ← lap_num + 1
        heading(i) ← heading(i) − 360
    else if heading(i) − last_heading < −Θ then
        lap_num ← lap_num - 1
        heading(i) ← heading(i) + 360
    last_heading ← heading(i)
    i ← i+1
    goto loop
```

After dealing with the three types of data cleaning, the whole data set is then imported into a database.

B. Data Analysis and Modeling

The data analysis and modeling component is the core of the scheme. It bridges the gap between the user and the generated database for better understanding and modeling the data set. The user in this stage plays the key role in analyzing the correlation of parameters and generating meaningful model for specific applications.

There are two potential ways to help users to analyze the data. First, the scheme provides the user with an interface to optimize the data set. This idea takes advantages of the fact that users who analyze the data set to generate the model for a specific application have expertise to establish a subset of the database. By properly selecting relevant parameters and filtering according to expert knowledge, a "query data set" will be formed through user interaction, as shown in Fig. 1. For instance, to generate predictive model for fine maneuvering, i. e., maneuvering in limited working space with a low speed, a speed constraint together with a position constraint can be added into the query. The resultant "query data set" will thus be more precise to be used for further modeling the ship motion. So far, we have implemented the constraints for all the parameters listed in Table I.

Second, the correlation between the measured parameters provides the user with an intuitive representation for data analyzing to some extent. For prediction purpose, it is more important to find out how significant the measured parameter to the predictive output compared to the correlation between two measured parameters. Sensitive analysis is therefore applied here to quantitatively determine how much the parameter can contribute to the changes of the output [17]. Fig. 3 is an example to illustrate the influence of parameters in M1 data module on the output of ship position. It is noted that if the significant value is smaller than the significance level of 0.05, the corresponding correlation coefficient is

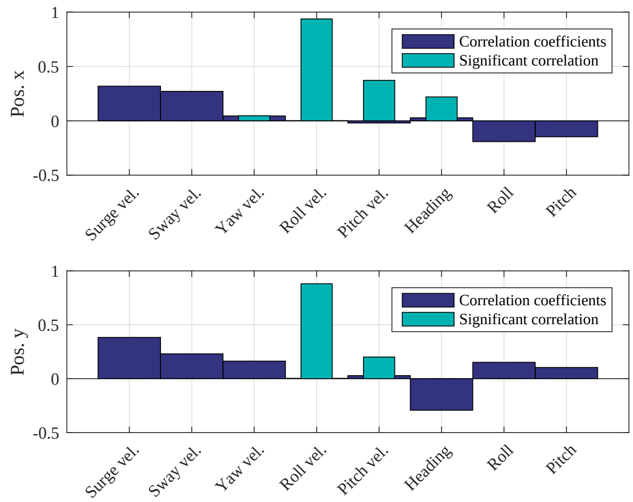

Fig. 3. Ship position correlation analysis.

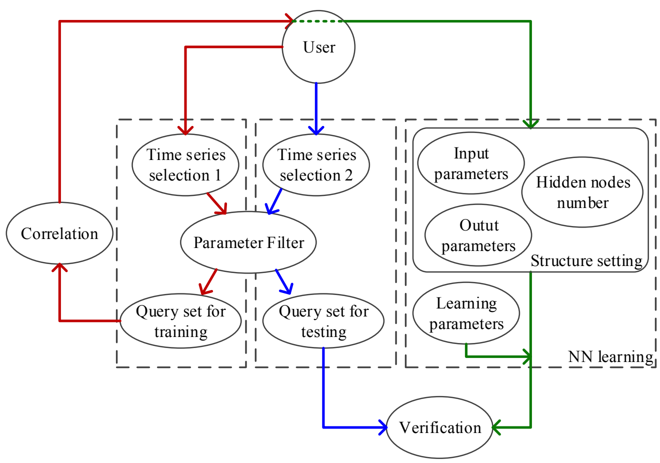

Fig. 4. Modeling procedure.

considered significant. Meanwhile, a higher absolute value of the correlation efficient indicates a stronger coupling between the two variables. From Fig. 3, the surge speed, the sway speed and the heading are deemed as the most significant parameters related to ship position.

In order to generate predictive models, we utilize a NN with three layers: the input layer, the hidden layer and the output layer. The concept used for modeling is primarily focused on giving maximum flexibility to users. To this end, the three layers are designed to be dynamic and changeable. The output parameters are selected according to the specific application. In addition, users can decide the number of hidden nodes, as well as determining the input parameters with the help of correlation analysis. The learning rate and the type of activation function are also optional for modeling. As a result of off-line training, the predictive model is established. To verify the model, users can further design a testing set by applying the same query to another time series and test the constructed NN. Fig. 4 illustrates the modeling procedure through user interaction.

C. Data Visualization

Data visualization involves two parts. One is the user interaction and the other is the model visualization, both of which are closely associated with “data analysis and modeling”. To realize the interaction, a web application is developed based on PHP [18] and MySQL [19]. With PHP, a database-driven web site allows the content to reside in the MySQL database, and to be dynamically pulled from the database to create web pages for users to view via a regular browser. In the application, we have implemented the time line selection, the constraints filtering and the NN modeling and representation for user interaction.

To better understand the resultant predictive model, the model visualization is responsible for animating the ship motion, as well as the predictive position over time. By using WebGL technology, e.g. Three.js [20], 3D graphics can be rendered within any compatible web browsers without the use of plug-ins. A control panel is attached to the animation so that users can drag and drop the animation for rapid assessment. In addition, the predictive performance or any raw parameters listed in Table I can be plotted taking advantages of Plotly Chart API [21]. In this way, users can easily figure out whether the NN modeling and off-line learning procedure is good enough for certain type of ship motion prediction.

IV. Experiment

An experiment based on the three years of raw data has been carried out to verify the feasibility of the proposed scheme. The case study aims to extract and verify the predictive model for fine maneuvering applications, in particular for ships working in a limited working space with a low speed. The following steps show how we used the developed system to predict ship position.

First, we have to import the raw data into the system. Considering the case study, there is no need to utilize the whole raw data set. For simplicity, we split the raw data by date and took one day’s data in CSV format as the source of the case study.

Data cleaning is then required to apply to the selected data set. In our case, all the three data cleaning methods including noise reduction, data continuing and resampling were performed and the purified data set was stored to MySQL.

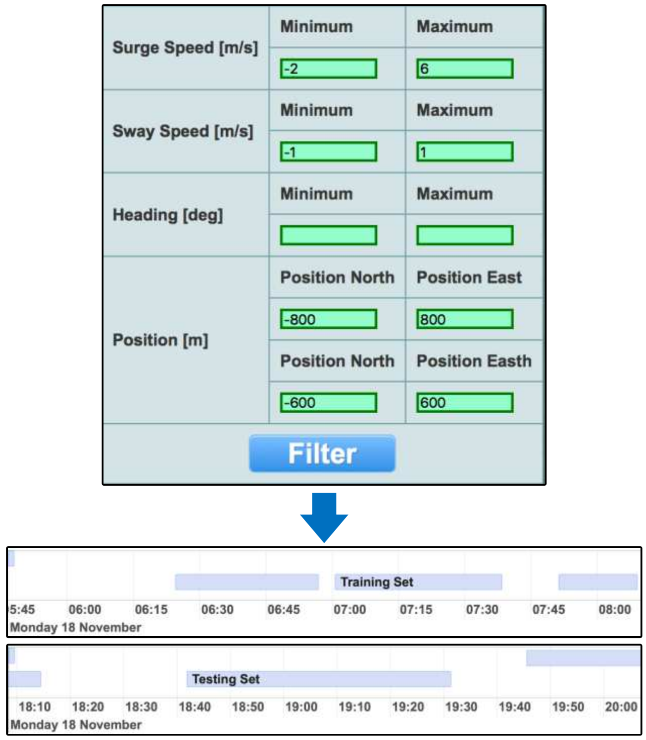

Fig. 5. Constraints and resultant data sets used for the position prediction during the fine maneuvering task.

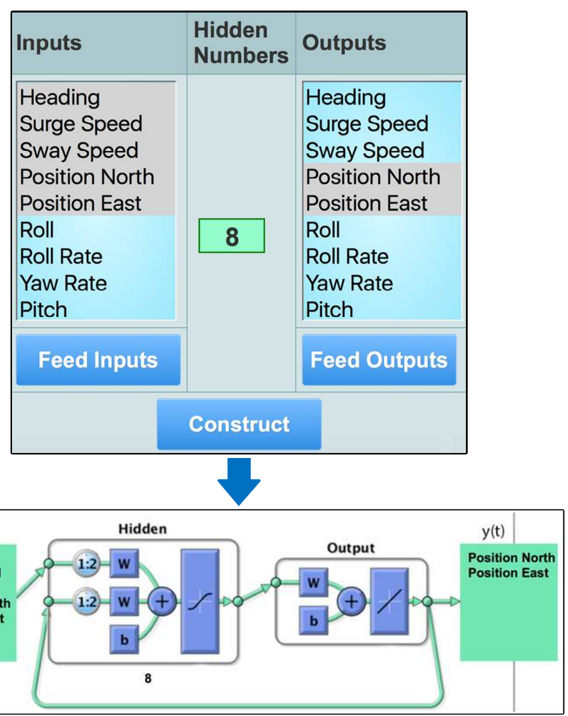

Fig. 6. Construct the NN based on the correlation analysis. ‘w’ and ‘b’ stand for the weight and the bias of NN. ‘1:2’ represents the time delay that allows the network to have a finite dynamic response to time series input data.

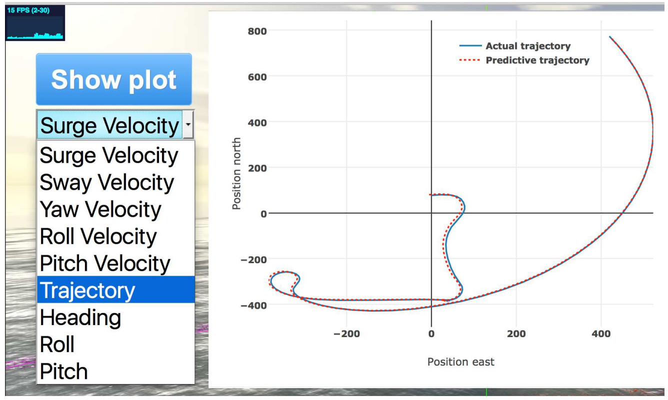

Fig. 7. Ship position prediction result.

According to the fine maneuvering task, we should further filter the data set. Fig. 5 illustrates the extra constraints that were used in the case study. Both the speed and the position constraints were considered. As a result, the data which satisfies the constraints were displayed on the timeline. We selected the data from time 7:02 to 7:37 as the training set and the data from time 18:42 to 19:32 as the testing set.

Once the data set is ready to use, it is time to construct the NN for off-line training. Fig. 6 shows how we set up the NN. First, the ship position was definitely selected as the NN outputs. Then, taking advantages of the correlation analysis from Fig. 3, the NN inputs were comprised of the ship speed, the heading direction, and the previous ship position. We took 8 hidden nodes for the NN construction. We also chose hyperbolic tangent as the activation function and set the learning rate to 0.1. By using the backpropagation algorithm on the training set, a matrix of NN weights was obtained.

Verification of the generated predictive model was done by applying the NN structure plus the matrix of weights to the testing set. The predictive results are automatically stored into MySQL and it is straightforward to compare between the actual and the predicted trajectory for the case study. For concise reason, part of the predictive positions for the testing set is illustrated in Fig. 7. The highlight here is that the prediction performance is straightforward. If the result is not satisfactory, the previous steps from NN modeling to verification can fast redo until reaching the expected prediction error. Besides the plot, animation for the test set was also generated, as shown in Fig. 8. From the experimental result, it reveals that the case study is complete and successful. As a result, we conclude the scheme works well in analyzing and modeling sensor data for ship motion prediction.

V. Conclusion

In this paper, a system structure from the import of the raw sensor data to the final verification of the generated predictive model for ship maneuvering applications is developed. First, the raw sensor data is purified and filtered to form a training set and a testing set according to the specific maneuvering task. Then, a correlation analysis is performed, enabling a better understanding of data association. Next, a flexible NN is constructed, from which the user can determine both the structure and the learning parameters. By applying the training set to the NN, a predictive model is obtained. After that, the testing set is used for verification. A case study about position prediction for fine maneuvering is carried out and the result shows the system is efficient in ship motion prediction.

Future work will focus on the modeling part, in particular that more modeling technologies will be concerned. Moreover, the prediction for other ship motions such as heading and rolling will be implemented.

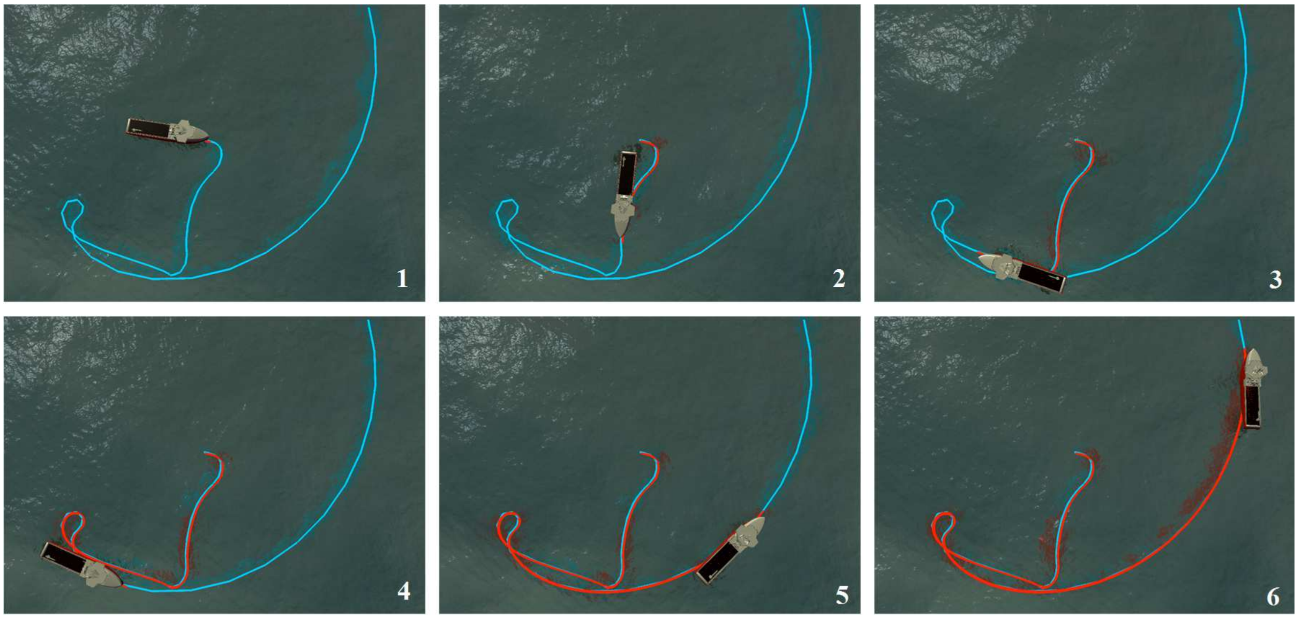

Fig. 8. Snapshots of the animation for ship position prediction. The blue and the red lines are the actual and the predictive trajectories, respectively.

Acknowledgement

This research is partially supported by the project "GCE BLUE Maritime Big Data" funded by Innovasjon Norge, and partially supported by the Mechatronics Lab at the Faculty of Maritime Technology and Operations, NTNU i Aalesund.

References

[1] T. I. Fossen, *Marine control systems: guidance, navigation and control of ships, rigs and underwater vehicles*, Trondheim: Marine Cybernetics, 2002.

[2] A. J. Sørensen, "A survey of dynamic positioning control systems", *Annual reviews in control*, vol. 35, no. 1, pp. 123-136, 2011.

[3] S. Kaisler, F. Armour, J. A. Espinosa and W. Money, "Big data: Issues and challenges moving forward", in *IEEE Hawaii Int. Conf. on System Sciences*, Hawaii, USA, pp. 995-1004, 2013.

[4] W. R. Blischke and D. P. Murthy, *Reliability: modeling, prediction, and optimization*, John Wiley & Sons, 2011.

[5] F. Salfner, M. Lenk and M. Malek, "A survey of online failure prediction methods", *ACM Computing Surveys (CSUR)*, vol. 42, no. 3, pp. 1-42, 2010.

[6] L. Al-Matarneh, A. Sheta, A, S. Bani-Ahmad, J. Alshaer and I. Al-oqily, "Development of temperature-based weather forecasting models using neural networks and fuzzy logic", *International Journal of Multimedia and Ubiquitous Engineering*, vol. 9, no. 12, pp. 343-366, 2014.

[7] S. Akoush and A. Sameh, A, "Mobile user movement prediction using bayesian learning for neural networks", in *Proc. of Int. Conf. on Wireless Communications and Mobile Computing*, New York, USA, pp. 191-196, 2007.

[8] T. Hong, M. Gui,M. E. Baran and H. L. Willis, H. L, "Modeling and forecasting hourly electric load by multiple linear regression with interactions", in *IEEE Power and Energy Society General Meeting*, Minnesota, USA, pp. 1-8, 2010.

[9] R. Skjetne, Ø. N. Smogeli, and T. I. Fossen, "A nonlinear ship manoeuvering model: identification and adaptive control with experiments for a model ship", *Modeling, Identification and Control*, vol. 25, no. 1, pp. 3-27, 2004.

[10] G. Li, W. Li, H. P. Hildre and H. Zhang, "Online learning control of surface vessels for fine trajectory tracking", *Journal of Marine Science and Technology*, DOI: 10.1007/s00773-015-0347-9, pp. 1-10, 2015.

[11] X. G. Zhang, and Z. J. Zou, "Identification of Abkowitz model for ship manoeuvring motion using ϵ-support vector regression", *Journal of Hydrodynamics, Ser. B*, vol. 23, no. 3, pp. 353-360, 2011.

[12] L. P. Perera and C. G. Soares, C. G, "Ocean vessel trajectory estimation and prediction based on Extended Kalman filter", in *Proc. of Int. Conf. on Adaptive and Self-adaptive Systems and Applications*, Lisbon, Portugal, pp. 14-20, 2010.

[13] T. Xu, X. Liu and X. Yang, “Ship trajectory online prediction based on BP neural network algorithm”, in *Int. Conf. on Information Technology, Computer Engineering and Management Sciences,* Jiangsu, China, pp. 103-106, 2011.
[14] J. C. Yin, Z. J. Zou and F. Xu, “On-line prediction of ship roll motion during maneuvering using sequential learning RBF neuralnetworks”, *Ocean engineering*, vol. 61, pp. 139-147, 2013.
[15] T. Abramowski, “Prediction of propeller forces during ship maneuvering”, *Journal of Theoretical and Applied Mechanics*, vol. 43, no. 1, pp. 157-178, 2005.
[16] Y. Liu, “Noise reduction by vector median filtering”, *Geophysics*, vol. 78, no. 3, pp. 79-87, 2013.
[17] D. M. Hamby, “A review of techniques for parameter sensitivity analysis of environmental models”, *Environmental monitoring and assessment*, vol. 32, no. 2, pp. 135-154, 1994.
[18] PHP manual [Online]. Available: `http://www.php.net/`. [Accessed: Feb. 22, 2016].
[19] MySQL [Online]. Available: `http://www.mysql.com/`. [Accessed: Feb. 22, 2016].
[20] Three.js [Online]. Available: `http://threejs.org/`. [Accessed: Feb. 22, 2016].
[21] Plotly chart [Online]. Available: `https://plot.ly/`. [Accessed: Feb. 22, 2016].