Chunlin Wang

Data analysis and modelling for onboard support of marine operations

NTNU
Norwegian University of
Science and Technology

Chunlin Wang

Data analysis and modelling for onboard support of marine operations

Thesis for the Degree of Philosophiae Doctor

Ålesund, August 2023

Norwegian University of Science and Technology
Faculty of Engineering
Department of Ocean Operations and Civil Engineering

NTNU
Norwegian University of
Science and Technology

Abstract

The ever-increasing exploration of ocean resources has led to more frequent and intensive marine operations. However, marine operations are vulnerable to accidents due to unpredictable environmental factors and human decisions. Therefore, providing onboard support for operators' decision-making is crucial to ensuring safe and sustainable marine operations. To achieve this, advanced sensors have been deployed on various offshore platforms to collect real-time data. As time goes on, this marine data accumulates and forms marine big data, characterized by five high Vs - volume, velocity, variety, veracity, and value. Marine big data speeds up the transition towards digitalization and automation, enabling onboard support for marine operations, such as ship motion prediction, path planning, and structural health monitoring.

The utilization of massive marine data to drive digitalization has become a significant topic in both research and industry. Data analysis and modelling offer a promising solution to address this issue. Data analysis is mainly divided into four categories: descriptive analytics, diagnostic analytics, predictive analytics, and prescriptive analytics. From a macro perspective, predictive analytics or modelling is a component of data analysis. Descriptive analytics answers the question:'What happened' while diagnostic analytics addresses the problem of 'why it happened'. Both these categories focus on discovering historical information in marine data. In contrast, predictive analytics and prescriptive analytics tend to analyze the future behaviours and events of a system by answering the questions: what will happen and how to make it happen. These four categories of data analysis involve a large variety of approaches, theories, and tools. Hence, they have been widely used to mine valuable knowledge and critical insights for onboard support of demanding marine operations.

Complex data types and different applications pose challenges to marine data analysis for onboard decision support. How to combine different analysis approaches to figure out these issues is the main concern in this dissertation. To show the importance of data analysis for onboard support of marine operations, three case studies are highlighted: ship dynamic positioning (DP) capability analysis, structural health monitoring, and Automatic Identification System (AIS) data analysis and modelling.

Ship DP capability is subject to the impact of environmental factors and the thruster's failure. Understanding the interaction between thrusters and environmental factors can provide support for DP capability improvement to prevent the occurrence of a loss of position. The objective of this study is to analyze the thrusters' significance under the influence of the thruster's failure and environmental disturbance via descriptive analytics, predictive analytics, and prescriptive analytics such as statistical analysis, Machine Learning (ML), and sensitivity analysis (SA). The experiment results show the feasibility of the proposed method.

Structural health monitoring aims to identify the modal parameters of offshore

structures imposed by drifting ice. The covariance-driven stochastic subspace algorithm (SSI-cov) was proposed to identify physical modes (natural frequency, damping ratio, and mode shape). Many uncertain parameters of SSI-cov bring uncertainties to the identified modal parameters. To address this issue, diagnostic analytics, such as clustering, and prescriptive analytics such as uncertainty analysis (UA) and SA, are combined to quantify the uncertainty of the identified modal parameters. The results present the proposed method can achieve an efficient and accurate uncertainty quantification of the identified modal parameters. Additionally, it outperforms the traditional slack values-based SSI-cov.

AIS data contains rich information about ship status, which has been widely used for ship behaviour analysis. To take full advantage of information-rich AIS data, this study gives three applications: COVID-19 impact analysis, probabilistic ship route prediction, and short-term ship trajectory prediction. First, descriptive analytics and diagnostic analytics are used to extract important features and analyze the statistics of these features in the case of the Coronavirus disease 2019 (COVID-19) raging. This study mainly analyzes the interaction between ship behaviours and COVID-19 impacts for the support of marine traffic management. Based on this work, the extracted features are further applied to the next application such as ship route prediction. Diagnostic analytics, including clustering and dynamic time warping (DTW), is chosen to make probabilistic ship route prediction. It is carried out in two steps. The first step is to cluster ship trajectories using clustering to render routes and the next step is to classify ship trajectories into different routes based on trajectory similarity estimated by DTW. Finally, the obtained route information is then used as prior knowledge for the ship trajectory prediction. A hybrid model is constructed based on historical trajectory information and online predicted ship positions obtained by Gaussian Process. The results demonstrate the proposed model outperforms the data-driven model and can obtain more accurate ship trajectory prediction.

Acknowledgment

Contents

List of Publications

This thesis is based on research resulting in four journal papers and two conference papers. They are all enclosed in the appendix section. In the following list of publications, the papers are listed chronologically, but in the main body of the text, a more thematic presentation order is prioritized over the chronological one.

    I C. Wang, G. Li, R. Skulstad, X. Cheng, O. Osen and H. Zhang, "A sensitivity quantification approach to significance analysis of thrusters in dynamic positioning operations", *Ocean Engineering*, vol. 223, p. 108659, 2021.

  II C. Wang, G. Li, P. Han, O. Osen and H. Zhang, "Impacts of COVID-19 on Ship Behaviours in Port Area: An AIS Data-Based Pattern Recognition Approach", *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, pp. 25127-25138, Feb 2022.

III C. Wang, T. Nord, and G. Li, "Automated Modal Parameters Identification During Ice-Structure Interactions", *American Society of Mechanical Engineers*, vol. 85864, issue 19, p. V002T02A020, 2022.

IV C. Wang, T. Nord, G. Ziemer and G. Li, "Towards uncertainty and sensitivity analysis for modal parameters identification during ice–structure interaction", *Ocean Engineering*, vol. 277, p. 114224, 2023.

  V C. Wang, M. Zhu, O. Osen, H. Zhang and G. Li, "AIS data-based probabilistic ship route prediction", *2023 IEEE 6th Information Technology, Networking*, pp.167-172, Feb 2023.

VI C. Wang, P. Han, M. Zhu, O. Osen and H. Zhang and G. Li, "AIS data-based Hybrid predictor for short-term ship trajectory prediction considering uncertainties", *Reliability Engineering and System Safety*, submitted 2023.

The following papers will not be discussed in this thesis. They may, however, be considered relevant due to co-authorship and similar topics:

i C. Wang, T. Wang, T. Nord, H. Zhang, O, Ottar and G. Li, "Uncertainty and Sensitivity Analysis for the Performance Evaluation of Ship Dynamical Model", *32nd International Ocean and Polar Engineering Conference*, Jun 2022.

ii S. Gao, C. Wang, S A. Tuenea, G. Li, and H. Zhang " In situ measurements and simulations of a net cage in currents", in *Aquaculture Engineering*, submitted, 2022.

iii M. Zhu, C. Wang, P. Han, R. Skulstad, G. Li, and H. Zhang, "An AIS data Data-driven based Hybrid Approach to Ship trajectory prediction- A Real in Oslofjord", in *IEEE Sensors Journal*, submitted, 2023.

ML	Machine Learning
DL	Deep Learning
SA	Sensitivity Analysis
UA	Uncertainty Analysis
UQ	Uncertainty Quantification
AIS	Automation Identification System
MMSI	Maritime Mobile Service Identity
DWT	Deadweight
DP	Dynamic Positioning
SVM	Support Vector Machine
SHM	Structural Health Monitoring
LSTM	Long Short-term memory
GP	Gaussian Process
SSI-cov	Covariance-driven Stochastic Subspace Identification
DTW	Dynamic Time Warping
DBSCAN	Density-based Spatial Clustering of Applications with Noise
MOGP	Multi-output Gaussian Process
NLML	Negative Log Marginal Likelihood
FAST	Fourier Amplitude Sensitivity Test
EFAST	Extended Fourier Amplitude Sensitivity Test
CDF	Cumulative Distribution Function
PDF	Probability Distribution Function
KS	Kolmogorov–Smirnov
HSVA	Hamburg Ship Model Basin

MDOF	Multi Degree Freedom
IC	Intermittent Crushing
FLI	Frequency Lock-in
CC	Continuous Crushing
SOG	Speed Over Ground
COG	Course Over Ground
Lat	Latitude
Lon	Longitude
COVID-19	Coronavirus Disease 2019
RELM	Regularized Extreme Learning Machine
BP	Back Propagation
RMSE	Root Mean Square Error
MAC	Modal Assurance Criteria
OMA	Operational Modal Analysis
RD	Robust Distance
MCD	Minimum Covariance Determinant
LHS	Latin Hypercube Sampling
NB	Number of Blocks
BR	Block Rows
SF	Sampling Frequency
SO	System Order
SC-I	Deviance in Frequency
SC-II	Deviance in Damping Ratio
SC-III	Normalized Standard Deviation of The Frequency
Std	Standard Deviation
Upper	Upper Bound of 95% Confidence Interval
Lower	Lower Bound of 95% Confidence Interval

Nomenclature

ϵ, C Two hyperparameters in the SVR

$k(\cdot, \cdot)$ Kernel function

$F_y(y)$ Unconditional CDF

$F_{y|\tilde{x}}(\cdot)$ Conditional CDF

γ Cumulative distance matrix

O Observability matrix

H Subspace matrix

f Frequency

ξ Damping coefficient

φ Mode shape

ω Frequency in radian

σ Normalized standard deviation

S_f Variance of frequency

S_ξ Variance of damping

S_{MAC} Variance of MAC

$\phi(x)$ Eigenvector by which x is mapped into high dimensional space

$\mathcal{N}(0, \sigma^2)$ The Gaussian distribution with the variance of σ^2

Σ_* Covariance matrix

$-log\,p(\boldsymbol{y}|X, \boldsymbol{\theta})$ The negative log marginal likelihood

$KL(\cdot)$ The Kullback-Leibler divergence

$\mathcal{O}$ The time complexity

$V(\cdot)$ Variance

$V(\sim i)$ The total variance of parameters except the ith parameter

S_i The main sensitivity index of factor i

S_{Ti}	The total sensitivity index of factor i
μ	Continuous time eigenvalue
λ	Eigenvalue
ϵ_d	Gaussian noise of the dth output
Σ_s	Diagonal matrix with elements σ_d^2.
K_{**}	The kernal matrix with elements k
R	Maximum circumference of the Earth
P	Probability
$\Re$	Feature space

List of Figures

List of Tables

Introduction

Recent years have seen an ever-increasing interest in the exploration, development, and utilization of ocean resources, which has made marine operations more frequent and intensive. To ensure safe marine operations, various advanced sensors are deployed on vessels, offshore platforms, etc., to collect real-time data. As time goes on, this data is piled up and forms marine big data characterized by five high Vs (volume, Value, Velocity, Variety, and Veracity) [1]. This data can provide stakeholders with valuable and well-informed insights into different aspects of decision making, health monitoring, traffic monitoring, etc. This dissertation mainly focuses on how to apply data mining technologies to discover potential patterns and knowledge, hence enabling onboard support of marine operations carried out in a safer and more efficient manner.

1.1 Background and motivation

Marine operations are subject to uncertain factors from the environment and human decisions. According to the annual overview 2022 of European Maritime Safety Agency[1], human elements account for 81.1% of maritime accidents based on the analysis of both human action events and human behaviours jointly. Therefore, offering onboard support for operators' decision-making is crucial to safe and sustainable marine operations. With the availability of data acquisition technologies, a huge volume of marine data has been collected over decades, forming marine big data. Marine big data speeds up the transition towards digitalization and automation to increase competitiveness and enhance operational efficiency [2, 3]. Digitalization involves building a digital model through data analysis and modelling, which is used for monitoring, diagnostics and prognostics, and maintenance of physical assets, systems, and manufacturing processes [4]. In this context, data analysis and modelling present a promising prospect for onboard support of marine operations, such as ship motion prediction [5, 6], path planning [7, 8], fault detection [9, 10], sea state estimation [11, 12].

Marine big data has the significant characteristics of high dimensionality, large volume, and noise pollution, which brings exceptional challenges to data mining [13]. To better manage marine big data and uncover its great values, the marine data value chain is abstracted and illustrated in Figure 1.1. It includes four parts: data acquisition and transition, data management, data analysis, and onboard support system development. Data acquisition and transition are carried out by means of satellites, aerial remote sensors, stations, ships, buoys, and sensors serving marine-related fields. A large volume of in-situ data is collected in real time, which contains abundant temporal and spatial

---

[1]Annual Overview of Marine Casualties and Incidents, https://www.emsa.europa.eu/newsroom/latest-news/item/4867-annual-overview-of-marine-casualties-and-incidents-2021.html, published on 30.11.2022
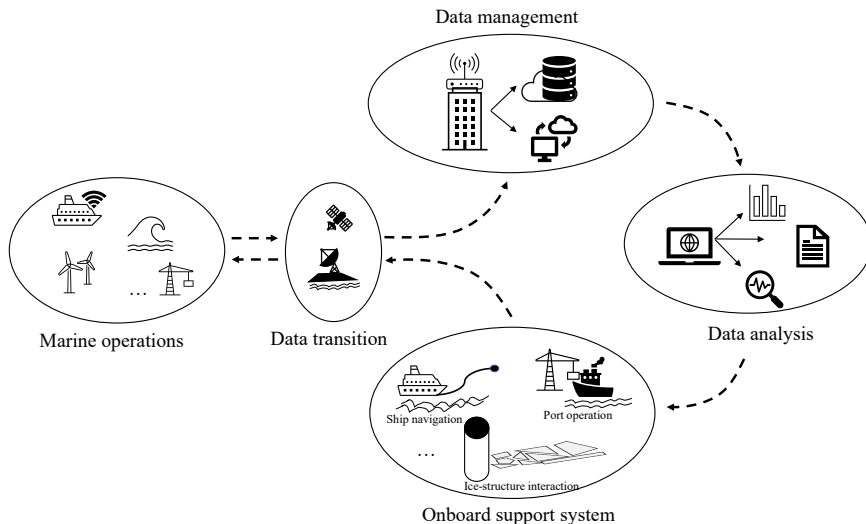
Figure 1.1: Marine data value chain.

information and enormous potential value [14]. However, it is intractable to directly observe informative and valuable insights from marine data due to complex data structure and types, high dimensionality, high volume, etc. Therefore, it is important to perform data processing in the phase of data management, such as data cleaning, data transformation, and integration. The processed data is then stored in the local database or cloud database. In order to serve marine operations, data analysis is used to extract valuable knowledge from complex and vast data via statistical analysis, Machine Learning (ML), Deep Learning (DL), etc. The extracted knowledge can be used to develop digital models and decision support systems for a wide range of marine applications such as ship collision avoidance, structural health monitoring, and behaviour prediction.

As shown in Figure 1.1, data analysis plays a key role in bridging the gap between marine data and onboard support systems. Data analysis tends to fall into four distinct categories: descriptive analytics, diagnostic analytics, predictive analytics, and prescriptive analytics [15]. Descriptive analytics answers the question: 'What happened' by the description of statistics in marine data. Diagnostic analytics addresses the problem of 'why it happened' through correlation analysis and regression analysis. They focus mainly on the summarization and interpretation of past facts. Consequently, they can provide limited support for marine operations since marine data features high volume, high dimensionality, and complex relationships.

To address this issue, predictive analytics is used to build some advanced predictive models using various modelling technologies. In recent years ML and DL are introduced in marine data mining due to their excellent modelling capability. ML and DL can build a data-driven model automatically using instance data or past experience to predict the future behaviour of a system without explicit knowledge of the physical behaviour of this system [16]. Therefore, predictive analytics has been widely used for the support

of marine operations such as trajectory prediction [17], ocean wave estimation [18], and power consumption estimation [19].

Despite the popularity of data-driven models, many new problems have also followed. One of these challenges is the black-box nature of data-driven models, which makes it difficult to interpret the results [20]. In addition, the outputs of data-driven models are subject to imprecision because 1) models are imperfect abstractions of real systems and full of various assumptions; 2) input data is biased in ways of noises, incomplete, obtained from inaccurate sampling; 3) models are used for situations very different from those in the historical knowledge base [21]. To address these issues in predictive analytics of marine data, prescriptive analytics is applied to explain uncertainty sources and quantify model uncertainty since it is good at answering the question of how to make it happen. The well-known approaches are sensitivity analysis (SA) and uncertainty analysis (UA). SA is to investigate how the variation of the output is apportioned to each input. UA is instead to quantify how uncertain is the output. The combination of SA and UA can assist in model optimization, model reliability improvement, risk assessment, etc., furthermore, guiding decision-makers toward a specific action to take for safe marine operations.

1.2   Research questions

The focus of this dissertation is concerned with data analysis and modelling for decision support of marine operations. This prompts the first question of this dissertation:

A significant portion of maritime accidents occurs due to human errors resulting from inadequate situational awareness and understanding of risk factors. Such accidents could be preventable if an onboard support tool provided decision-makers with a better understanding of potential risks in marine operations. The availability and accessibility of marine big data open up the opportunity to drive the development of onboard support tools. However, it is intractable to obtain knowledge directly from marine data as it is full of various noises, faults, outliers, and complex relationships among data features [22]. Therefore, data analysis is proposed to discover the valuable knowledge and critical insights hidden in the marine data, in order to support stakeholders to make better-informed decisions and take timely actions to ensure safe and efficient operations.

- **Why is modelling indispensable for marine data mining and how to perform it?**

Traditional data analytics, such as descriptive analysis and diagnostic analysis, are limited to simple tasks such as statistical analysis of a small amount of data and correlation analysis. It will lose effectiveness to analyse marine big data characterized by five V's [1]. To address this problem, predictive analytics, such as ML and DL, are introduced to construct advanced models driven by marine data to predict what is likely to happen. They can perform faster, more accurately, and more precisely for massive volumes of marine data [23]. Despite their advantages, data-driven models are susceptible to uncertainties from environmental factors, out-of-distribution data, etc. As a consequence, models' performance can not be guaranteed in real applications. This leads to the next research question.

- **How does uncertainty analysis benefit the predictive analytics of marine**

data?

In order to answer this question, it is necessary to comprehend what types of uncertainties arise during data analysis and modelling. Uncertainty can be broadly classified into two categories: aleatory uncertainty and epistemic uncertainty. Aleatory uncertainty arises due to the inherent randomness or noise in data, while epistemic uncertainty is due to the lack of knowledge or understanding of the model's behaviour in certain regions of the sample space. [24]. Aleatory uncertainty is intrinsic and not reducible but can be estimated by analyzing the noise in the data [25]. The more random the data, the larger the aleatory uncertainty. Epistemic uncertainty, on the other hand, can be neutralized by collecting more data or knowledge. However, it cannot be completely eliminated in marine data analysis. For marine data analysis, it is difficult to remove epistemic uncertainty completely. Therefore, it is necessary to conduct uncertainty quantification (UQ) of total uncertainty (aleatory uncertainty and epistemic uncertainty), to ensure the reliability and trustworthiness of data analysis results.

- **What are the limitations of UA in marine data analysis?**

UA is carried out via Monte Carlo sampling from the data distribution. However, the real data distribution exhibits coupling or dependence between variables and is hard to be estimated. To mitigate this problem, data distribution is often assumed to be independent and identically distributed. Such simplification might introduce additional uncertainty for UQ. Another challenge lies in the high dimension of marine data brings more computational cost. Of these data features' variation, some do not change the value of variables of interest. Removing these uninfluential factors can speed up UA to provide onboard support for marine operations. This leads to the next research question.

- **What are the advantages and limitations of SA for decision support of marine operations?**

SA, in the most general sense, is to investigate how the outputs of a system or model are affected by its inputs [26]. SA is particularly useful in identifying which factors may be redundant and can be fixed or removed in subsequent analysis [27]. In recent years, research and practice in SA have gained significant momentum as many researchers contribute to designing various theoretical frameworks based on different applications in their domains. It has been widely used for dimensionality reduction in UA, model-based policy-making, decision support, etc. In the context of marine operations, SA can also provide decision-makers with a deep understanding of the processes, hypotheses, parameters, scales, and their interactions that affect their operations to ensure safe and successful outcomes.

However, one limitation of SA is its reliance on a function or model which is often not available in marine data mining due to the complexity and high dimensionality of marine data. It is extremely hard to use a mathematical model to represent the non-linear relationship among observations. Thus, researchers often rely on constructing a surrogate model to approximate complex patterns and relationships in marine data. The use of surrogate models can help improve the accuracy and efficiency of SA.

1.3 Scope of work

1.3.1 Research objectives

In seeking to answer the above research questions, this dissertation seeks to obtain the following research objectives (ROs):

√ **RO1: Apply descriptive and diagnostic analytics to analyze marine data for decision support of marine operations.**

Descriptive and diagnostic analytics are two essential categories of marine data mining that provide decision-makers with valuable insights into what happened in the past and why it happened. Descriptive analytics allows stakeholders to examine marine data statistically and understand the patterns and trends. On the other hand, diagnostic analytics goes a step further and identifies the root causes of certain events or behaviors. For example, data generated from Automatic Identification System (AIS) contains abundant ship behaviour information concerning Maritime Mobile Service Identity (MMSI), types, deadweight (DWT), the time of arrival and departure, as well as other ship's states. Through AIS data analysis, stakeholders can have a better understanding of the interactions among ship's behaviour, regional economy, port operations, and other factors. Thus, they can take timely actions to fit future development and challenges. The descriptive and diagnostics of AIS data are described in paper II.

√ **RO2: Create surrogate models with ML and DL to find the map among observations.**

Marine data is characterized by high dimensionality and nonlinearity. As a result, input-output patterns are complicated and are hard to be represented with simple functions or mathematical models. For example, the unknown relationship between thrusters and ship DP capability poses challenges to thrusters' importance analysis. Hence, it is necessary to construct a surrogate model between thrusters' thrust and ship DP capability, which can be used further for ship DP capability improvement. ML and DL have the powerful ability to fit nonlinearity among variables in marine data. The corresponding exploration is shown in Papers I and VI.

√ **RO3: Develop an SA-based method to identify the most influential factors to expected outcomes in marine operations.**

Marine operations are influenced by a variety of factors. Some might lead to destructive and detrimental impacts on the environment, personnel, and physical assets. For example, offshore vessel operations are affected by the interaction of the environment and the vessel's dynamic positioning (DP) system performance. Specifically, when an emergency occurs during DP operations, the requirement of rapid response poses challenges to decision-makers. Understanding the sensitivity of the vessel's DP operations to different options or factors could assist operators in making a well-informed decision so as to avoid risks. The corresponding exploration is discussed in Paper I.

√ **RO4: Quantify the uncertainty of observations of interest by a probabilistic model in marine operations.**

Decision-making heavily relies on the confidence level of the model outputs. Due to the complexity and uncertainty of marine operations, the precision of models is prone to the impacts of uncertain factors such as data errors, and out-of-distribution data. Inaccurate predictions may result in wrong judgments of the current conditions, putting the safety of personnel and assets at risk. For example, when AIS data-based model is used to make ship trajectory predictions, uncertainties arising from the pilot's decision, environmental impacts, and data quality can affect the accuracy of the developed probabilistic model with DL. How to quantify model uncertainties to provide trustworthy predictions is the main concern of papers VI and VII.

$\sqrt{}$ **RO5: Propose a UA and SA framework to improve the efficiency of uncertainty quantification for online support of marine operations.**

As mentioned previously, UA incurs huge computational costs in evaluating the re-sampled data variables or model parameters, making it unsuitable for real-time support in marine operations. However, some uncertain inputs do not have an impact on the model outputs and are therefore redundant, and can be fixed or removed for UA. For instance, in the case of the covariance-driven stochastic subspace identification (SSI-cov) algorithm used to identify modal parameters of structures, some parameters bring large uncertainty to the result while the remaining parameters can be fixed to implement more efficient UA. This issue will be discussed and addressed in papers IV and V.

The interconnection between research objectives and research questions is clarified as follows. RO1 is to apply descriptive and diagnostic analytics for marine data mining, which answers the first research question. RO2 focuses on data modelling, which aims to address the second and fifth research questions. RO3 investigates the application of SA for decision support, which is related to the fifth research question. RO4 corresponds to the third research question for UQ. RO5 proposes an analysis framework based on UA and SA for UQ, which is related to the fourth research question.

1.3.2 Interconnection between the research objectives and publications

The interconnection between the research objectives and the published papers is shown in Figure 1.2. In order to satisfy RO1, two case studies are given, one of which uses AIS data to analyse ship behaviours and the other takes structure response data to conduct modal parameters identification. Paper I regards the problem of how the coronavirus disease 2019 (COVID-19) causes an impact on ship behaviours through the descriptive analytics and diagnostic analytics of AIS data and uses clustering technology combined with statistical approaches to discover abnormal ship behaviours. For structural response data, clustering technology is proposed to identify modal parameters automatically in paper III.

For RO2, modelling is carried out with ML or DL. For thrusters' significance analysis, A support vector machine (SVM) is chosen to construct a surrogate model between thrusters and ship DP capability in the paper I. For ship trajectory prediction, a Gaussian Process model (GP) is used to build the model for ship position prediction in paper VI.

A SA method is combined with ML to analyze thrusters' importance under the cases of different thrusters' failures and sea states during ship DP operations in the paper I,

which is related to RO3. The same SA approach is also used to pick up important parameters to accommodate the issues of UA as suggested by RO5. In paper IV, a UA and SA framework is proposed to quantify the uncertainty of identified modal parameters to support online structural health monitoring (SHM).

For UA, probabilistic ship trajectory prediction is given to present uncertainties of model predictions in marine operations. This case contains two parts: one is to estimate the probability of which route a ship will track in paper V and another applies a GP model to predict the ship's uncertain movement in the short term in paper VI. These two papers are related to RO4. All of the above forms part of the module that provides decision support for marine operations.
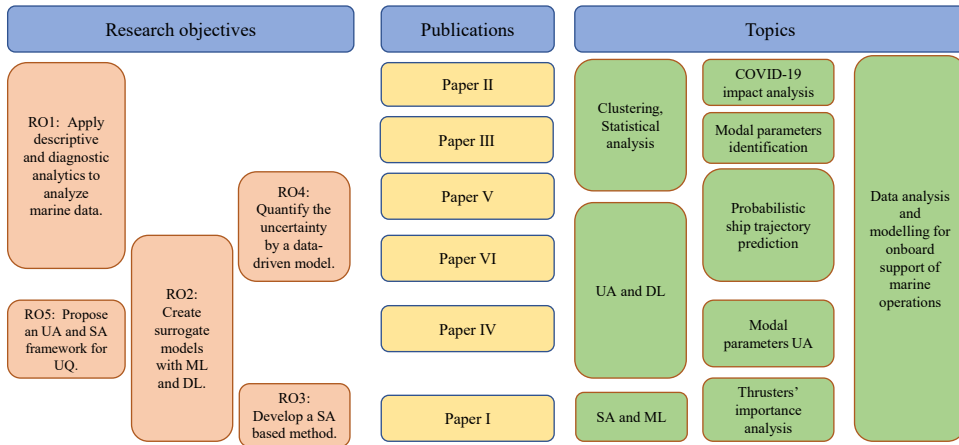
Figure 1.2: Interconnection of published paper in the thesis.

1.4 Contributions of the dissertation

The major contributions of this dissertation are as follows, which are related to the research objective above:

- Descriptive analytics and diagnostic analytics are utilized in marine data analysis to uncover potential patterns and support decision-making. Two case studies demonstrate the effectiveness of these methods, which are closely related to RO1.

- ML technologies are employed to construct a surrogate model from marine data, which enables SA to analyze the importance of each factor to a specific operation. It is connected to RO2 and RO3.

- A hybrid model is proposed to enhance the confidence level of predicted results through UQ for online decision support. It is corresponding to RO4.

- A framework based on UA and SA-based framework is proposed to implement rapid UQ of observations of interest for onboard support of SHM. It pertains to RO5.

1.5 Structure of the dissertation

This introductory chapter presented the background for the dissertation research, establishing its main goals and defining the scope of work. The rest of this dissertation unfolds as follows. Chapter 2 introduces the foundation of data analysis methods to support marine operations and the experimental platforms that are used to develop and test the model. Chapter 3 presents the first case study, which focuses on thrusters' importance analysis when the vessel is in dynamical positioning operation. This chapter is based on the paper I. Chapter 4 relates to papers III and IV, and uses clustering technology for automated modal parameters identification, and proposes a UA and SA framework for UQ of identified modal parameters. Chapter 5 presents AIS data mining and its application for probabilistic ship trajectory prediction. This chapter is based on papers II, V, and VI. Chapter 6 concludes the dissertation, summarizes the contributions, and indicates the directions for future works.

Data analysis and modelling methods to support marine operations

This chapter introduces data analysis as a fundamental technology to support marine operations online. Fig. 2.1 presents a schematic illustration of applying data analysis for decision support of marine operations. As shown in the blue picture [1], a variety of marine operations are carried out at sea to generate different marine data such as structural response data from offshore structure, sea states, ship status data from dynamic positioning vessel, AIS data from the sailing vessels. These data are, in return, used for onboard support of marine operations by data analysis. Data analysis involves a combination of various technologies that are grouped into different categories [28].
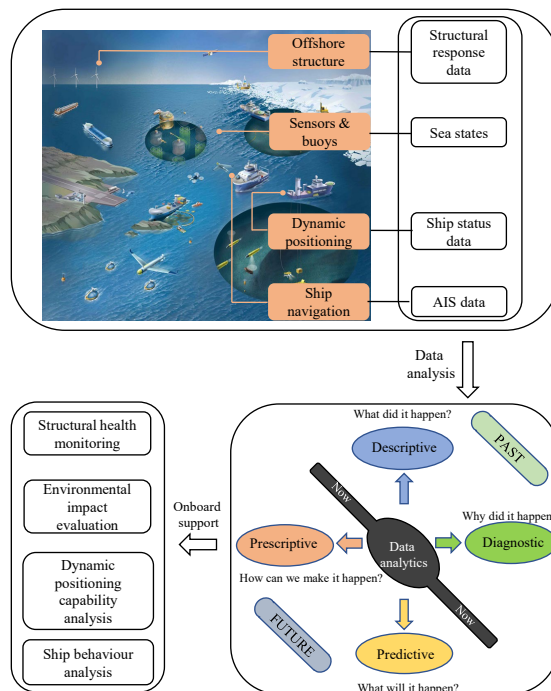
Figure 2.1: Schematic illustration of data analysis to support marine operations.

---

[1]Centre for Autonomous Marine Operations and Systems, https://www.ntnu.edu/amos/research

Section 2.1 introduces the related marine applications to data analysis. Section 2.2.1 presents the data analysis methods that are used in this dissertation. Section 2.3 explains the experimental platforms and data sources that are used to build and validate the models.

2.1 Applications of data analysis for marine operations

As shown in Fig. 2.1, data analysis involves four categories: descriptive analytics, diagnostic analytics, predictive analytics, and prescriptive analytics. Descriptive analytics focuses on summarizing and presenting historical marine data to gain insights into past events and trends. It uses statistical analysis to organize data into meaningful and understandable formats such as tables, charts, and graphs.

Diagnostic analytics often involves a deep dive into data, including examining trends, patterns, and relationships between different variables using regression analysis or clustering technologies. Predictive analytics aims to create advanced models with ML and DL to predict future behaviours or events. Prescriptive analytics takes a step further by recommending the best course of action to achieve the desired outcome. This analysis can provide insights and knowledge into the factors that affect the outcome via SA, which can help identify potential solutions to drive decision-making. These four types of data analysis are widely used in, but not limited to, the following applications:

- **Ship behaviour analysis:** With the acquisition of large amounts of AIS data, ship behaviour analysis has elicited considerable attention in the literature. Clustering was employed to find normal patterns and ship behaviour is monitored using real-time AIS [29]. Sheng et al. created a logic regression model to identify illegal, unreported, and unregulated ships [30]. Zhang et al. estimated the marine traffic volumes by analyzing the historical trajectories of ships in Singapore port [31].

- **Navigation safety:** Better situation awareness of ship surroundings can help operators manipulate the ship in a safe manner. Therefore, Tang et al. train a Long short-term memory (LSTM) model to predict ship trajectory 10 minutes ahead [32]. Considering uncertain ship movement, Rong et al. proposed a Gaussian process model to make a probabilistic ship trajectory prediction [33].

- **Structure health monitoring:** Understanding the dynamic interaction between ice and structure is critical for the operational safety of offshore structures. Hence, the SSI-cov algorithm was proposed to estimate the frequencies, damping ratios, and their uncertainties [34]. Nord et al. applied SSI-cov to identify modal parameters on the Norströmsgrund lighthouse under a variety of ice-structure interaction modes [35].

- **Ship DP capability evaluation:** The evaluation of DP capability can reduce the risk of ship offshore operation. Wang et al. proposed a DP capability polar plot program to determine the maximum environmental loads of DP vessels [36]. Mauro et al. assessed the operability of a ship DP system by Quasi-Monte Carlo sampling from the joint distribution of wind and waves [37].

Generally, the above applications require a combination of different analysis approaches, including descriptive analytics, diagnostic analytics, predictive analytics, and prescrip-

tive analytics. The following section will introduce the core analysis methods used in this dissertation.

2.2 Data analysis methods

Diagnostic analytics mainly uses two clustering technologies (hierarchical clustering and density-based spatial clustering of applications with noise (DBSCAN)), dynamic time warping (DTW), and SSI-cov algorithm to analyze marine data in different applications. Predictive analytics is to construct data-driven data using a support vector machine and multi-task Gaussian process. Prescriptive analytics mainly analyzes the uncertainty of model outputs and the sensitivity of input parameters by means of UA and SA.

2.2.1 Clustering analysis

Hierarchical clustering

Hierarchical clustering is represented by a rooted tree where each leaf represents a data point and each internal node represents a cluster containing its descendant leaves. The tree is constructed based on the distance information between different data points [38]. This tree is constructed by iteratively merging the two closest data points or clusters until all the data points have been merged into a single cluster. Since it is not sensitive to initialization conditions and less sensitive to outliers, it has a wide range of applications in data mining.

Let $Q = q_1, q_2, ..., q_n$ be a set of objects. The clustering process is divided by the following steps [39]:

- Calculate the proximity matrix: calculate the distance between each pair of objects (q_i, q_j).

- Initialize the clusters: assign each data point to its own cluster.

- Merge clusters: merge the two similar clusters iteratively until all the data points have been merged into a single cluster. At each step, the proximity matrix is updated correspondingly.

- Create a dendrogram: create a dendrogram to show the hierarchy of clusters and to help identify the appropriate number of clusters to use.

- Determine the number of clusters: use the cut-off value to partition the hierarchical tree into clusters.

Hierarchical clustering is relatively slow as a result of pairwise distance computation. As a consequence, it is not suitable for large dataset analysis. In addition, it needs a user-specified number of clusters and a cut-off value. That could affect the accuracy of clustering results.

Density-based spatial clustering of applications with noise

The DBSCAN identifies distinctive clusters in the data based on their density [40, 41]. It divides the dataset into three categories: core points, boundary points, and noise points. Core points are those that have a sufficient number of other points within a certain radius, boundary points are those that do not satisfy the core point condition

but have a core point in their neighborhood, and noise points are those that do not have any points within their neighborhood. The algorithm starts with core points and expands clusters until all core points have been visited, resulting in a series of clusters.

These three points are distinguished by two parameters: minPts and eps. 'minPts' refers to the minimum number of points clustered together. if a point is considered as a core point if it has at lear minPts points in its surrounding area. Otherwise, it is taken as boundary points or noise points. 'eps' represents the maximum distance between two points for them to be considered neighbors. Points that are within the eps distance of a core point are assigned to the same cluster. The algorithm steps are shown as follows:

- Choose an arbitrary point from the dataset that has not been visited.

- Determine if the point is a core point. A core point has at least minPts points (including itself) within its eps radius.

- If the point is a core point, create a new cluster and add the point and all its neighbors to the cluster. The neighbors are found by recursively repeating the same process starting from each neighbor point.

- If the point is not a core point, mark it as a noise point.

- Repeat the above steps for all unvisited points in the dataset until all points have been visited.

DBSCAN has several advantages over traditional clustering algorithms, including the ability to identify clusters of arbitrary shape, the ability to automatically detect outliers without requiring prior knowledge of their number or location, and the ability to adaptively determine the number of clusters. The pseudocode of DBSCAN is shown in Algorithm 2.

2.2.2 Dynamic time warping

The objective of DTW is to compare time-dependent sequences $X = (x_1, x_2, ..., x_n)$ and $Y = (y_1, y_2, ..., y_m)$. Warping path information is contained in a $n \times m$ distance matrix:

$$D = \begin{pmatrix} d(x_1, y_1) & d(x_1, y_2) & \cdots & d(x_1, y_m) \\ d(x_2, y_1) & d(x_2, y_2) & \cdots & d(x_2, y_m) \\ \vdots & \vdots & \cdots & \vdots \\ d(x_n, y_1) & d(x_n, y_2) & \cdots & d(x_n, y_m) \end{pmatrix} \tag{2.1}$$

where d is a distance measure function. DTW aims to find out the warping path $W = \{w_1, w_2, ..., w_k, ...w_K\}$ by minimizing the following function:

$$\begin{cases} DTW(X, Y) = min(\sqrt{\sum_{k=1}^{K} d(w_k)}) \\ d(w_k) = d(i, j) = d(x_i, y_j) \end{cases} \tag{2.2}$$

Given $w_k = (i, j)$ and $w_{k-1} = (i', j')$. The optimal warping path is subject to the following three conditions: (i) Boundary condition: $w_1 = (1, 1)$ and $w_K = (n, m)$; (ii)

Algorithm 1: DBSCAN algorithm

    Input: D: Data, ϵ, minPts, Euclidean_distance
    Input: label

1 **for** *point p in D* **do**
2   **if** *label(p) is undefined* **then**
3   neighbours N $\leftarrow$ RangeQuery(D,Euclidean_distance,p,ϵ)
4   **end**
5   **if** $|N| < minPts$ **then**
6   label(p)=Noise
7   **end**
8   l $\leftarrow$ next cluster label
9   label(p) $\leftarrow$ l
10   Seed set S $\leftarrow$ N $\cup$ {p}
11   **for** *q in S* **do**
12   **if** *label(q) is Noise* **then**
13   label(q) $\leftarrow$ l
14   **end**
15   **end**
16   **if** *label(q) is undefined* **then**
17   neighbours N $\leftarrow$ RangeQuery(D,Euclidean_distance,q,ϵ)
18   label(q) $\leftarrow$ l
19   **end**
20   **if** $|N| \geq minPts$ **then**
21   S $\leftarrow$ S $\cup$ N
22   **end**
23 **end**
24 **Function** `RangeQuery`(*D, dist,Q, ϵ*)**:**
25   neighbours = empty list **if** *point p in D* **then**
26   **if** *dist(Q,p)$\leq \epsilon$* **then**
27   neighbours = neighbours $\cup$ {p}
28   **end**
29   **end**
30   **return** neighbours

Monotonicity condition: $i - i' \leq 1$ and $j - j' \leq 1$; (iii) Step size condition: $w_k - w_{k-1} \in \{(1,0),(0,1),(1,1)\}$ for $k \in [1, K-1]$. The smaller $DTW(X,Y)$ represents the larger similarity of X and Y.

In order to solve (2.2), dynamic programming is introduced with a computational complexity of O(nm). By this method, a cumulative distance matrix γ of the same dimension as the D, is created using (2.3). A more detailed introduction can be referred to [42].

$$\gamma(i,j) = min\{\gamma(i-1,j-1), \gamma(i-1,j), \\ \gamma(i,j-1)\} + d(x_i, y_j) \tag{2.3}$$

2.2.3 Covariance-driven stochastic subspace identification algorithm

To obtain the structural properties, an SSI-cov algorithm was proposed to estimate the frequencies, damping ratios, and uncertainties. The linear time-invariant system is described by a discrete-time stat-space model

$$\begin{cases} x_{k+1} = Ax_k + w_k \\ y_k = Cx_k + v_k \end{cases} \tag{2.4}$$

where w_k and v_k are the process and output noise, respectively. In order to identify matrices A and C from which the modal frequencies, damping, and mode shapes can be obtained, the eigenvalues and eigenvectors of the system in Eqn.2.4 are calculated by the following equations

$$\begin{cases} (A - \lambda_i I)\phi_i = 0 \\ \varphi_i = C\phi_i \end{cases} \tag{2.5}$$

from which the μ_i, f_i, and ξ_i can be obtained:

$$\mu_i = \frac{ln\lambda_i}{T}, f_i = \frac{|\mu_i|}{2\pi}, \xi_i = -100\frac{\Re(\mu_i)}{|\mu_i|} \tag{2.6}$$

where T is the sampling period. The stochastic subspace algorithm is a prevalent method to estimate the matrices A and C. The algorithm uses the output data to build a subspace matrix $H_{p+1,q} \in \mathbb{R}^{(p+1)r \times qr_0}$. Therein, r is the number of sensors, r_0 is the number of reference sensors, and p and q are the parameters chosen such that $pr \geq qr_0 \geq n$, where n is the model order. The subspace matrix $H_{p+1,q}$ can be truncated at a user-defined model order n via singular value decomposition

$$H_{p+1,q} = \begin{bmatrix} U_1 & U_0 \end{bmatrix} \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_0 \end{bmatrix} \begin{bmatrix} V_1^T \\ V_0^T \end{bmatrix} \tag{2.7}$$

and

$$O_{p+1} = U_1\Sigma_1^{1/2} \tag{2.8}$$

The C matrix can be directly extracted from the first block of r rows of the observability matrix O_{p+1}, while the A matrix can be obtained from a least-squares solution of

$$O_{p+1}^{\uparrow}A = O_{p+1}^{\downarrow} \tag{2.9}$$

14

where $O^{\uparrow}_{p+1} = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{p-1} \end{bmatrix}, O^{\downarrow}_{p+1} = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{p} \end{bmatrix}$

The principle of SSI-cov is to propagate the covariance of the subspace matrix, Σ_H, to the modal parameters through first-order perturbations. The covariance of the modal parameters is obtained as

$$
\begin{aligned}
cov\left(\begin{bmatrix} f_i \\ \xi_i \end{bmatrix}, \begin{bmatrix} f_j \\ \xi_j \end{bmatrix}\right) &= \begin{bmatrix} J_{f_i,A} & 0_{1,rn} \\ J_{\xi_i,A} & 0_{1,rn} \end{bmatrix} \Sigma_{AC} \begin{bmatrix} J_{f_i,A} & 0_{1,rn} \\ J_{\xi_i,A} & 0_{1,rn} \end{bmatrix}^T \\
cov\left(\begin{bmatrix} \Re(\phi_i) \\ \Im(\phi_i) \end{bmatrix}, \begin{bmatrix} \Re(\phi_j) \\ \Im(\phi_j) \end{bmatrix}\right) &= \begin{bmatrix} \Re(J_{\phi_i,A,C}) \\ \Im(J_{\phi_i,A,C}) \end{bmatrix} \Sigma_{AC} \begin{bmatrix} \Re(J_{\phi_i,A,C}) \\ \Im(J_{\phi_i,A,C}) \end{bmatrix}^T
\end{aligned}
\tag{2.10}
$$

2.2.4 Support vector machine

For m training samples (x_i, y_i) where $x_i \in \Re^k$ and $y_i \in \Re$ The basic form of SVM is as follows:

$$
\begin{aligned}
&\min_{\omega,b} \frac{1}{2}||\omega||^2 \\
&s.t.\ y_i(\omega^T \phi(x_i) + b) \geq 1,\ i = 1, 2, \ldots, m
\end{aligned}
\tag{2.11}
$$

Note that $\phi(x)$ is the eigenvector by which x is mapped into high dimensional space, $\phi(x)^T$ is the transpose of $\phi(x)$, ω is the normal vector of a plane, b is an offset. Lagrange multiplier method is used for transforming (2.11) to its dual problem

$$
\begin{aligned}
&\max_{\alpha} \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_i \alpha_j y_i y_j \phi(x_i)^T \phi(x_j) \\
&s.t. \begin{cases} \sum_{i=1}^{m} \alpha_i y_i = 0 \\ \alpha_i \geq 0,\ i = 1, 2, \ldots, m \end{cases}
\end{aligned}
\tag{2.12}
$$

With the help of kernel function $K(x_i, x_j) = \phi(x_i)^T \phi(x_j)$, it makes a great reduction of computational cost owing to avoidance of gaining the concrete feature vector and mapping function. After α is calculated from solving (2.12), the model can be obtained as follows:

$$
\begin{aligned}
f(x) &= \omega^T \phi(x) + b \\
&= \sum_{i=1}^{m} \alpha_i y_i K(x, x_i) + b
\end{aligned}
\tag{2.13}
$$

In addition to regression and time series prediction applications, excellent performances were introduced by [43, 44].

2.2.5 Multi-task Gaussian process regression

The GP is a probability distribution over functions and inference taking place directly in the space of functions. GP regression originated in geostatistics in 1967 and is known as

kriging [45]. It was introduced to the machine learning community from the inspiration to construct GP from neural networks [46]. Since GP is a nonparametric and interpretable Bayesian model, it was soon applied to learn the forward or inverse dynamics of robotic systems. Later modifications were mainly to improve its scalability and handle sparse data [47]. The advantage of GP is that it provides a well-calibrated uncertainty of the prediction. However, ordinary GP can only model output separately and ignore the correlation in outputs [48]. To benefit related tasks and not hurt performance when these tasks are unrelated, a multi-output GP (MOGP) was proposed to improve prediction accuracy[49, 50]. Here is a brief description of MOGP.

Given inputs $\boldsymbol{x} \in \Re^{n \times P}$ and D outputs $\boldsymbol{y} = \{\boldsymbol{y_1}, \boldsymbol{y_2}, ...\boldsymbol{y_D}\}$, its relationship is defined as follows:

$$\boldsymbol{y_d} = f_t(\boldsymbol{x}) + \epsilon_d, d = [1, D] \tag{2.14}$$

where $\boldsymbol{y_d} \in \Re^{n \times 1}$, $\epsilon_d \sim \mathcal{N}(0, \sigma_d^2)$ is the iid Gaussian noise of the dth output. The likelihood function of the D outputs is defined as follows:

$$p(\boldsymbol{y}|\boldsymbol{f}, \boldsymbol{x}, \Sigma_s) = \mathcal{N}(f(\boldsymbol{x}), \Sigma_s) \tag{2.15}$$

where Σ_s represents a diagonal matrix with elements σ_d^2.

Given D inputs $X = \{X_1, X_2, ..., X_D\}$, $X_D \in \Re^{n \times P}$, and M outputs $\boldsymbol{y}$, the posterior distribution of $\boldsymbol{f}(\boldsymbol{x_*}) = \{f_1(x_*), ..., f_M(x_*)\}$ at a new test point $\boldsymbol{x_*}$ is derived analytically as

$$\boldsymbol{f}(\boldsymbol{x_*})|X, \boldsymbol{y}, \boldsymbol{x_*} \sim \mathcal{N}(\hat{\boldsymbol{f}}(\boldsymbol{x_*}), \Sigma_*) \tag{2.16}$$

where $\hat{\boldsymbol{f}}(\boldsymbol{x_*})$ is a mean value and Σ_* is a covariance matrix. They are computed by Eq. (2.17) and Eq. (2.18)

$$\hat{\boldsymbol{f}}(\boldsymbol{x_*}) = K_{\boldsymbol{f*}}^T (K_{ff} + \Sigma_M)^{-1} \boldsymbol{y} \tag{2.17}$$

$$\Sigma_* = K_{**} - K_{\boldsymbol{f*}}^T (K_{ff} + \Sigma_M)^{-1} K_{\boldsymbol{f*}} \tag{2.18}$$

where $K_{\boldsymbol{f*}} = K(X_d, \boldsymbol{x_*}) \in \Re^{nD \times D}$ for d = [1,D], it has blocks $K_{dd'} = [k_{dd'}(\boldsymbol{x_i}, \boldsymbol{x_*})]$ for i = [1,n] and d,d'=[1,D]; $K_{**} \in \Re^{D \times D}$ has elements $k_{dd'}(\boldsymbol{x_*}, \boldsymbol{x_*})$ for d,d'=[1,D]; k is a specifying kernel function; $\Sigma_M = \Sigma_s \otimes I_n \in \Re^{nD \times nD}$ is a diagonal noise matrix; the block partitioned matrix $K_{ff} \in \Re^{nD \times nD}$ is shown by

$$K_{ff} = \begin{pmatrix} K_{11}(X, X) & \cdots & K_{1D}(X, X) \\ K_{21}(X, X) & \cdots & K_{2D}(X, X) \\ \vdots & \cdots & \vdots \\ dK_{D1}(X, X) & \cdots & K_{DD}(X, X) \end{pmatrix} \tag{2.19}$$

In order to estimate parameters $\boldsymbol{\theta}$ and σ in kernel function k and ϵ separately, the negative log marginal likelihood (NLML) is minimized as

$$\boldsymbol{\theta}_{opt} = \underset{\boldsymbol{\theta}}{argmin}(-log \, p(\boldsymbol{y}|X, \boldsymbol{\theta})) \tag{2.20}$$
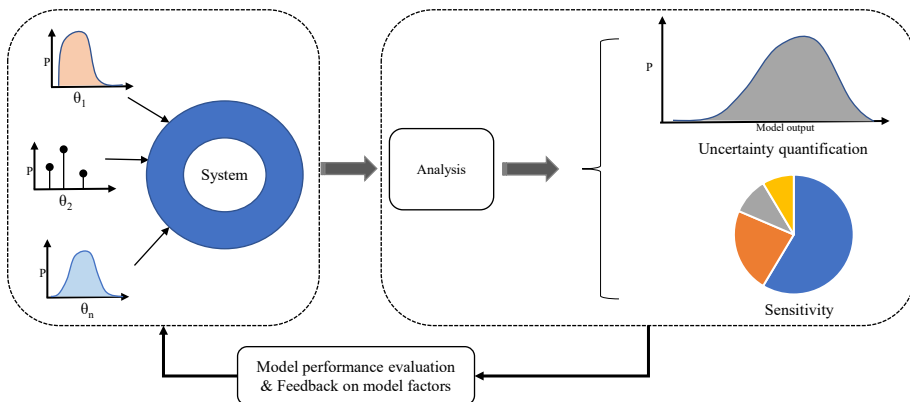
Figure 2.2: Schematic illustration of UA and SA.

where $-log\,p(\boldsymbol{y}|X,\boldsymbol{\theta})$ is defined as

$$
\begin{aligned}
-log\,p(\boldsymbol{y}|X,\boldsymbol{\theta}) &= \frac{1}{2}\boldsymbol{y}^T(K_{ff}+\sigma_d^2 I)^{-1}\boldsymbol{y} \\
&+\frac{1}{2}\,log\,|K_{ff}+\sigma_d^2 I| + \frac{n}{2}\,log\,2\pi
\end{aligned}
\tag{2.21}
$$

NLML has a well-known drawback of huge computational cost ($\mathcal{O}(n^3)$). As MOGPs need to consider the covariances between D outputs, leading to $\mathcal{O}(n^3 D^3)$ scaling. Therefore, the sparse variational inference was developed to approximate the posterior distribution[51]. It learn a set of $M \ll n$ inducing variables $\boldsymbol{u} \in \Re^M$, defined at inducing locations $Z = \{z_m\}_{m=1}^{M}, z_m \in \Re^P$ [52]. We then collect function values $f(Z)$ in variables $\boldsymbol{u}$, and specify a free mean and variance:

$$
q(\boldsymbol{u}) = \mathcal{N}(\boldsymbol{u},\boldsymbol{m},S)
\tag{2.22}
$$

where $\boldsymbol{m} \in \Re^M$ and $S \in \Re^{M \times M}$. Next, the evidence lower bound (ELBO) is maximized to learn inducing locations and variational parameters:

$$
\mathcal{L} = \mathbb{E}_{q(\boldsymbol{f})}[log\,p(\boldsymbol{y}|\boldsymbol{f})] - KL(q(\boldsymbol{u})||p(\boldsymbol{u}))
\tag{2.23}
$$

where $q(\boldsymbol{f}) = \mathcal{N}(\boldsymbol{f},\mu_{\boldsymbol{f}},\Sigma_{\boldsymbol{f}})$ is the approximated posterior over the function values at the data points implied by conditioning on $\boldsymbol{u}$,

$$
\mu_{\boldsymbol{f}} = \boldsymbol{k}_Z^T K_Z^{-1}\boldsymbol{m}
\tag{2.24}
$$

$$
\Sigma_{\boldsymbol{f}} = K_{ff} + \boldsymbol{k}_Z^T K_Z^{-1}(S - K_Z)K_Z^{-1}\boldsymbol{k}_Z
\tag{2.25}
$$

where $\boldsymbol{k}_Z$ and K_Z is analogous to K_{f*} and K_{ff} respectively. Sparse variance inference can implement an efficient GP modelling with $\mathcal{O}(M^3 D)$

2.2.6 Uncertainty analysis

UA aims to analyze how much the uncertainty of model output is caused by the variation of input parameters. In contrast, SA is to investigate how much the variation of model output is influenced by each input parameter. UA is a forward analysis by the propagation of uncertainty in input parameters to the output while SA is a backward analysis by apportioning output uncertainty to input parameters. First of all, UA is to sample data points from the assumed data distributions, such as Gaussian distribution and uniform distribution. Next, the sampled data is fed into a system or model to obtain the output values. After that, probability distribution and statistics are used to represent the uncertainty of output values.

2.2.7 Sensitivity analysis

SA, in general, is made up of variance-based and density-based methods. Variance-based methods include Sobol [53], the Fourier Amplitude Sensitivity Test (FAST) [54], and the Extend-FAST (EFAST) [55] and so on. A well-known advantage of variance-based methods is their ability to quantify the individual parameter contribution and the contribution resulting from parameter interactions [56]. However, variance-based methods do not completely represent the output uncertainty when the model output is highly skewed [57]. Francesca et al. came up with a novel SA method called 'PAWN' that characterizes the output distribution by its cumulative distribution function (CDF) instead of probability distribution function (PDF) [58]. One advantage of PAWN is that it hugely reduces computational cost because there is no need to compute unknown parameters for the approximation of empirical CDF. Another advantage is that sensitivity indices can be easily obtained, by considering the entire range of variation of the model output or a sub-range. This dissertation mainly introduces Sobol and PAWN used in the research papers.

A generic model is described as follows.

$$Y = f(X_1, X_2, ..., X_M) \tag{2.26}$$

where Y is the model output of interest; $X = (X_1, X_2, ..., X_M) \in \mathbb{R}^{M \times 1}$ is the model input which contains M factors; $f(X)$ can represent abstract models (data-driven models, mathematical model, or defined function) and mechanical models (robots). In this study, $f(X)$ corresponds to SSI-cov and Hierarchical clustering methods.

Sobol's method is based on the total variance decomposition.

$$V(Y) = \sum_{i=1}^{M} V_i + \sum_{i=1}^{M-1} \sum_{j=i+1}^{M} V_{ij} + \cdots + V_{1,\cdots,M} \tag{2.27}$$

where $V(Y)$ is the variance of model output Y; V_i is the variance contribution of X_i to the model output; V_{ij} is the variance from the interaction between X_i and X_j; $V_{1,\cdots,M}$ represents the variance induced by the interaction between M parameters. The V_i is addressed as the first-order or main effect of X_i on Y. Therefore the first-order sensitivity index (S_i) of X_i is computed by Equ. 2.28. The total sensitivity index (S_{Ti}) is obtained by Equ.2.29.
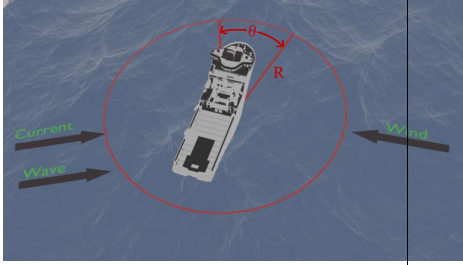
(a) DP operations of a vessel at sea.
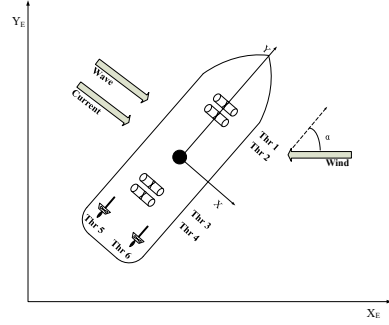
(b) Thrusters' configuration of DP vessel.

Figure 2.3: Illustration of DP simulator.

$$S_i = \frac{V_i}{V(Y)} \tag{2.28}$$

$$S_{Ti} = \frac{V_i + V_{ij} + \cdots + V_{1,\cdots,M}}{V(Y)} = 1 - \frac{V_{\sim i}}{V(Y)} \tag{2.29}$$

where $V_{\sim i}$ represents the total variance contribution of remaining parameters to Y given X_i. The detailed description can be referred to [59].

'PAWN' is a CDF-based GSA method [57]. The main principle is to estimate the difference between unconditional CDF and conditional CDF using Kolmogorov–Smirnov (KS) test.

$$\begin{cases} \hat{S}_i = \max_{k=1,\ldots,M} KS(I_k) \\ KS(I_k) = \max_y |F_y(y) - F_{y|\tilde{x}_i}(y|\tilde{x}_i \in I_k)| \end{cases} \tag{2.30}$$

where KS is Kolmogorov-Smirnov statistic; $F_y(y)$ is unconditional CDF where $y \subseteq Y$ and $F_{y|\tilde{x}_i}(y|\tilde{x}_i \in I_k)$ is conditional CDF where $\tilde{x}_i$ is fixed. The detailed information can be found in [60].

2.3 Experimental platforms and data collection

This dissertation focuses on three aspects of decision support of marine operations: ship DP capability analysis, structural health monitoring, and ship behaviour analysis. For DP capability analysis, a simulator is used to generate simulation data to verify the feasibility of the proposed method. For structural health monitoring, both model-scale data and full-scale data are used to analyze modal parameters. For ship behaviour analysis, real AIS data is used for analysis and modelling. This section introduces the experimental platform and data collection procedure.

Table 2.1: Sea states

Beaufort description	Wind velocity (m/s)	Wave height (m)	Wave period (s)	Current speed(m/s)
Fresh breeze	7.90	1.30	6.50	0.75
Strong breeze	13.80	3.10	8.50	0.75

2.3.1 Ship dynamic positioning data

the DP data are generated from a professional simulator in the Offshore Simulator Centre — the world's most advanced provider of simulators for demanding marine operations[2]. Figure 2.3a illustrates the simulator conducting DP operation under the impact of environmental disturbances. Its position is limited within a red circle whose diameter is set as R = 3m. The limit of the heading is restricted by the red sector whose angle is represented as $\theta = 6°$. Figure 2.3b shows the environmental effects on the ship. Wind with an attack angle of α can be changed and set as $[45°, 90°, 135°]$ for different scenarios. Current and waves coming from other directions are fixed in the study. In Figure 2.3b, the Earth-fixed reference frame is denoted as (X_E, Y_E). The body-fixed reference frame (X,Y) is fixed on the body of the vessel. Its origin is the vessel's center of gravity. The DP vessel is equipped with six thrusters including four tunnel thrusters (Thruster 1-4) and two main thrusters (Thruster 5 and 6). In the simulator, the sea state, the thruster state, and the desired position are all adjustable.

Two different sea states are investigated as shown in Table 2.1. The desired position is set to (0,0). Thruster states involve various thruster failure modes. The combination of sea states and thruster failures is shown in Table 2.2. They are 'strong breeze 45°', 'strong breeze 90°', 'strong breeze 135°', and 'fresh breeze 45°'. For 'strong breeze 45°', there are seven different thruster failure modes represented by binary string: '011111', '101111', '110111', '111011', '111101', '111110', '110110'. Here, '0' denotes the thruster is malfunctioning; '1' denotes the thruster is working normally. For example, '101111' indicates the second thruster is malfunctioning while the others are working normally. The sampling frequency is set as 20HZ.

2.3.2 Structural response data

Ice-structure interaction data are collected from model-scale experiments and full-scale experiments.

Data collection from model-scale experiment

All model-scale tests are carried out in the Hamburg Ship Model Basin's (HSVA) large ice model basin [3]. The experiment setup was designed with a flexible foundation with adjustable mass and stiffness to mimic certain dynamic characteristics of the structure and a rigid model. Three Triax accelerometers are used to monitor the ice-induced vibrations of the structures in x- and y-direction (loading direction and perpendicular in-plane motion). The setup is shown in Figure 2.4a ([62]). The data were collected under different structural and ice-related properties. The main tests and corresponding properties are shown in Table 2.3.

---

[2]https://osc.no/
[3]https://www.hsva.de/

Table 2.2: Environment and thruster failures setting for significance analysis.

Sea states	Attack angle [deg]	Thruster failure
Strong breeze	45	011111
		101111
		110111
		111011
		111101
		111110
		110110
Strong breeze	90	101111
		110110
Strong breeze	135	101111
		110110
Fresh breeze	45	101111
		110110

Table 2.3: The main test runs and the corresponding parameters settings.

Run	Model	Ice type	Ice drift velocity	Ice thickness	Flexural strength
25010	9200 MDOF	ICMI	4-150 mm/s	23mm	86kPa
32010	9500 MDOF	Model ice	4-150 mm/s	41mm	56kPa

Different runs are designed with different global stiffness of the setup. Hydrodynamic added masses for models '9200' and '9500' are 16kg and 19kg separately. 'MDOF' represents the structure as multi-degree-freedom vibration which has two dominant frequencies ($f_1 = 2.81HZ$ and $f_2 = 3.77HZ$). Ice types are HSVA's standard model ice and an improved crushing model ice [61]. Model ice is generated by exposing the water surface to cooled air. The current model ice was not always ideal for crushing failure type of dynamic ice-structure interaction tests. Therefore, an alternative wave ice was proposed by simulating sea wave effects during water freezing. The first three test runs contain ice drift velocity from 4-150 mm/s whilst the ice velocity of '460101' starts from 14 to 150mm/s. Based on different ice velocities, the measurements are grouped into different corresponding ice failure types (intermittent crushing (IC), frequency lock-in (FLI), and continuous crushing (CC)). The full data set is described by Stange et al. [62].

Data collection from full-scale data

The full-scale experiments are conducted on the Norströmsgrund lighthouse which is located in the Gulf of Bothnia. As shown in Figure 4.9, It is a gravity-based concrete structure with a wall thickness varying between 0.2m at the top and 1.4m at the mean water level ([35]). The diameter of the structure at the mean water level was 7.5 m at an elevation of +14.2 m. Nine panels were installed across the outer surface at the mean water level to measure the ice forces. Four accelerometers were equipped at different

(a) Model-scale experiment ([61])   (b) Full-scale experiment ([35])
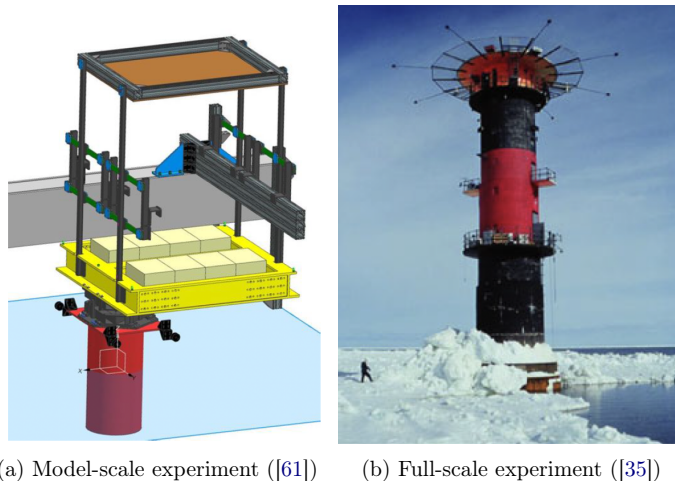
Figure 2.4: The experiment setup for model-scale and full-scale test

positions of the lighthouse. The detailed description can be found in [35, 63]

2.3.3 AIS data

Ship behaviour analysis uses the three-year historical AIS data in the Olso area from 2017-2019. AIS data is open source and can be downloaded from marine traffic [4], marine cadastre [5]. AIS data contains abundant information, including the time-stamp for each recorded data point and the ship's MMSI, DWT, speed over ground (SOG), course over ground (COG), and ship length. Because it offers rich information in real time, AIS data is widely applied for navigation safety, collision avoidance, and trajectory analysis [64].

For the sake of ship behaviour analysis for port management, some important features need to be extracted such as ship type, deadweight, ship destination, arrival time, and berthing time. Sea-web [6], marine traffic [1], and myshiptracking[7] is accessed to match ship type and deadweight information based on MMSI. In the next step, ship types need to be changed into numerical values for the sake of the convenience of data analysis. Numbers from '1' to '6' were assigned to 'Cargo', 'Carrier', 'Container', 'Tanker', 'Cruise', and 'Ferry', respectively. Other ship types, such as yacht and fishing vessels, are excluded from this study.

To obtain the ship's arrival time and its berthing time, it is necessary to pinpoint the ship's destination, i.e. berth. To this end, position points (Latitude (Lat), Longitude (Lon)) with 'SOG = 0' are selected. 'SOG = 0' means the ship is mooring. However, ships may moor at a position outside of a berth. In order to address this problem, a clustering approach (DBSCAN) is introduced to identify real berths.

---

[1]https://www.marinetraffic.com/

[2]https://marinecadastre.gov/ais/

[6]https://maritime.ihs.com/

[7]https://www.myshiptracking.com/vessels

For ship trajectory prediction, features like the ship's position, COG, and SOG are used as model inputs. Considering missing data, non-identical time steps, and anomaly data, the ship trajectories need processing further. Therefore, an interpolation method (UnivariateSpline) is introduced to unify the time step of AIS data to 5 seconds. Such unification facilitates determining the future position of a sip and provides a foundation for model training. In order to improve the accuracy of interpolation, Geographical coordinates (Lat and Lon) are transformed to meters by Pythagoras formula:

$$\begin{cases} dist = \sqrt{(\Delta lat)^2 + (\Delta long)^2} \\ \Delta lat = R \times (lat2 - lat1) \\ \Delta long = R \times \cos(central\_long \times \frac{2\pi}{360}) \end{cases} \tag{2.31}$$

where R is the maximum circumference of the Earth (40075000 m); $(lat1, long1)$ and $(lat2, long2)$ are the geographical coordinates of two data points. $dist$ is the distance (m). $central\_long$ denotes the longitude at which Olso locates.

2.4 Chapter summary

Table. 2.4 shows how models and data sources are contained in different chapters. Chapter 3 presents the investigation of thrusters' importance to the vessel's DP capability, which uses data in Section 2.3.1. Chapter 4 describes the modal parameters identification of offshore structures, covering the data in Sections 2.3.2. Chapter 5 demonstrates case studies on ship behaviour analysis, the data in Section 2.3.3 are used. These three case studies form the main part of this dissertation, with the main goal to implement data analysis and modelling for onboard support of marine operations.

Table 2.4: Interconnection of the data analysis methods and data source contained in different Chapters.

Method \ Data	SVM	SA	UA	SSI-cov	clustering	DTW	GP
Section 2.3.1	Chapter 3						
Section 2.3.2			Chapter 4				
Section 2.3.3					Chapter 5		

This chapter introduces the fundamentals of data analysis methods. The statistical analysis methods (hierarchical clustering, DBSCAN, DTW, and SSI-cov), machine learning models (SVM and MOGP), UA, and SA that are used in this dissertation are also briefly introduced. It is worth noting that different data types have different errors and hence need different data pre-processing approaches. This chapter does not give a detailed description of data pre-processing but it can be found in research papers.

Case study: Thrusters' significance analysis in ship dynamic positioning operations

This chapter presents research results from paper I. Vessels with DP systems can automatically maintain the desired position under the influence of environments, which is critical to ensure safe offshore operations. In order to prevent the occurrence of a loss of position, DP 2 and DP 3 are designed with redundant power systems to guarantee the vessel's DP capability [65]. Once a thruster failure occurs, the power allocation system could reallocate the desired thrust to the remaining thrusters so as to make a vessel still keep its position.

Environmental factors cause an impact on the thrust allocation of thrusters. In other words, thrusters' importance varies over different environmental forces. Understanding the interaction between thrusters and environments is crucial to a vessel's DP capability improvement, especially in the case of thruster failures.

In this chapter, a novel methodology is proposed to analyze the significance of each thruster on DP capability under the influence of environments and thruster failures. It could not only provide onboard support for improving DP capability but also give guidance for power system design as well as thrusters' maintenance with the help of statistical analysis and SA.

3.1 Workflow of thrusters' significance analysis

Figure 4.5 outlines the procedure of thrusters' significance analysis in DP operations. DP data is generated by a DP simulator that is considered a digital model of a real vessel. The inputs to the simulator are adjustable, such as sea states, desired position, and thrusters' states. The details have been elaborated in 2.3.1. The second part is data analysis composed of data pre-processing and significance analysis. The final part is onboard support of the real vessel's DP operations as well as system optimization.

Significance analysis is to identify the importance of thrusters to the vessel's DP capability. It includes statistical analysis and sensitivity analysis that analyzes the significance of thrusters from different aspects. Statistical analysis, as a supplementary instruction for SA, focuses on statistics of DP data such as mean value, variance, and PDF [66]. SA is responsible for estimating the sensitivity of each thruster to the DP capability. It comprises three steps: 1) proposing a synthesized criterion to quantify DP capability; 2) applying an optimal ML method to create a surrogate model between thrusters and DP capability; 3) using the SA method (PAWN) to analyze sensitivity indices of thrusters [58].
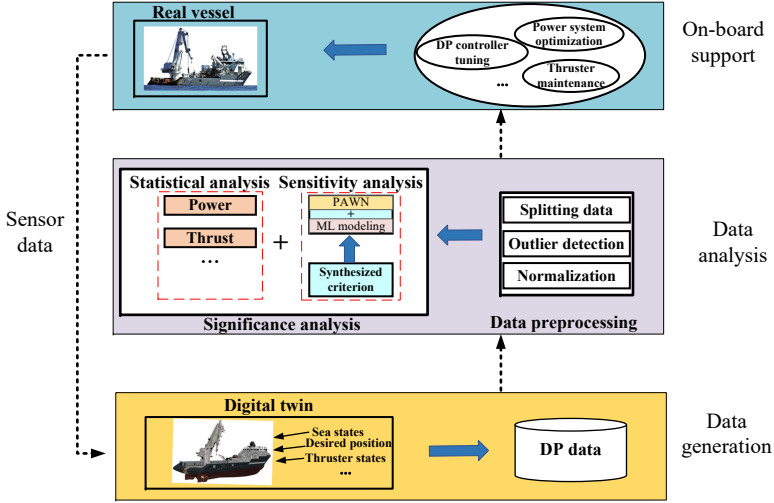
Figure 3.1: The system structure of significance analysis of thrusters in DP operations.

3.1.1 Synthesized criterion of DP capability

According to the DP capability level in 'DNVGL-ST-0111' standard, the assessment
of station-keeping capability is mainly based on statistics of the position deviation and
heading deviation. Therefore, position and heading should be integrated into the synthe-
sized criterion. In addition, for ensuring the safety and accuracy of DP operations, the
DP vessel has a higher power requirement than other conventional vessels [67]. There-
fore, power consumption is also taken into consideration in this criterion. As a result, we
create a synthesized criterion by Eq. (3.1) to lump the above-mentioned ship parameters
together, with extra modification to make it adapt to the SA method.

$$
\begin{cases}
V = \omega_1 \times D + \omega_2 \times A + \omega_3 \times P \\
\omega_1 + \omega_2 + \omega_3 = 1 \\
Cri = -ln(V) \qquad\qquad V > 0.
\end{cases}
\tag{3.1}
$$

where ω_1, ω_2, and ω_3 are weighting factors within [0,1]; D is position deviation computed
by the distance between current and original position; A denotes the heading angle vari-
ation; P represents total power consumed by thrusters; Cri is the synthesized criterion
computed by the inverse of the monotone increasing function 'ln'. The larger V is, the
worse the positioning capability (Cri). Compared to the exponential function in the
interval [0,1], the minus of 'ln' function can amplify the value of V to better reflect the
distinction of positioning capability [68]. Cri will be used as the model output when
ML trains a surrogate model between thrusters' parameters and DP capability.

3.1.2 Sensitivity analysis

The process of SA executed by PAWN combined with SVM is shown in Algorithm 2.
Note, This algorithm can be applied for other ML methods as well, but some adjustments

to the training parameters may be necessary. In this algorithm, 'LIBSVM' is used as
an SVM tool to train the surrogate model [69]. The model training parameters like 's',
't', 'bestc', 'bestg', 'p', 'v', and the introduction of functions like 'SVMcgForRegress',
'libsvmtrain', and 'libsvmpredict' can be found in [69]. This algorithm mainly includes
three parts. The first part is modelling (lines 9-13). The thrust of all thrusters is
the model input, and the positioning capability as defined by Cri above is the model
output. SVM is employed to construct a surrogate model between the model input
and output. The second part is resampling (lines 14-15). 'Unconditional_sampling' is
used to generate unconditional samples; 'PAWN_sampling' is used to gain conditional
samples. The last part is sensitivity index computation (line 16). The 'PAWN' indices of
all thrusters are computed by 'PAWN_index'. Its function is shown in lines 1-7. Lines
2-3 represent the model evaluation of the unconditional and conditional samples. Lines
14-16 are to compute the 'PAWN' index.

Algorithm 2: SA algorithm

Input: $Thrust$,Cri s, t, p, v

Output: SA_index

1 **Function** $PAWN\_index(Xu, XX, model)$:
2     $Yu \leftarrow libsvmpredict(Xu, model)$
3     $YY \leftarrow libsvmpredict(XX, model)$
4     $[YF, Fu, Fc] \leftarrow PAWN\_cdf(Yu, YY)$
5     $KS \leftarrow PAWN\_ks(YF, Fu, Fc)$
6     $index \leftarrow max(KS)$
7   **return** index
8 **for** $i = 1 : num$ **do**
9     $X \leftarrow Thrust$
10     $Y \leftarrow Cri$
11     $[bestc, bestg] \leftarrow SVMcgForRegress(X, Y)$
12     $cmd \leftarrow [s, t, bestc, bestg, p, v]$
13     $model \leftarrow libsvmtrain(X, Y, cmd)$
14     $U \leftarrow Unconditional\_sampling$
15     $C \leftarrow PAWN\_sampling$
16     $index(i) \leftarrow PAWN\_index(U, C, model)$
17 **end**
18 $SA\_index \leftarrow index/num$

3.2 Experimental results

3.2.1 An optimal ML selection based on Ishigami function

This experiment compares three common ML methods Backpropagation (BP) neural
network, regularized extreme learning machine (RELM), and SVM. In order to select
the optimal ML model, the Ishigami function is used as a ground truth. It is shown in
Eq. (3.2).

$$y = sin(\chi_1) + asin(\chi_2)^2 + b\chi_3^4 sin(\chi_1) \tag{3.2}$$

where a and b are random constants that can influence the sensitivity index of χ_i, $i \in \{1, 2, 3\}$. χ_i follows a uniform distribution over $[-\pi, \pi]$. Here, we set $a = 2$ and $b = 1$.

Figure 3.2 displays SA results as well as a benchmark value. The x-axis is the size of the sample and the y-axis represents the estimated PAWN index. The dotted line is the benchmark value of sensitivity indices of the three parameters χ_i in Eq. (3.2). The corresponding sensitivity indices are S_1=0.53, S_2=0.19, and S_3=0.35, respectively. It is evident that both BP and RELM cannot figure sensitivity index out correctly; whereas PAWN combined with SVM has a better approximation to the benchmark. Therefore, SVM is selected as modelling method in the follow-up sensitivity analysis of thrusters in different scenarios.
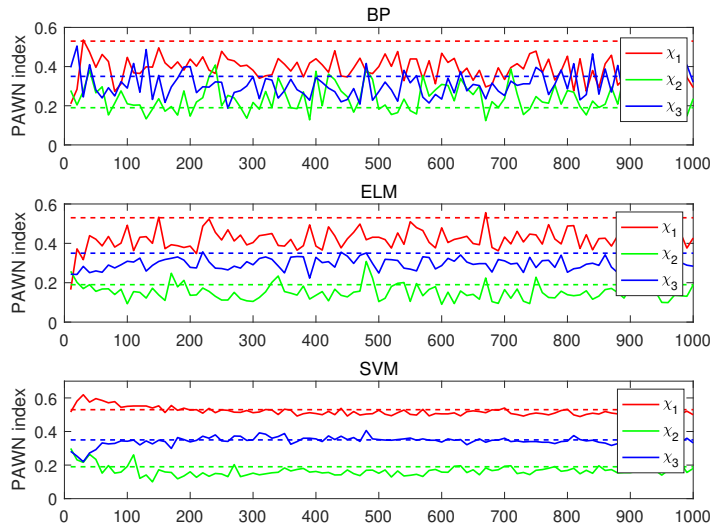
Figure 3.2: SA results computed by PAWN based on different ML methods.

3.2.2 Significance analysis in different thruster failure modes at two sea states

This section mainly analyzes and compares SA results in different environmental factors and thruster conditions. Table 3.1 lists the SA results of thruster failures at the strong breeze and fresh breeze sea states. It is found that thruster 5 is more significant than the rest of the thrusters in most cases. Its contribution accounts for around 30% $\sim$ 40%. Especially, when thruster 6 fails to work, the significance of thruster 5 exceeds 35% because thruster 5 as the only main propeller must generate much more thrust to counteract the influence of environmental disturbances. When one thruster failure occurs, the significance of thrusters that play a complementary role will have a significant

Table 3.1: SA results of thruster failures in the strong breeze and fresh breeze.

Sea states	Direction (deg)	Thruster failure	PAWN index					
			Thr1	Thr2	Thr3	Thr4	Thr5	Thr6
Strong breeze	45	111111	0.1342	0.1040	0.1576	0.2246	**0.2976**	0.0817
		011111	0	**0.3701**	0.1448	0.1480	0.2284	0.1087
		101111	0.2642	0	0.0604	0.1069	**0.3222**	0.2459
		110111	0.1992	0.2080	0	0.0850	**0.3058**	0.2019
		111011	0.1415	0.2483	0.0853	0	**0.3472**	0.1775
		111101	0.2629	0.1839	0.1433	0.1098	0	**0.3000**
		111110	0.1674	0.1225	0.1435	0.1456	**0.4209**	0
		110110	0.2106	0.2026	0	0.2050	**0.3818**	0
	90	111111	**0.2877**	0.1211	0.0829	0.1485	0.1313	0.2283
		101111	0.2723	0	0.1103	0.1100	**0.2985**	0.2089
		110110	0.0737	0.3337	0	0.2392	**0.3534**	0
	135	111111	0.1832	0.1638	0.1224	0.1888	**0.2544**	0.0873
		101111	0.0987	0	**0.3491**	0.3273	0.1268	0.0980
		110110	0.2285	**0.4016**	0	0.0997	0.2702	0
Fresh breeze	45	111111	0.1373	0.0729	0.1317	0.0771	**0.3460**	0.2350
		101111	0.2591	0	0.1007	0.0901	**0.3401**	0.2099
		110110	0.1826	0.2113	0	0.2282	**0.3780**	0

increase as shown in Table 3.1. For example, the PAWN index of thruster 6 increases from 8% to 30% when thruster 5 fails in 'strong breeze 45°'. The same happens to thrusters 1 and 2. For the case of '101111' in 'strong breeze 45°', for instance, the significance of thruster 1 rises by 13% up to 26.42%. For dual thruster failure '110110' in all sea states, at least two of tunnel thrusters' significance go up to over 20% compared with one thruster failure. That possibly results from the drastic variation of the ship heading. It is reflected from the above analysis that the significance of thrusters depends on the conjunction of sea states, wind direction as well as thruster failures.

From the perspective of statistics, thruster 6 has as much thrust as thruster 5 as shown in Figure 3.4. The mean and variance of thrust generated by thruster 5 are the same as those generated by thruster 6. The two thrusters also consume the same amount of power and have similar statistical features. But observing results obtained by the proposed SA method in Figure 3.3, in which all SA indices are drawn as blue bars, shows that thruster 6 is far less significant than thruster 5. It is even less than thruster 2. Such a finding does make sense because thruster 5 generated the maximum moment to resist the wind forces when the wind attack angle is 45°. In other words, thruster 5 plays a more important role than thruster 6 in terms of DP capability improvement. The comparison reveals that SA did discover valuable information that statistical analysis can not give.

3.2.3 Real-time computation of thrusters' sensitivity

Figure 3.5 shows the variation of sensitivity indices of thrusters over time. The horizontal axis denotes sensitivity index is computed at a window time of 25s that comprises 500 sample points. Evidently, the proposed method is able to gain the contribution of each thruster to the DP capability in the process of vessel counteracting against environmental forces. Especially, when thruster 1 shuts down at 650s depicted by a red circle, the importance of thruster 1 becomes 0 thereafter. On the other hand, thruster 2 plays a
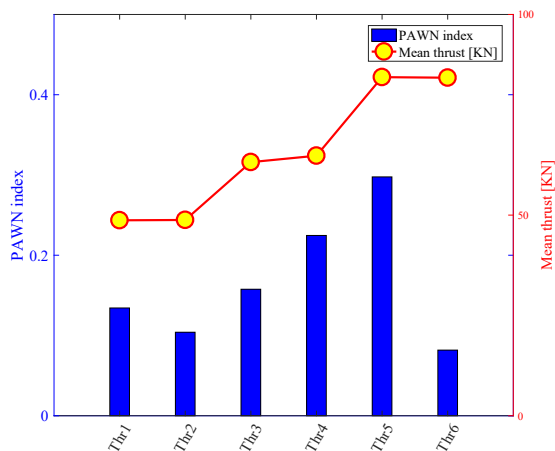
Figure 3.3: The SA result and average thrust of 6 thrusters for '111111' in 'strong breeze 45°'.

more and more important role since this point. This is because thruster 1 and 2 are
bow thrusters, as shown in Figure 2.3b, the malfunction of thruster 1 leads to the rise
of thruster 2 importance in the long term. To sum up, the proposed method is capable
of finding the contribution of all thrusters in a real-time manner.

3.3 Chapter summary

This chapter investigates the significance of thrusters to a vessel's DP capability under
the impacts of thruster failures and environmental factors. Due to DP capability being
an abstract conception, it is, hence, quantified by a synthesized criterion. Next, the
Ishigami function is used to select an optimal ML method to build a surrogate model
between thrust and DP capability. Through the comparison with other ML methods,
SVM is selected for the subsequent SA. Based on the significance analysis of thrusters, it
can be found that thrusters' importance varies over different environments and thruster
failures. It is verified that the proposed method can identify the importance of thrusters
during DP operations. Through an online case study, the proposed SA combined with
SVM did offer onboard support of DP operations.
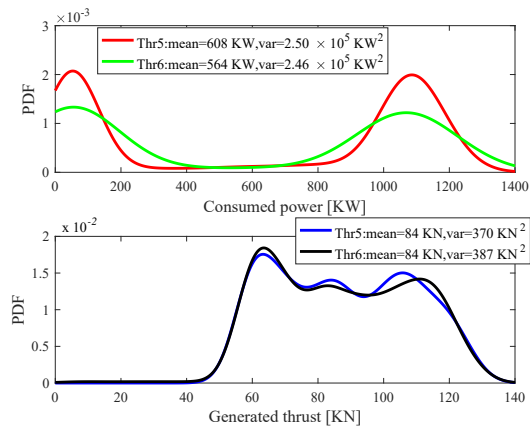
Figure 3.4: The PDF of consumed power and thrust generated by thrusters 5 and 6 for '111111' with strong breeze and $\alpha = 45°$.
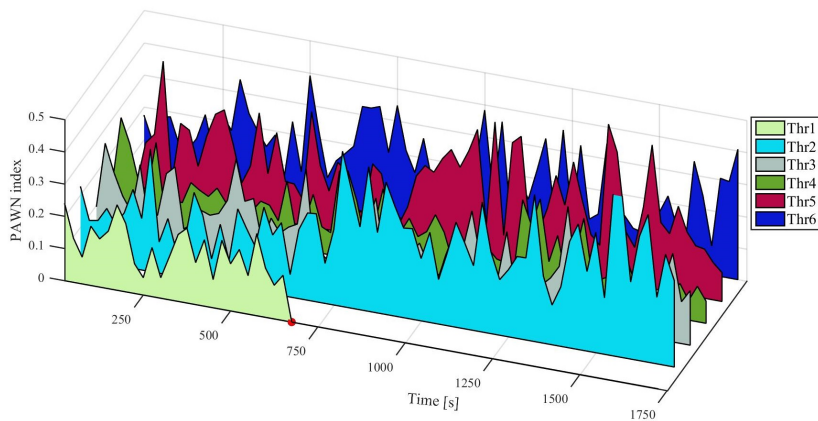
Figure 3.5: Real-time computation of the significance of thrusters.

Case study: Modal parameters identification for structural health monitoring

This chapter presents research results from papers III and IV.Offshore structures are prone to damage caused by ice-induced vibrations. It is presently unknown to what extent different ice conditions change the properties of the structure, such as natural frequency, damping ratio, and mode shape. Understanding the dynamic interaction between ice and structures are important for the operational ability of offshore structures [70, 71]. To address this problem, the covariance-driven stochastic subspace identification algorithm (SSI-cov) is introduced to identify modal parameters of a scale-model structure during ice-structure interactions. However, many user-defined parameters in SSI-cov could lead to inherent bias to the identified modal parameters[72]. Therefore, clustering technology is proposed to replace the slack value criterion for automatic parameters identification. Nevertheless, there are other parameters that may affect the identified parameters. On this basis, uncertainty analysis is, furthermore, applied to quantify the total uncertainty caused by other parameters selection, algorithm uncertainty.

4.1 Clustering for automated modal parameters identification

After SSI-cov analysis, poles at different system orders are obtained. A pole is considered stable if the deviances in frequency, damping, and normalized standard deviation of the frequency fulfill the predefined stability criterion. After that, a stabilization diagram is constructed by stable poles via taking frequency as abscissa and system order as ordinate [73]. Physical modes should then show up as vertical lines in the diagram.

To date, there are many suggested methods to automatically determine the modal parameters. Magalhaes et al. applied hierarchical clustering to identify the modes successfully based on the data from concrete arch bridge [74]. Reynders et al. introduced how to use hierarchical clustering to identify the physical modes based on single-mode validation criteria [75]. It does not require any user-specified parameter values. The validation example shows that hierarchical clustering has better robustness to identify modal parameters than the traditional identification approach. Inspired by this research, hierarchical clustering is used to identify the parameters of the ice-structure interaction model.

4.1.1 Modal parameters identification procedure

As shown in Figure 4.1, The procedure includes three parts: data preprocessing, SSI-cov analysis, and physical mode identification. Structural response data is collected from the model-scale test as described in 2.3.2. After data preprocessing, the SSI-cov is used to identify modes in the data. Physical modes are clustered by hierarchical clustering.
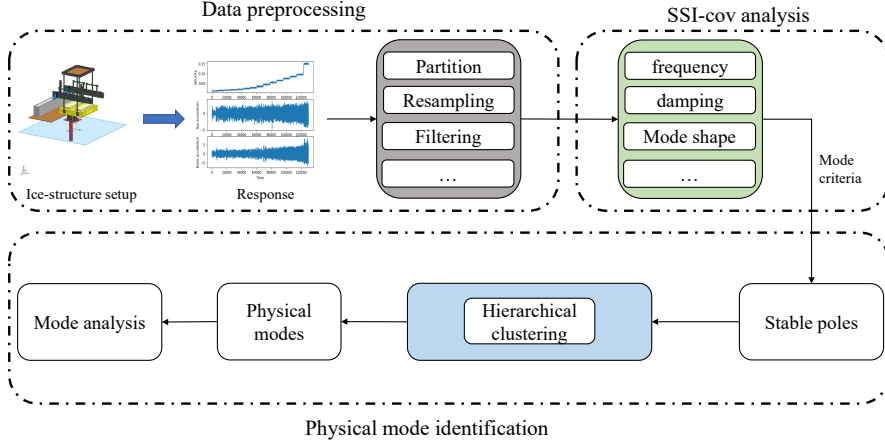
Figure 4.1: The procedure of automated modal parameters identification.

After the SSI-cov analysis, many poles can be obtained including stable/unstable poles. We need group poles with similar modal characteristics. This is commonly performed in a stabilization diagram, which shows the frequency of the poles on the horizontal axis and the order of the system on the vertical axis. A physical mode appears as a straight vertical line of poles, and the line with the corresponding lowest frequency is the first eigenfrequency, the column with the corresponding second lowest frequency is the second natural frequency, and so on. Poles that are not stacked on a vertical line are usually what is referred to as spurious poles/modes, i.e. modes without physical interpretation.

The major challenge lies in the process of choosing the poles that should be counted as part of the column of poles (mode), due to the fact that some lie at a slightly different frequency, have different damping values or mode shapes, and have different corresponding uncertainties. Therefore different techniques have emerged to handle the physical mode selection, where clustering algorithms have been suggested as an efficient technique to determine the physical modes. One of the popular methods is Hierarchical clustering [75, 39].

Hierarchical clustering depends on the proximity metric as described in 2.2.1. Therefore, in this study, eigenfrequency difference and modal assurance criteria (MAC) are used as distance measures in [76]. Its form is shown in Eq. (4.1)

$$d(k, l) = |f_k - f_l| + (1 - MAC(\phi_k, \phi_l)) \tag{4.1}$$

where f_k is the eigenfrequency of mode k; MAC is computed by Eq. (4.2)

$$MAC(\phi_k, \phi_l) = \frac{|\phi_k^T \phi_l|^2}{||\phi_k||_2^2 ||\phi_l||_2^2}, ||\phi_k||_2^2, ||\phi_l||_2^2 \neq 0 \tag{4.2}$$

where ϕ_k is the mode shape of mode k. The schematic diagram of physical modes identification using hierarchical clustering is shown in Fig. 4.2.
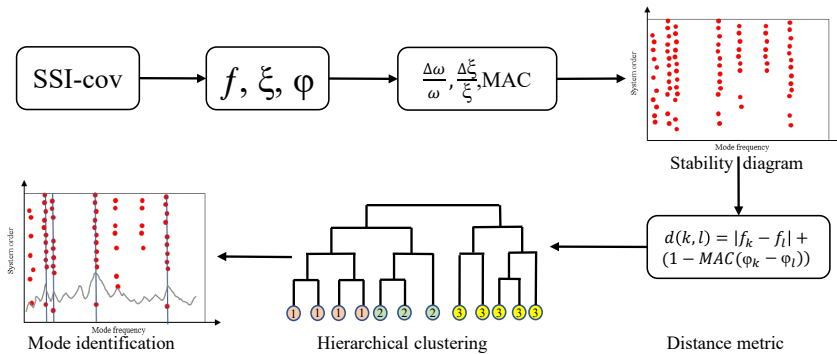
Figure 4.2: The procedure of automated modal parameters identification.

4.1.2 Experiment results

This experiment compares the hierarchical clustering method and slack value-based identification method with model-scale test 32010 as described in Table. 2.3. First,

The data whose ice velocities are from 4 mm/s and 150mm/s in datafile '32010' is chosen as two cases to compare these two methods. The benchmark values of the first two eigenfrequencies are 21.352 and 29.516 rad/s separately, which are estimated when the structure was moving in the open water [62]. Considering the uncertain factors, the benchmark values are expanded by 10% deviation to an interval: [19.22, 23.49] for the first frequency, [26.56, 32.47] for the second frequency. The cutoff and depth of the Hierarchical clustering algorithm are chosen as 0.1 and 5 separately based on data characteristics.

The focus of this study falls on the first three natural frequencies that represent the most concerning modes. The ice velocities range from 4-150 mm/s. The identified frequencies are shown in Table. 4.1. The bold numbers represent the successful identifications of natural frequencies by two methods. For IC, the identified first frequencies by slack value and Hierarchical clustering are 70.8, and 14.86, separately for different ice speeds. Results show that both methods fail to identify the first frequency. For FLI and CC, the first two natural frequencies identified by Hierarchical clustering are around 21 rad/s and 29 rad/s among different ice speeds. From this table, Hierarchical clustering renders bold numbers than the slack value-based approach. Table. 4.2 shows the identified damping of the first two modes. 'NULL' means the corresponding damping can not be obtained due to the failure of parameters identification. It can be seen that both methods achieve similar results. For 'FLI', the damping of the first mode is quite lower than that of the second mode, whereas it is the opposite for other cases. This trend probably results from the increase in ice velocity. Figure. 4.3a and Figure. 4.3b show the identified frequencies by these two methods in the case of ice velocity being 8 mm/s. Based on the referenced values of the first two natural frequencies, it is easily found that hierarchical clustering can identify these two frequencies correctly while the slack value-based method can not. As shown in Fig. 4.3a, some missing modes are supposed

Table 4.1: The identified frequencies by slack value and hierarchical clustering based on the data in '32010' under the ice failures of IC, FLI, and CC.

Frequency	Method	ice velocity (mm/s)														
		4	6	8	10	12	14	16	18	20	28	45	65	80	95	150
		IC	IC	FLI	FLI	FLI	FLI	FLI	FLI	FLI	FLI	FLI	CC	CC	CC	CC
First	Slack value	70.80	14.86	20.25	29.76	21.02	21.07	30.71	21.06	21.19	21.27	21.57	21.38	21.50	21.38	21.61
	Hierarchical	70.80	14.86	20.24	20.99	21.02	21.07	21.03	21.06	21.19	21.27	21.57	21.43	21.51	21.50	21.61
Second	Slack value	119.20	29.80	61.01	41.85	30.80	28.39	31.75	29.05	28.55	29.79	30.29	30.13	30.14	21.41	26.83
	Hierarchical	120.38	29.80	24.90	29.76	30.80	28.38	31.32	29.17	28.55	29.79	30.30	30.08	30.14	25.96	26.83
Third	Slack value	122.92	36.88	80.66	70.58	42.01	31.30	31.79	29.22	31.61	42.57	59.68	61.41	59.33	25.94	30.39
	Hierarchical	152.26	36.88	29.95	41.86	42.01	31.26	42.02	42.14	31.61	42.57	53.21	61.35	49.06	30.08	30.39

Table 4.2: The identified damping by slack value and hierarchical clustering based on the data in '32010' under the ice failures of IC, FLI, and CC.

Damping (%)	Method	ice velocity (mm/s)														
		4	6	8	10	12	14	16	18	20	28	45	65	80	95	150
		IC	IC	FLI	FLI	FLI	FLI	FLI	FLI	FLI	FLI	FLI	CC	CC	CC	CC
First	Slack value	NULL	NULL	0.11	NULL	0.08	0.01	NULL	0.08	0.03	0.02	0.64	2.44	1.67	1.64	2.47
	Hierarchical	NULL	NULL	0.13	0.05	0.08	0.01	0.08	0.08	0.03	0.02	0.64	2.35	1.73	1.51	2.47
Second	Slack value	NULL	0.54	0.65	2.57	3.23	1.82	2.69	3.69	1.89	2.53	0.92	0.79	1.17	1.02	0.37
	Hierarchical	NULL	0.54	1.02	2.57	3.23	1.57	3.26	3.58	1.89	2.51	0.91	0.82	1.17	1.09	0.38

(a) Slack value

(b) Hierarchical clustering

Figure 4.3: The comparison of slack value and hierarchical clustering when ice velocity is 8 mm/s.

to be identified in the stabilization diagram. For example, at the position of frequency 30 rad/s, there is supposed to be a mode that appears at the peak of the power spectrum. Based on the aforementioned analyses, it is concluded that hierarchical clustering outperforms the slack value-based approach as a whole.

37

4.2 Uncertainty and sensitivity analysis for modal parameters identification.

Although SSI-cov can effectively remove modes with large variances to obtain physical modes, its accuracy of modal parameter identification can not be guaranteed due to various assumptions and uncertain input parameters. For example, ice forces usually can not be represented as Gaussian white noise, and therefore the input violates the random white noise assumption of SSI-cov. In addition, the structure is described as a linear time-invariant system. Such assumptions could result in errors in the modal parameters identification. Besides that, SSI-cov contains many uncertain input parameters such as slack values, stability criterion, system orders, etc. The parameters' uncertainty is propagated to the identified parameters and results in the poor estimation of their values.

To reduce the uncertainty of slack values, clustering technologies were introduced for automated operational modal analysis (OMA) ([75]). Clustering approaches can implement parameter identification in an automatic way and avoid the artificial selection of slack values. Nevertheless, clustering algorithms will bring additional uncertainties as they contain several uncertain parameters and algorithm structure uncertainty (known as model uncertainty). As a consequence, a robust outlier detection was proposed to reduce statistical uncertainty caused by the clustering algorithm ([72]). However, it can not remove the algorithm uncertainty completely. In addition, clustering technologies did not consider the uncertainty from other input parameters of SSI-cov. Therefore, a modal parameters analysis framework based on UA and SA is proposed in this section. SA aims to pick up those factors that account for the most contributions to the model output. That is beneficial to lower the burden of UA caused by large amounts of variables.

4.2.1 Uncertainty and sensitivity analysis framework

This section is to introduce the proposed UA framework that is applied to identify modal parameters. Compared with traditional OMA, it can estimate the uncertainty bound of parameters induced by uncertain input variables. As shown in Fig. 4.4, the framework is mainly divided into three parts. The first part is to do a model evaluation based on the input parameters sampled from a hypothetical distribution. The second one is automatic parameter identification using the clustering method. The identified modal parameters are used for the following sensitivity analysis and uncertainty analysis. The workflow is first conducting SA to select the most important input variables, second doing UA based on the selected uncertain variables.

Modal parameters identification

In order to eliminate uncertainties in the Hierarchical clustering algorithm, two strategies are employed to improve the robustness of the algorithm. The first method is to use the Silhouette value to evaluate the clustering results. It measures how similar a point is to points in its own cluster when compared to points in other clusters. The higher it is, the better the samples are clustered. The principle is shown in Eq. 4.3.

$$s(i) = \frac{b(i) - a(i)}{max(b(i), a(i))} \tag{4.3}$$

where b(i) denotes the average distance of point 'i' with all points in the closest
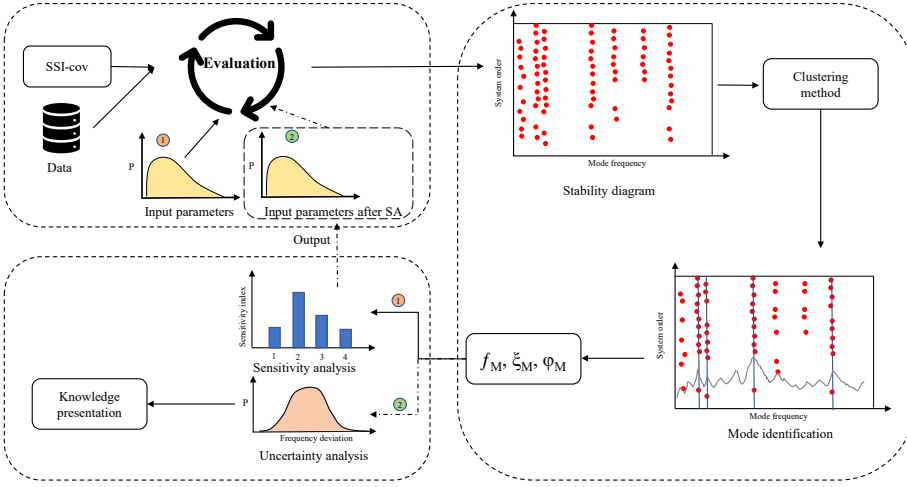
Figure 4.4: The framework of uncertainty analysis and sensitivity analysis on modal parameters identification.

cluster to its cluster; a(i) is the average distance of point i with other points in the same clusters; s(i) is the silhouette coefficient that ranges from [-1,1].

Due to the variability of modal estimates, outlier detection is used for penalizing undesirable modes in the final clusters to reduce identification uncertainties [72]. Equ. 4.4 defines the robust distance (RD).

$$RD(x) = d(x, \hat{\mu}_{MCD}, \hat{\Sigma}_{MCD}) \tag{4.4}$$

where MCD is the minimum covariance determinant that is used for outlying values detection. x represents frequency in this study; $\hat{\mu}_{MCD}$ denotes the MCD estimates of location; $\hat{\mu}_{MCD}$ is the covariance of MCD.

After that, a hierarchical tree could be created as shown in Fig. 4.2. The color of the leaves in the tree represents different clusters. If a cluster contains a pre-defined number of poles, the poles in this cluster render a physical mode as shown by the straight line in the stability diagram. After clustering, the stability diagram could show physical modes.

Uncertainty and sensitivity analysis

This study compares two SA approaches: Sobol and PAWN. Through the comparison of the parameters' importance computed by Sobol and PAWN, the optimal results are used for the subsequent UA. The flowchart of the two SA methods is shown in Fig. 4.5. First of all, some variables are defined manually. 'N' is the size of Sobol's sample. 'M' is the number of parameters in modal parameters identification. 'Unif' assumes the distribution function is uniform. 'LHS' is a Latin hypercube sampling strategy. 'NU' is the size of unconditional samples while 'NC' is the size of conditional samples. 'n' is the

conditional point from which conditional samples are sampled. Next, different sampling
methods are applied to generate different data samples. 'XA' is the sampled data that
is used to estimate 'YA' through model evaluation while 'XB' is used to compute 'YB'.
Likewise, 'XU' is to estimate 'YU', and 'XX' is to obtain 'YC'. After that, 'YA' is used
to estimate V_i by the Sobol method while 'YB' is employed to estimate $V_{\sim i}$. 'YU' is
to obtain $F_y(y)$ while is to get $F_{y|\tilde{x}_i}$. Followed by Sobol and PAWN analysis, two sets
of SA results can be obtained. They are compared to obtain convincing sensitivities
of all parameters. After important parameters are picked, LHS sampling is used to
generate a sample, and Model evaluation is conducted to carry out UQ of identified

Figure 4.5: The flowchart of Sobol and PAWN methods for sensitivity analysis.

4.2.2   Experiment results

For sensitivity analysis, first of all, the sampling strategy and distribution function is
selected as 'LHS' and uniform distribution. $\sigma = 40\% \times \mu$, where μ represents the
predefined values of seven input parameters [78]. Next, due to the time-consuming SSI-
cov computation, it is not practical to set up a larger 'N'. Therefore, the Sobol index
needs evaluation when 'N' is chosen as 1000 and 2000. 'M' means 7 parameters, including
the number of blocks (NB), block rows (BR), sampling frequency (SF), system order
(SO), and stability criteria (the deviance in frequency (SC-I), the deviance in damping

ratio (SC-II), and normalized standard deviation of the frequency (SC-III)). 'NU = 150', 'NC=100', and 'n=10' are referred to [60]. In Equ. 4.3, s is 0.5. As the first two modes (f_M, ξ_M, φ_M) are the modes in the model-scaled experiment that were designed to be easily excited by the ice force. Hence they are used for the output of interest in SA.

For UA, the sample size of uncertain parameters is 400. The benchmark values of the first and second natural frequencies are 21.35 and 29.52 (rad/s). SA is carried out using MATLAB toolbox ([77]). All MATLAB programs are run on the high-performance computer at the Norwegian University of Science and Technology ([79]).
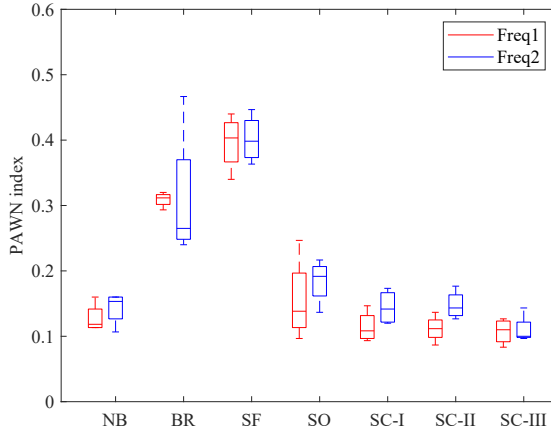
Figure 4.6: The statistics of PAWN indices of seven parameters for different ice velocities in data file 32010.

Sensitivity analysis

Due to the desperately high time cost of model evaluations, several random cases are chosen for SA. For data file 32010, SA is applied for four different cases in which ice velocities are 14, 16, 18, and 20 (mm/s) separately. For data file 25010, ice velocities are from 18 to 26 (mm/s) with the step of 2 (mm/s). For full-scale measurements, ice velocities vary from 30 to 200 (mm/s).

Figure 4.6 presents statistics of parameters' PAWN indices among different velocities 14, 16, 18, 20 mm/s while parameters' Sobol indices are shown in Figure 4.7. In both figures, the box represents the quantiles of parameters' sensitivity on the first and second natural frequencies. The PAWN indices of BR and SF are larger than the remaining overall. Therefore, BR and SF are chosen as uncertain factors for UA on data file 32010. In addition, for different cases, PAWN indices of different parameters have small variations. Nevertheless, in Figure 4.7, Sobol indices of parameters vary drastically and are hard to rank. Therefore, the PAWN method has better performance regarding SA of parameters for ice-structure interaction analysis.

Figure 4.8 shows the PAWN indices of BR and SF are the largest, varying from 0.25 to 0.55. The PAWN indices of the remaining parameters are below 0.2. Hence, BR and
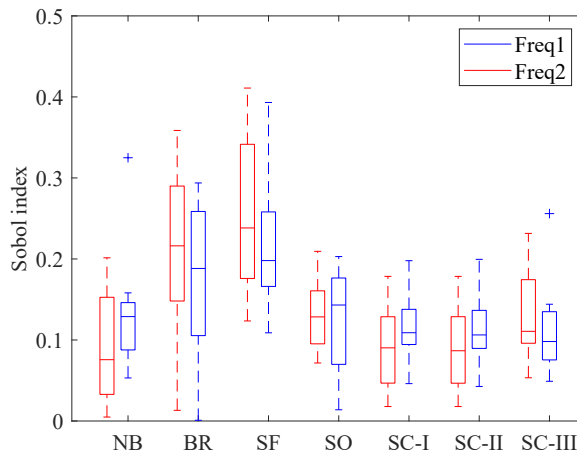
Figure 4.7: The statistics of Sobol indices of seven parameters for different ice velocities in data file 32010.
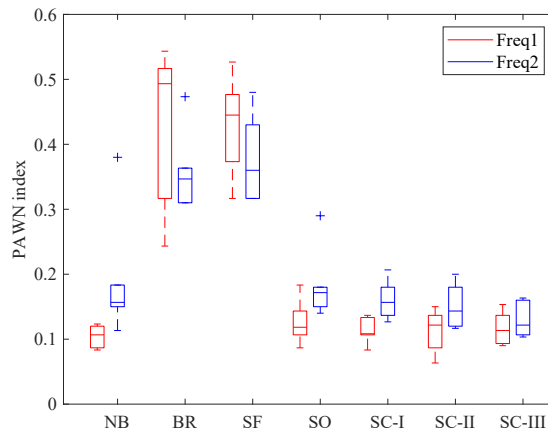
Figure 4.8: The statistics of PAWN indices of seven parameters for different ice velocities in data file 25010.
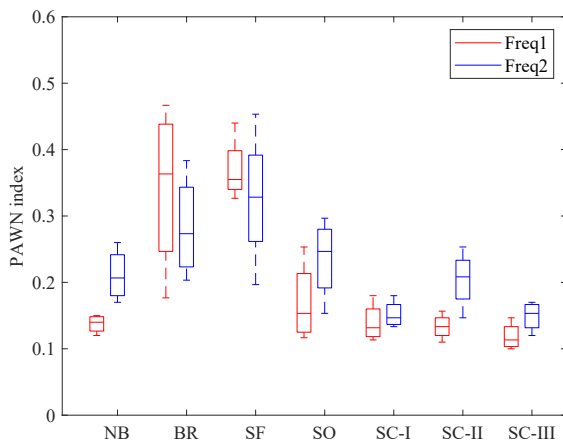
Figure 4.9: The statistics of PAWN indices of seven parameters for different ice velocities in the full-scale test.

SF are chosen as uncertain factors for UA on data file 25010.

Figure 4.9 shows the top two sensitive parameters are BR and SF whose PAWN indices are around 0.4 and 0.35 separately. Their indices are a bit smaller than the result in data files 32010 and 25010. This is possibly caused by the larger noise from the full-scale experiment in the real world.

To sum up, the most influential parameters are chosen as BR and SF for the following UA on data files 32010, 25010, and the full-scale experiment.

Uncertainty analysis

Table 4.3 lists the statistics of the identified first two natural frequencies. The statistics include mean value (Mean), standard deviation (Std), upper bound of 95% confidence interval (Upper), and lower bound of 95% confidence interval (Lower). The bold number means the correct identification. For the identified first natural frequency, the proposed method fails to identify it when ice failure is IC. Hence, it is not discussed in the following analysis. After the ice failure mode changes to FLI, the first natural frequencies are 19.84, 29.76, and 23.53 rad/s when ice velocities are 8, 10, and 20 mm/s, respectively. The frequencies in the remaining cases are around 21.5 which is close to the benchmark value. In contrast to the first natural frequency, the second frequency identification becomes more intractable. 5 out of 15 identified second frequencies are over 30. One is less than 29, around 26.66 rad/s. Other frequencies can be identified correctly.

Next, prior knowledge is used to pick up anomalies in the results. The benchmark values of the first two natural frequencies are 21.352 and 29.516 rad/s separately, which are estimated when the structure was moving in the open water [62]. Considering the difference between open water and ice, the benchmark values are expanded by 10% deviation to an interval: [19.22, 23.49] for the first frequency, [26.56, 32.47] for the second frequency. The processed UA results are shown in Figure 4.10 for data file

Table 4.3: The uncertainty analysis results of identified frequencies based on the data in '32010' under the ice failures of IC, FLI, and CC.

Frequency	Statistics	ice velocity (mm/s)														
		4	6	8	10	12	14	16	18	20	28	45	65	80	95	150
		IC	IC	FLI	FLI	FLI	FLI	FLI	FLI	FLI	FLI	FLI	CC	CC	CC	CC
First	Mean	45.59	17.57	19.84	29.76	**21.69**	**21.02**	**21.92**	**21.06**	23.53	**21.28**	**21.59**	**21.50**	**21.25**	**21.96**	**21.81**
	Std	28.04	5.68	1.64	2.41	0.01	1.57	4.55	0.01	3.39	0.42	0.02	0.25	1.22	2.10	1.23
	Upper	102.56	28.71	23.05	26.41	21.04	24.42	30.85	21.08	30.18	22.11	21.63	21.98	23.64	26.08	24.21
	Lower	-7.38	6.44	16.63	16.97	21.00	18.28	13.00	21.05	16.88	20.47	21.55	21.02	18.85	17.84	19.40
Second	Mean	88.00	**29.73**	26.66	32.00	32.49	**29.68**	32.43	**29.01**	30.38	**29.71**	**29.32**	**29.93**	**29.42**	**29.76**	**29.57**
	Std	28.84	14.48	3.44	10.59	3.88	3.62	9.07	1.60	5.12	0.66	1.82	1.92	2.43	12.02	4.47
	Upper	144.53	58.10	33.40	52.76	40.10	36.77	50.20	32.16	40.41	31.01	32.89	33.71	34.19	53.32	38.33
	Lower	31.47	1.35	19.91	11.24	24.88	22.58	14.65	25.87	20.35	28.42	25.74	26.16	24.65	6.19	20.80

32010, Figure 4.11 for data file 25010, and Figure 4.12 for the full-scale measurements.

Figure 4.10a shows the filtered first natural frequency and corresponding damping ratio. The left Y axis is the value of the identified natural frequency. The right Y axis is the corresponding damping ratio. The green dotted line represents the identified first natural frequency in [78]. Black points are the mean value of identified frequencies. The length of the vertical line at each point means the estimated 95% confidence interval. Based on the varying trend of the mean value, UA has a more sensible estimation than the method without UA, especially when ice velocities are larger than 45 mm/s. The superiority can also be verified in Figure 4.10b. Overall, the first natural frequency goes up with the increase in ice velocity. This trend is the same as the damping ratio as shown by the blue line. However, this trend does not appear on the identified second natural frequency as shown in Figure 4.10b. With the increase of ice velocity, the second natural frequency goes up until velocity reaches 16 mm/s. Next, it plummets when ice velocity changes to 18 mm/s. After that, it keeps an increasing trend from 20 mm/s. The damping ratio shows an overall decreasing trend when ice moves toward higher velocity. In addition, the uncertainty of the identified second frequency and damping ratio is significantly larger than that of the first frequency and corresponding damping ratio.

Figure 4.11a shows the significant increasing trend of the first natural frequency in data file 25010. The value rises from 19.3 to 21. For the second natural frequency, its values fluctuate vary drastically as shown in Figure 4.11b. The same trend can be found in Figure 4.12a and Figure 4.12b. In Figure 4.12a, the value of the first natural frequency starts from 18 to 22 as the increase of ice velocity. However, there is no obvious trend of the second natural frequency in Figure 4.12b. In addition, the value of the second natural frequency is far larger than that in the data files 32010 and 25010.

The mined knowledge is summarized as follows.

1. Compared with traditional SSI-cov, UA based on prior knowledge does have superiority regarding the accuracy of modal parameter identification. In addition, the
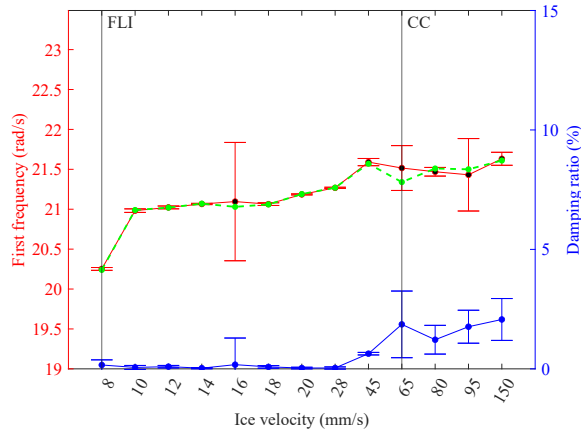
(a) The identified first frequencies and corresponding damping
ratios.
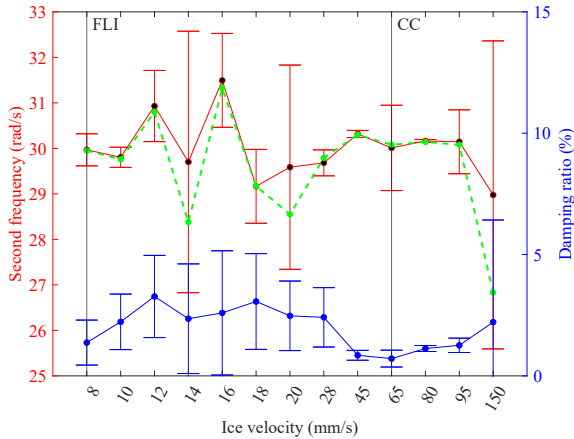
(b) The identified second frequencies and corresponding damp-
ing ratios.

Figure 4.10: The identified first frequencies and corresponding damping ratios over different
ice velocities in data file 32010 (Green dotted line: identified frequencies without UA; Black
points: the mean value of identified frequencies; Vertical lines: 95% confidence interval).

confidence interval could provide more valuable support for the SHM.

2. The identified first natural frequency and corresponding damping ratio present a
stable upward trend with the increase of ice velocity for three different experiments.
No difference between model ice and wave ice was observed in relation to that trend.
A small uncertainty interval demonstrates SSI-cov and Hierarchical clustering have
very good robustness for the first frequency identification.

(a) The identified first frequencies and corresponding damping
ratios.
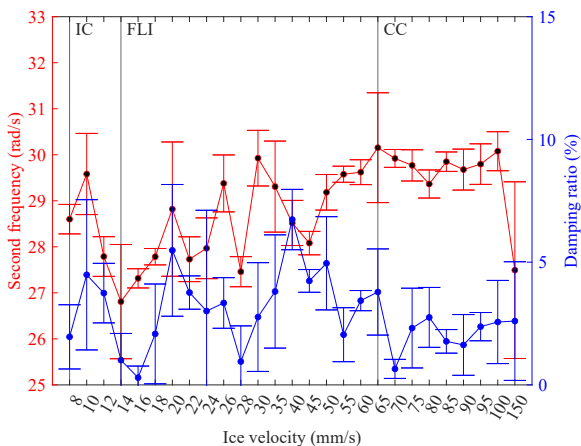
(b) The identified second frequencies and corresponding damp-
ing ratios.

Figure 4.11: The identified frequencies and corresponding damping ratios over different ice
velocities in data file 25010.

3. For the second frequency, its trend is not significant for all three experiments. Large
   uncertainty means that SSI-cov and Hierarchical clustering work worse in terms of
   the second frequency identification.

4. For data files 32010 and 25010, the damping ratio of the first mode is quite small,
   varying within [0% 3%]. However, it becomes larger in the full-scale experiment,
   varying within [5% 10%].
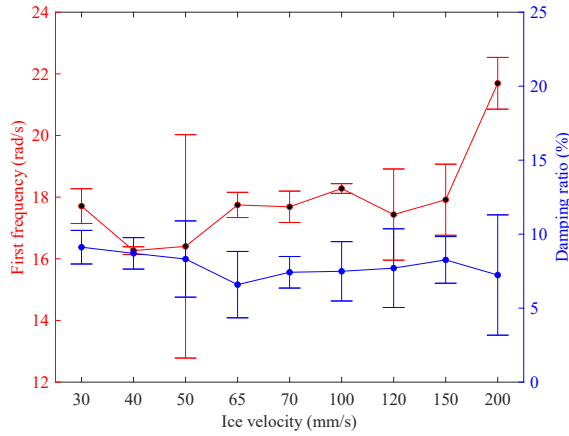
(a) The identified first frequencies and corresponding damping
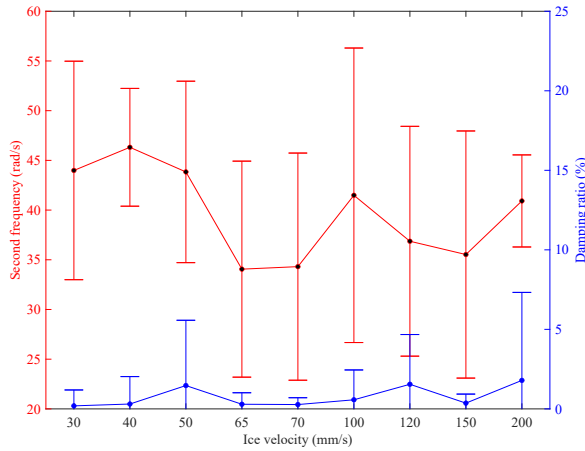
(b) The identified second frequencies and corresponding damping ratios.

Figure 4.12: The identified second frequencies and corresponding damping ratios over different ice velocities in the full-scale test.

4.3 Chapter summary

This chapter uses diesel engines on ships as an example. Two application scenarios are considered:

- Clustering is introduced to improve the efficiency of SSI-cov algorithm. Human interactions for the selection of slack values may cause a bias to the modal parameters' values. Therefore, the slack value-based method is replaced with Hierarchical

clustering to identify physical modes. The case study shows that Hierarchical clustering did achieve a better performance on modal parameters identification than slack values.

- Considering the uncertainties from parameters of SSI-cov and clustering technology, UA and SA -based framework is proposed to quantify the uncertainty of identified modal parameters. From Figure 4.6-4.9, PAWN can find out that 'BR' and 'SF' are the top two influential factors. Through the analysis on modal-scale tests and full-scale tests, results show that the proposed framework can identify the varying trend clearly in terms of the first natural frequency and its corresponding damping ratios as shown in Figure 4.10a.

Modal parameters identification is important to the SHM. The model presented in this chapter can be used to implement online modal parameters identification with high accuracy for onboard support of SHM.

5

Case study: AIS data analysis and modelling

AIS data contains abundant information, including the time-stamp for each recorded data point and the ship's MMSI, DWT, SOG, ship length. This information gives decision-makers extensive knowledge about a ship's behaviour and its surroundings. Furthermore, It can benefit for navigational safety of ships.

This chapter mainly introduces AIS data-related applications. They are COVID-19 impacts analysis, probabilistic ship route prediction, and short-term ship trajectory prediction as shown in Figure 5.1. COVID-19 impact analysis mainly focuses on the statistics of ship types, the number of ships, etc. The extract features are used as inputs to the next application. Probabilistic ship route prediction is to infer the possible route
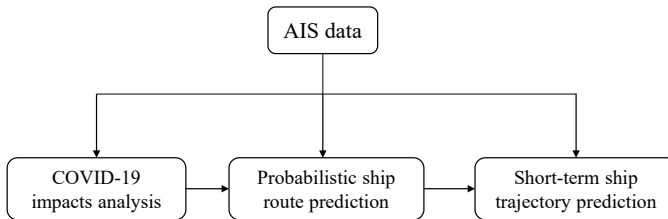
Figure 5.1: The relationship of AIS data-related applications.

5.1 COVID-19 impacts analysis

The advent of the COVID-19 pandemic disrupted global commercial activities and the tourism industry heavily [80]. Impacts on maritime transportation were huge, as seaborne trade represents over 80% of global merchandise trade [81]. Investigating how COVID-19 has affected ship behaviours is significant for economic condition evaluation, and port management. This section introduces how descriptive and diagnostic analytics are used for AIS data analysis to investigate how COVID-19 impacts the behaviour of different ships.

5.1.1 System structure of AIS data analysis

This subsection outlines the impact analysis of COVID-19 on ship behaviours and port operations. The whole analysis procedure is mainly divided into three parts, as shown in Figure 5.9. The first part is data preprocessing, including ship information matching.
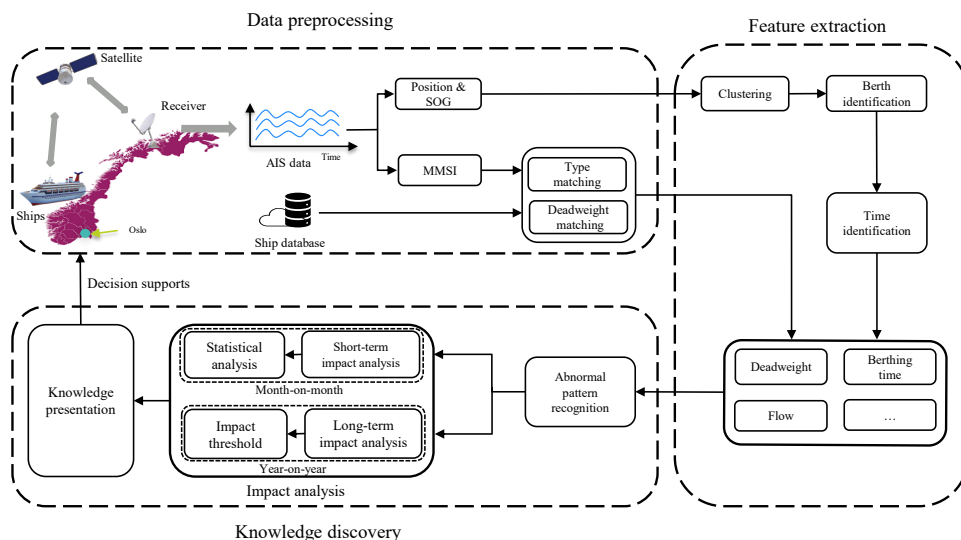
Figure 5.2: The workflow of impacts analysis of COVID-19 on ship behaviours based on AIS data.

The second one introduces how several useful features are extracted from raw AIS data. The knowledge discovery process is described in the last part. The analysis results could be used for the decision support of port management and preplanning in the next health emergency crisis.

Data preprocessing aims to match ship types and deadweight information based on MMSI. After that, clustering and statistical approaches are applied to obtain important features such as the berths and quays in Olso port and the dwelling time of each ship. Based on the extracted features, a similarity measuring method (DTW) is selected to carry out abnormal pattern recognition. It is combined with statistical analysis to analyze behaviour patterns of ships under COVID-19 effects.

The impact analysis is mainly composed of long-term analysis and short-term analysis. The long-term impact analysis aims to investigate the year-on-year difference caused by COVID-19 and derive an impact threshold for the following short-term analysis. The short-term impact analysis estimates the variation of month-on-month impact and explores the trend based on historical knowledge.

5.1.2 Experiment results

This experiment gives two case studies. One is long-term impact analysis, i.e. year-on-year analysis, in terms of ship flow, daily throughout, and average dwelling time of quays. Another is short-term impact analysis, i.e. month-on-month analysis, on the variation trend of the ship's flow.
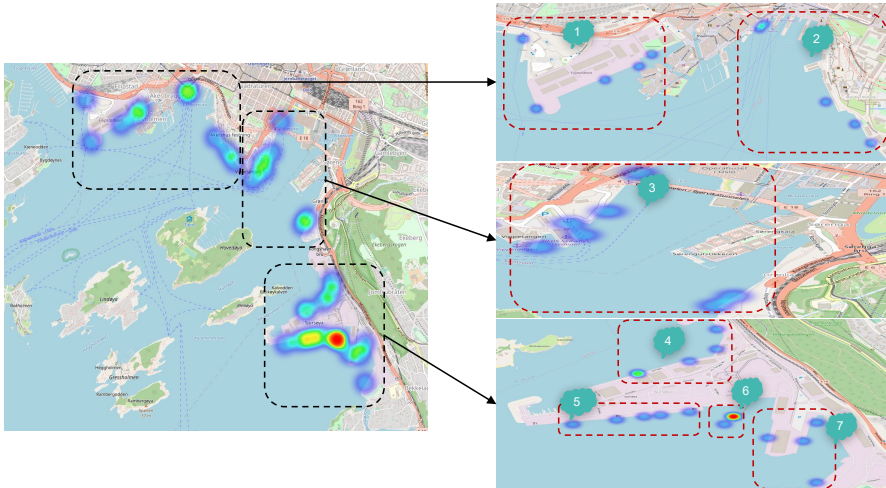
Figure 5.3: The distribution of berths in the Oslo port.

Table 5.1: DTW distance of ship flow, quay's daily throughout, and quay's average berthing time in different years.

Feature	Category	Y'17 VS Y'18	Y'17 VS Y'19	Y'18 VS Y'19	Average variation	□ VS Y'20		
						Y'17 (%)	Y'18 (%)	Y'19 (%)
Ship flow	General cargo	171.0	185.0	192.0	182.7	187.0 (2.4)	194.0 (6.2)	193.0 (5.7)
	Carrier	8.0	17.0	13.0	12.7	18.0 (42.7)	14.0 (10.0)	23.0 (**81.6**)
	Container	147.0	151.0	157.0	151.7	137.0 (-9.7)	149.0 (-1.8)	151.0 (-0.4)
	Tanker	54.0	53.0	54.0	53.7	49.0 (-8.7)	49.0 (-8.7)	44.0 (-18.0)
	Cruise	25.0	32.0	27.0	28.0	69.0 (**146.4**)	56.0 (**100.0**)	74.0 (**164.3**)
	Ferry	123.0	131.0	123.0	125.7	255.0 (**102.9**)	247.0 (**96.5**)	245.0 (**95.0**)
Throughout of quays	Quay 1	18.1	101.8	96.5	72.1	74.8 (3.7)	71.6 (-7.3)	46.7 (-35.3)
	Quay 3	3.1	4.6	4.8	4.3	10.3 (**146.0**)	11.7 (**179.4**)	10.7 (**155.4**)
	Quay 4	65.8	67.3	55.2	62.8	66.8 (6.4)	59.7 (-4.8)	57.8 (-8.0)
	Quay 5	220.4	216.8	216.2	217.8	200.6 (-7.9)	205.0 (-5.9)	209.6 (-3.8)
	Quay 6	139.7	117.4	135.1	130.7	121.8 (-6.8)	136.7 (4.6)	116.1 (-11.2)
	Quay 7	15.9	18.1	19.4	17.8	11.0 (-38.4)	17.9 (0.6)	17.3 (-2.7)
Berthing time of quays	Quay 1	8.5	11.6	12.4	10.8	11.9 (9.8)	13.3 (25.3)	12.9 (23.7)
	Quay 3	5.7	6.5	5.9	6.0	9.7 (**60.0**)	9.0 (**50.0**)	9.7 (**61.2**)
	Quay 4	33.0	32.2	29.3	31.5	35.9 (14.1)	33.6 (6.6)	30.5 (-3.1)
	Quay 5	27.3	25.8	25.9	26.3	23.1 (-12.1)	26.2 (-0.6)	26.3 (0.0)
	Quay 6	16.6	16.2	17.9	16.7	14.1 (-15.9)	15.1 (-9.6)	16.6 (-0.8)
	Quay 7	4.9	5.6	5.3	5.3	4.5 (-15.3)	4.7 (-12.0)	5.0 (-5.1)

Year-on-year impact analysis

DTW analysis results are shown in Table 5.1. The distribution of quays is shown in The units of DTW distance with respect to the ship's flow, daily throughout of quays, and berthing time is the number of times, 10000 tons, and 100 hours separately. 'Average variation' is the average value of the former three-year variations. '□ VS Y'20' denotes '2017 vs. 2020', '2018 vs. 2020', '2019 vs. 2020'. '(%)' denotes the increasing percentage of current value against the benchmark. For example, 2.4% represents the variation between 2017 and 2020 rising by 2.4% against the average variation. That also means the impact caused by COVID-19 is about 2.4%. On the other hand, -9.7% means the variation caused by COVID-19 is less than the average variation. In other words, COVID-19 does not influence its behaviours.

The mined knowledge is shown as follows:

- The drastic variation appears in the flow of cruises and ferries as a result of the big decrease of ship frequencies.

- Cargo ships are influenced slightly.

- Most of the port operations are not affected heavily.

- Quay 3 has a significant change with respect to daily throughout and berthing time before and during the pandemic.

The first point indicates that passenger ships are sensitive to the global pandemic. That brings revelations to policy-makers that service industries are supposed to gain more financial support. The second point implies port managers should guarantee the sufficient cargo handling capability of quays to speed up logistics. The third point tells us that few ships were stopping this berth during COVID-19. This is a rational outcome, possibly because quay 3 is close to ferry berths and at Oslo port center. Therefore cargo ships are moved to other cargo quays. There are some abnormal patterns that are not caused by COVID-19. For example, the variation of carrier's flow is quite large (81.6%) between 2019 and 2020.

A significance interval test is introduced to estimate the threshold of impacts caused by COVID-19 based on the analysis results above. Although these three features represent different physical meanings, throughout and dwelling time, in essence, results from the variation of ship flow. Hence, it is reasonable to use the average changing rate of impacts as a threshold that represents the severity of impacts. In this study, α is set as 0.05. As the p-value of the Shapiro test is 5.1×10^{-8} far less than 0.05, data are normally distributed. Therefore, the significance interval is set as [10.1, 39.7]. In this study, the upper significance interval is selected as a threshold. That means, if the impact caused by COVID-19 is larger than 39.7, it is regarded as a severe impact of which stakeholders should be cautious.

Month-on-month impact analysis

In order to analyze the short-term trend of ship's flow variation, first of all, January 2019 is selected as a referenced month. And then, the fluctuation of the ship's flow between January and February is counted as a normal vibration. In other words, it is a benchmark value that equals zero. If the variation is larger than zero, the ship

flow varies abnormally in the current time period. The window time is 15 days. Some are 14 or 16 days, depending on the month. The blue bar shows the increasing rate of change in the remaining months against the benchmark scatters in Figure 5.4. The boxes represent the percentile value of the ship's flow variation in the past three months. At the bottom of Figure 5.4, green bars mean that the current variation is less than the average variation of the past three months; orange bars indicate that it is a bit larger; red bars show that its variation is over 39.7, reaching a serious level. For example, due to COVID-19, the drastic variation takes place from the middle of March 2020. It starts to recover to the normal level from the middle of June.
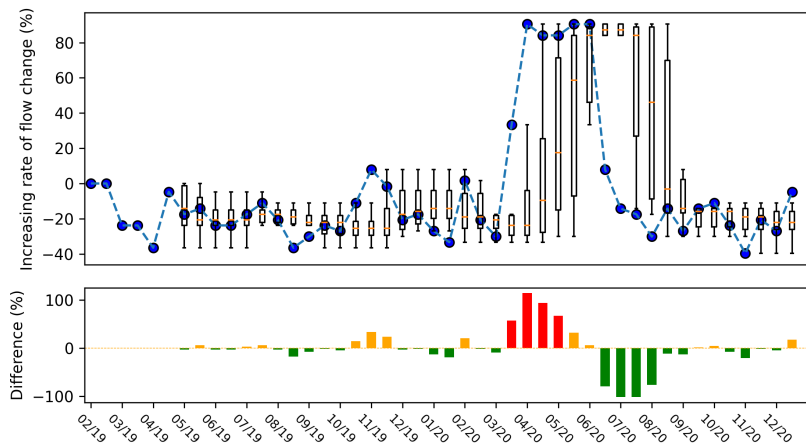
Figure 5.4: The short-term impact analysis on ferry's flow using DTW.

To sum up, the case study exhibits that the proposed method is able to find abnormal patterns caused by COVID-19 and quantify its impacts. The discovered knowledge could be used for port management and preplanning in the next emergency crisis. In addition, it is capable of providing support to stakeholders for short-term decisions through month-on-month impact analysis.

5.2 Probabilistic ship route prediction

Maritime transportation suffers from uncertainties caused by pilots' decisions and environmental factors. The changing environment and surroundings of ships would pose high risks to safe ship navigation. Due to rich information on ship behaviours in AIS data, it can be applied to estimate the possibility of a ship tracking certain routes, which can be further used to support decision-makers for collision avoidance.

5.2.1 System structure of route prediction

The system of online ship trajectory probabilistic prediction includes two parts as shown in Fig. 5.5. The first part is to cluster ship trajectories in a regional waterway. Three steps are demanded to cluster historical ship trajectories. The first step is to collect
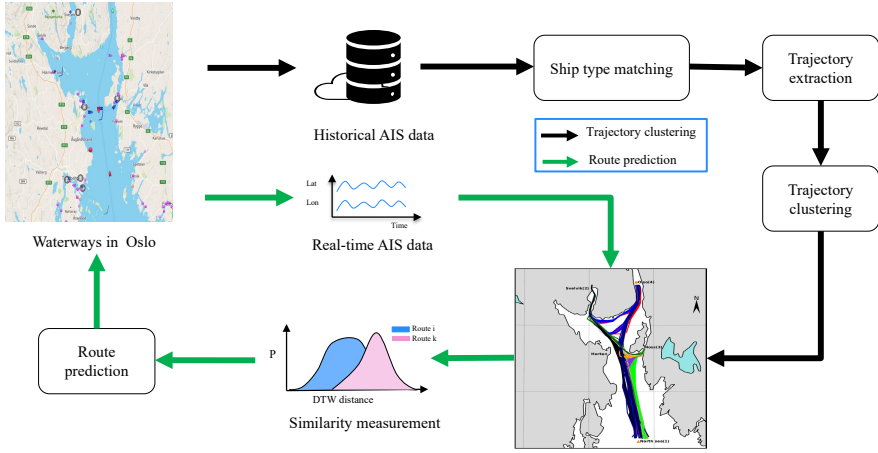
Figure 5.5: The system structure of online probabilistic route prediction of ships.

AIS data in this area and match ship types based on MMSI using the database in my previous paper [82]. The second step is to cut off ship trajectories based on ship behaviours such as port call and port departure. The last step is trajectory clustering to group ship trajectories according to the trajectory shape. The second part is real-time trajectory probabilistic prediction. AIS data of a ship is obtained with a specific window time. After that, this ship's trajectory is classified to the corresponding route based on the trajectory similarities. When a new ship is navigating in this area, its trajectories could be classified into one of the clustered trajectories. This is done by computing the DTW distance between the ship trajectory and clustered trajectories at the same time interval. Then, the probability distribution of DTW distance to each cluster can be obtained. Generally, the smaller the DTW distance, the larger probability this ship follows this route. Therefore, the minimum value in the distribution of DTW distance is used as the metric of trajectory similarity. Given the minimum DTW distance (DTW_i) of this ship's trajectory to route i, Eq. 5.1 is used to estimate the probability (P_i) of the ship following the route i separately.

$$\begin{cases} Sum = \sum_i^n \frac{1}{DTW_i} \\ P_i = \frac{1}{DTW_i \times Sum} \end{cases} \tag{5.1}$$

where n is the number of identified routes. Due to the varying ship behaviours, such trajectory prediction is carried out with a specific window time. At the same time, the probability is updated synchronously.

5.2.2 Experiment results

This section mainly introduces how to use historical AIS data to cluster ship routes and how to predict ship routes probabilistically.
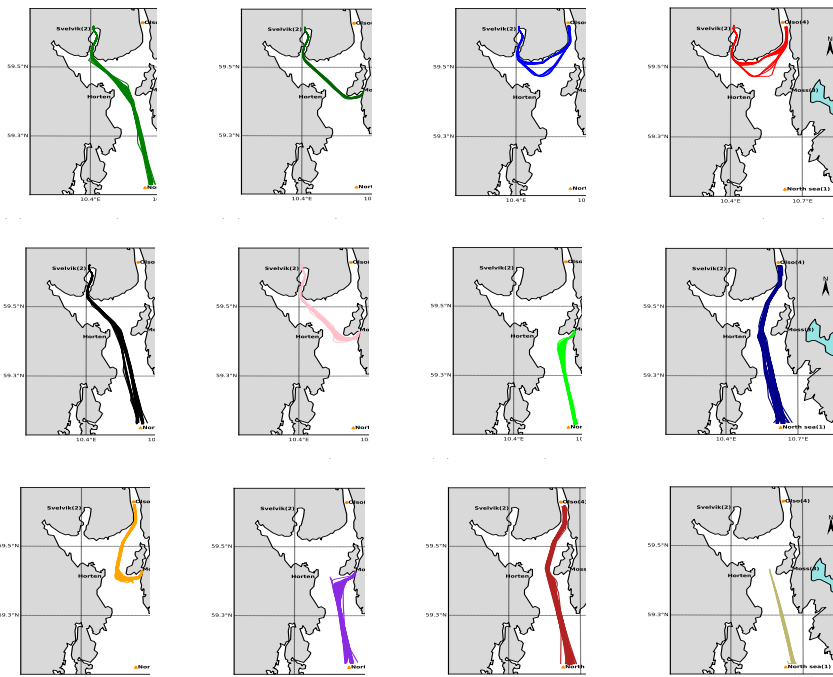
(i) Route9 (4 to 3).   (j) Route10 (3 to 1).   (k) Route11 (1 to 4).   (l) Route12 (1 to 4).

Figure 5.6: The identified ship routes by HDBSCAN.

Identified ship routes

Figure 5.6 shows the clustered trajectories by HDBSCAN. In this figure, the orange circle refers to the direction of Olso which is labelled as 4; the orange triangle is located at the north sea which is labelled 1. The city Svelcik and Moss are labelled 2 and 3 separately. In these figures, 12 routes are found from different directions. The correspondingly identified trajectories are shown in Figure 5.6a - Figure 5.6l. In each subfigure, for example in Figure 5.6a, '1 to 2' means the ship is calling the port of Svelvik from the North sea. Therein, an abnormal route is found in Figure 5.6l. This possibly results from the process of trajectory extraction. When some ships are stopping for avoiding collisions, the trajectories of these ships are cut off into two segments. We can remove this route manually or not consider this route for the ship route classification in the next section.

Probabilistic route prediction

In the case study of ship route prediction, there are four ships sailing in this area as shown in Figure 5.7. These trajectories are extracted from the AIS data on June 7th, 2020. In the legend of this figure, Ship1$(1->3)$ represents ship1 moved from the direction of 1 to 3.

These six figures show the four ships' trajectories at different time moments. The
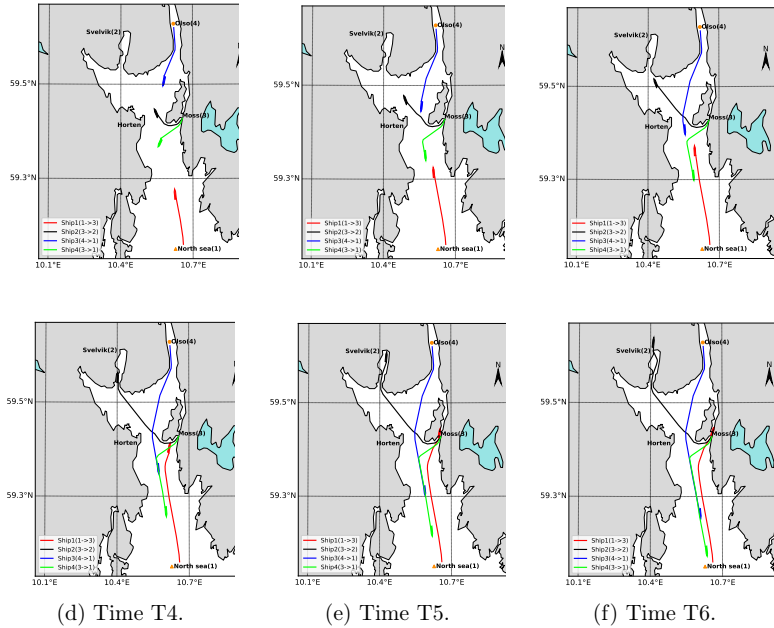
Figure 5.7: The four ships' trajectories at different time moments.

time T1-T9 is a relative time. The 10-minute timestamp is defined as 0. Every 10 minutes is regarded as a window time. That means the time interval between T1 and T2 is 10 minutes. At time T6, ship1 and ship2 arrive at their destination whereas the remaining two ships are still sailing toward their destination.

For ship1, at six time moments, the DTW distance between its trajectory and routes is shown in Table 5.2. Note that all values were multiplied by 10^3 in Table 5.2. The probability of ship1 tracking Route1, route7, and route11 is shown in Figure 5.8a. The estimated probability of different routes has a very large variation over different time moments. Ship1 follows route11 with a relatively large probability of 47.34% at the beginning. Route1 ranks second and route7 has the smallest probability. As time goes on, the probability of route7 becomes greater than the others from T4 and keeps arsing up to 96.49% at T6. On the contrary, the probability of route11 and route1 decreases all the way to 1.70% and 1.81%, separately. The predicted probability of routes is identical to our observation of ship1 movement in Figure 5.7a - Figure 5.7f.

For ship3, the DTW distance between its trajectory and routes is shown in Table 5.2. The corresponding probability estimation is shown in Figure 5.8b. At T1 and T2, Ship4 tracks Route9 with the largest possibility over 40%. On the contrary, route3 has the smallest possibility, which means Ship4 is unlikely to track route3 at the beginning. From T3, route8's possibility goes up quickly and starts to be larger than route9's. As time goes on, route8 accounts for the dominant possibility which is over 80% from T4. As shown in Figure 5.7d, we can observe that Ship4 is moving toward the direction of the North Sea. In other words, it is tracking route8 from now on.

Table 5.2: The DTW distance between ship's trajectory and identified routes.

Ship	Course	T1	T2	T3	T4	T5	T6
Ship1	Course1	2.78	2.89	3.28	14.59	49.67	78.03
	Course7	5.27	3.05	4.93	5.74	8.03	1.37
	Course11	2.03	2.27	2.72	11.80	34.28	73.10
Ship2	Course2	3.06	7.28	2.66	2.37	5.01	1.71
	Course10	23.55	60.51	126.63	195.16	219.57	275.56
Ship3	Course3	2.82	19.55	49.4	111.97	210.36	283.13
	Course8	1.73	2.83	3.58	3.74	3.66	2.93
	Course9	1.66	1.93	4.99	16.71	45.43	95.56
Ship4	Course2	29.77	68.05	84.05	128.70	233.42	345.34
	Course10	1.83	2.09	1.07	1.00	1.93	1.84
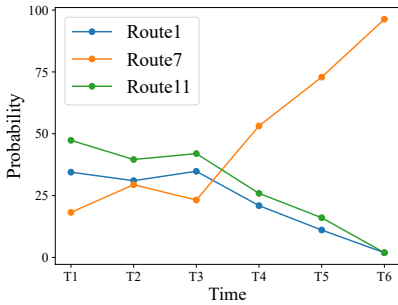
(a) Ship1
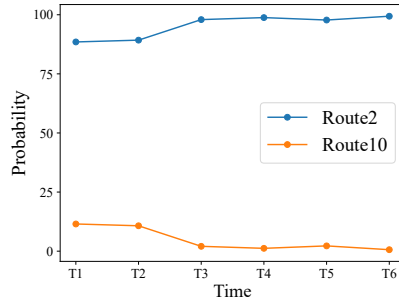
(b) Ship2
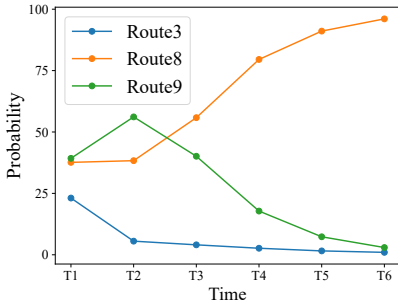
(c) Ship3
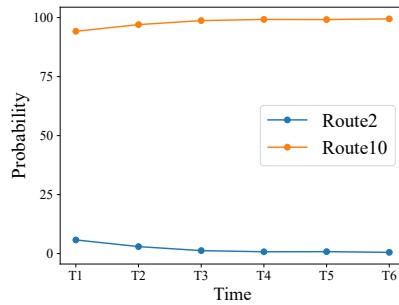
(d) Ship4

Figure 5.8: The probabilistic route prediction of four ships at six time moments.

For ship2 and ship4, the DTW distance between them and their routes presents an inverse varying trend in Table.5.2. Figure 5.8b and Figure 5.8d show that ship2 tracks route2 whereas ship4 tracks Route10 from beginning to end.
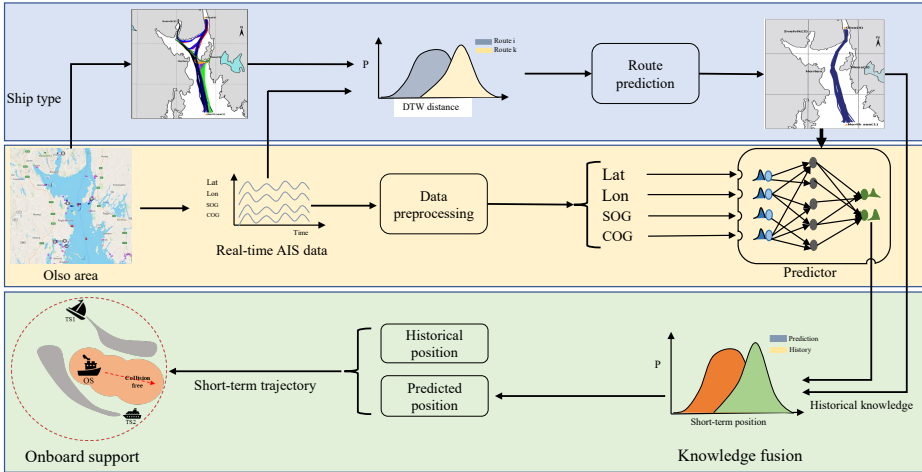
Figure 5.9: The knowledge fusion framework for the uncertain ship motion prediction.

Through real-time ship trajectory prediction, it can give decision-makers support in terms of the inference of the future trajectory. It can benefit for them to make ship path planning for collision avoidance.

5.3 Short-term ship trajectory prediction

AIS data have been widely used for ship trajectory prediction. The prediction accuracy is prone to the impact of AIS data quality. On the hand, noises and data missing problems exist in the AIS data. On the other hand, many uncertainties could be introduced during data pre-processing, such as interpolation. Therefore, this paper proposes a hybrid model to quantify the ship position uncertainty by integrating historical trajectory information and online prediction of the GP model.

5.3.1 System structure of ship trajectory prediction

This section outlines how to quantify uncertainties of ship motion-based knowledge fusion of online position prediction and historical information. The whole analysis procedure is mainly composed of three parts, as shown in Figure 5.9. The first part is to identify the route that this ship is tracking, which can be used to extract position behavior from the identified route as prior knowledge. The second one constructs a data-driven model which is used to predict future positions online based on current ship states. The last part is to make predictions by knowledge fusion from historical information and real-time position prediction.

After route identification, the historical trajectory information could be extracted based on the current ship position. It is represented by a Gaussian distribution. Meanwhile, the constructed GP model can predict the future position based on the current ship states. The predicted position is also denoted by a Gaussian distribution. Next, Gaussian fusion is to fuse these two Gaussian distributions. Assuming $Lat_{hist} \sim \mathcal{N}(\mu_1, \sigma_1^2)$
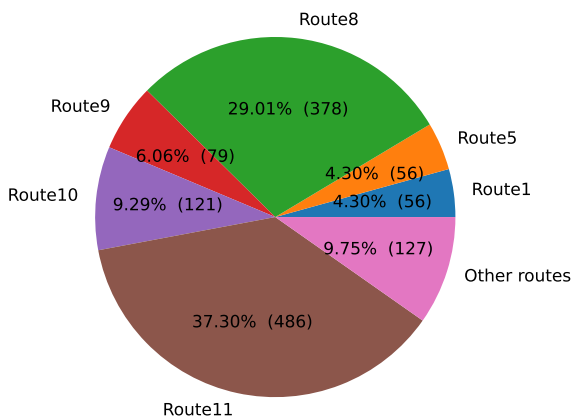
Figure 5.10: The number of trajectories in different routes.

and $Lat_{online} \sim \mathcal{N}(\mu_2, \sigma_2^2)$, Gaussian fusion is given by

$$\mathcal{N}(x, \mu_1, \sigma_1^2) \otimes \mathcal{N}(x, \mu_2, \sigma_2^2) =$$
$$S \cdot \frac{1}{\sqrt{2\pi}\sigma'} \cdot e^{-\frac{(x-\mu')^2}{2\sigma'^2}} \tag{5.2}$$

where

$$\begin{cases} S = \frac{1}{\sqrt{2\pi(\sigma_1^2+\sigma_2^2)}} \cdot e^{-\frac{(\mu_1-\mu_2)^2}{2(\sigma_1^2+\sigma_2^2)}} \\ \mu' = \frac{\sigma_2^2\mu_1+\sigma_1^2\mu_2}{\sigma_1^2+\sigma_2^2} \\ \sigma' = \sqrt{\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}} \end{cases} \tag{5.3}$$

Via Gaussian fusion, the new mean value (μ') and standard variance (σ') of Lat can be obtained. Likewise, the new mean value and standard variance of Lon can also be derived.

5.3.2 Experiment results

As shown in Figure 5.10, route11 contains 486 trajectories, accounting for 37.30. It is followed by route8 around 378 trajectories. The number of other remaining 5 routes amounts to 127. Therein, route11 has a similar behavior, apart from trajectory direction to route8. Therefore, only route11 is chosen for the following experiment. Likewise, route9, route10, and route5 are selected while others are not considered due to the limited number of trajectories. The training set accounts for 80% while the test set is 20%. For GP modelling, the variational distribution is the Cholesky distribution and the inducing points of variational inference are (3,16,36). The epochs of Pytorch training are 50 at which the training loss converges to the smallest value. LSTM has 6 layers network with 128 hidden neurons. The code was implemented based on Pytorch and

Table 5.3: The RMSE of the whole trajectory prediction on four routes

Route	Methods	1min (m)	2min (m)	3min (m)	6min (m)
11	GP	73	84	100	158
	LSTM	89	104	121	173
	Kinermatic	469	915	4310	6815
	Hybrid	**61**	**75**	**93**	**148**
5	GP	137	137	154	227
	LSTM	82	91	116	208
	Kinermatic	492	950	4270	6728
	Hybrid	**60**	**78**	**99**	**162**
9	GP	49	65	105	196
	LSTM	47	55	73	150
	Kinermatic	216	419	1746	2691
	Hybrid	**47**	**52**	**74**	**112**
10	GP	94	146	188	297
	LSTM	84	99	132	225
	Kinermatic	483	944	4313	6837
	Hybrid	**57**	**85**	**115**	**195**

Gpytorch [83]. Ship trajectory contains different ship behaviors such as turning and straight sailing. For straight movement, the model has relatively higher accuracy than the turning maneuvers. Hence, two types of error metrics are chosen in this study. One uses the root mean square error (RMSE) of distance for a whole trajectory.

Quantitative analysis

Table 5.3 displays the RMSE values between the reference and the predicted positions at the time horizon of 1 min, 2 min, 3 min, and 6 min using the entire trajectory for evaluation. The trajectory includes both turning and straight sections, where 'turning' denotes a ship-turning segment. We compare the RMSE of different models including GP, LSTM, Kinermatic model, and the proposed hybrid model.

The results in Table 5.3 show that the hybrid model achieves consistently low RMSE values for 1-minute position prediction, with values less than 65 m. As the time horizon increases to 2 and 3 minutes, the RMSE values of the hybrid model increase, but remain smaller than those of other models, peaking at 115 m. These values are approximately 20 m smaller than those of LSTM. For a 6-minute prediction, the hybrid model's RMSEs range from 112 m to 195 m. The difference in RMSE between the hybrid model and LSTM grows larger at around 40 m. Overall, the hybrid model outperforms other models.

Table 5.4 shows the RMSE of models for the turning trajectories prediction in different routes. Compared with the whole trajectory prediction, the RMSE values of GP, LSTM, and hybrid model increase overall for turning trajectory prediction, especially

Table 5.4: The RMSE of the turning trajectory prediction on four routes

Route	Methods	1min (m)	2min (m)	3min (m)	6min (m)
	GP	80	98	119	183
11	LSTM	106	123	138	208
	Hybrid	**47**	**70**	**95**	**163**
	GP	133	149	180	279
5	LSTM	107	136	170	283
	Hybrid	**66**	**93**	**125**	**207**
	GP	83	136	212	432
9	LSTM	82	128	201	503
	Hybrid	**54**	**108**	**173**	**328**
	GP	122	183	237	325
10	LSTM	171	200	236	394
	Hybrid	**105**	**124**	**169**	**274**

for 6-min position prediction. When the predictor makes a turning trajectory prediction for route9, the RMSE of the hybrid model is 200m larger than that of the whole trajectory prediction. For other routes prediction, the difference is not obvious due to the small turning angle. From this table, it can be seen that the hybrid model shows an obvious advantage and its RMSE is over 100m smaller than the LSTM for route9 and route10. Therefore, the hybrid model did have a better performance on ship trajectory prediction, particularly for the ship-turning trajectory prediction.

Qualitative analysis

Figure 5.11 shows the predicted trajectory in route9 at different time moments, in which the green point means the current ship position, the black points are the reference positions at 1 min, 2 min, 3 min, and 6 min; the blue dashed line represents the predicted trajectory by GP; the yellow line is the history information; the red line is the predicted trajectory by the hybrid model; the blue shaded area is the uncertainty area within 95% confidence interval of the GP prediction while the red area is the uncertainty area of the hybrid model prediction.

In Figure 5.11a, the red line coincides with the referenced trajectory. In addition, its uncertainty area is smaller than the blue area. As the predicted time horizon becomes longer, the offset between the predicted trajectory and the reference goes larger and its uncertainty area turns to bigger as well. From this figure, the uncertainty area obtained by the hybrid model is smaller than GP's. Even in the worst case, its uncertainty is equivalent to the GP's.

Based on the above analysis, the pure data-driven GP model can not predict ship-turning manoeuvers perfectly. This does make sense because the turning data only account for a small part of the training dataset. Therefore, historical trajectory information plays a key role to calibrate the GP's prediction to render a more accurate prediction.

(a) T1

(b) T2

(c) T3

(d) T4
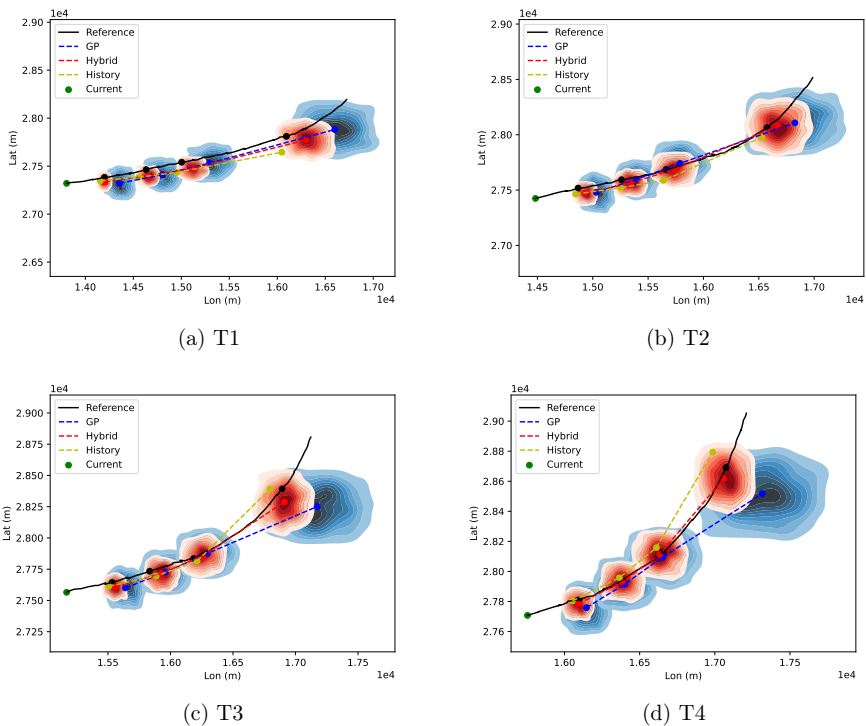
Figure 5.11: The turning trajectory prediction for route9.

5.3.3   Online ship trajectory prediction

A ship is sailing from Olso to the north sea. The real-time ship trajectory prediction is shown in Figure 5.12 and Figure 5.12. The left subfigures are the probabilistic route prediction at six time moments. The dotted lines are the possible routes of this ship will track in the following time. The pie chart at the left bottom of the subfigure shows the probability of each route this ship is following. This probability changes over time. For example, in Figure 5.12a, a ship tracking route8 has the largest probability around 51.9% at T1. In Figure 5.12c, the probability of route9 goes up from 30.2% to 64.0%. As time goes on, its probability keeps increasing up to 100% in Figure 5.12k.

The right subfigures are the ship trajectory prediction at different time moments. Route prediction aims to identify which predictor is supposed to be used for ship position prediction. For example, a ship following route8 has a large probability at T1 and consequently, a predictor trained by trajectories in route8 is used to make trajectory prediction. The corresponding predicted trajectory is shown in Figure 5.12b. Therein, the solid black line is the past trajectory of this ship and the dotted line represents the future trajectory that is taken as the ground truth. The red dotted line is the predicted trajectory by the proposed hybrid model and the red-shaded areas are the estimated
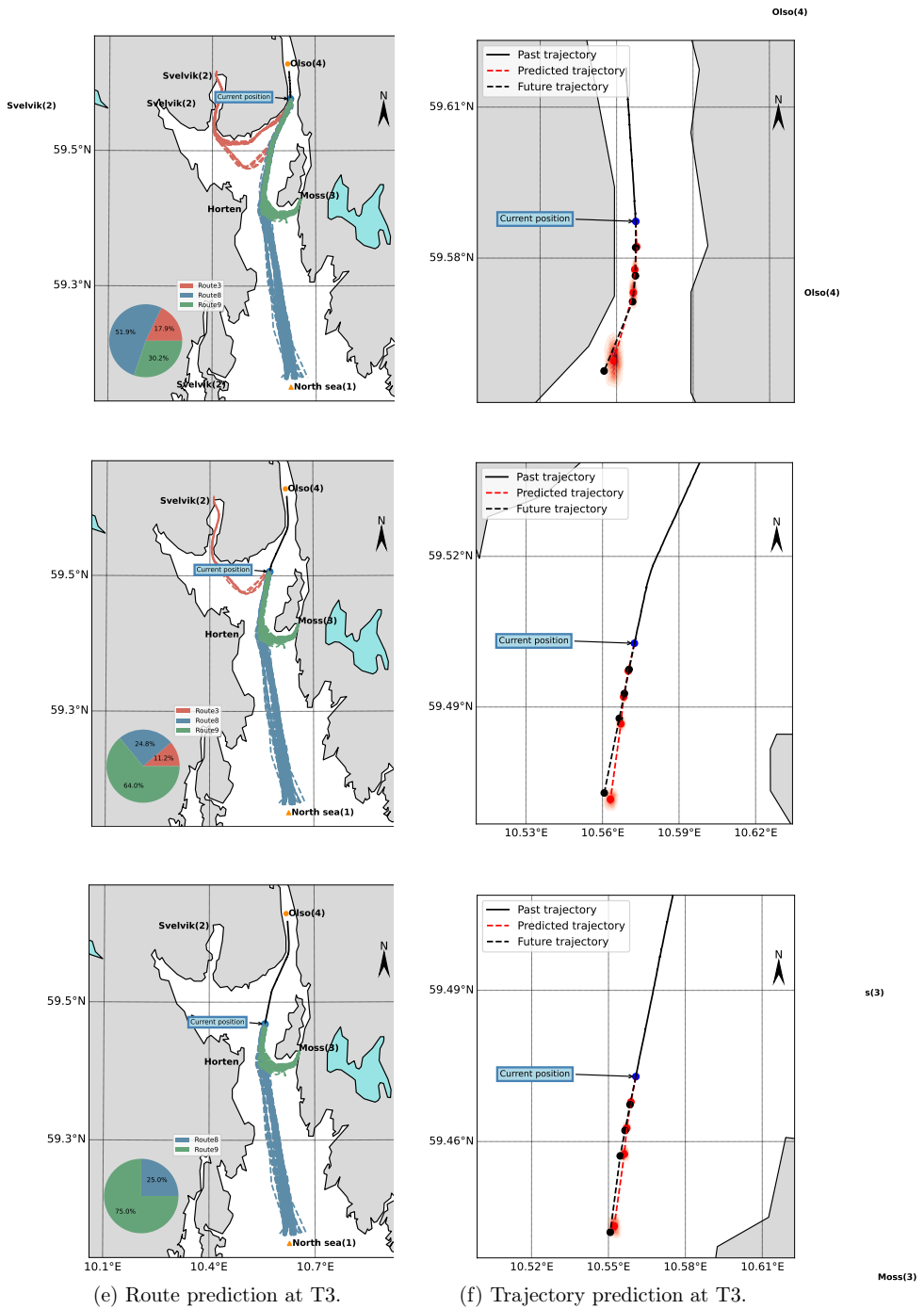
(e) Route prediction at T3.

(f) Trajectory prediction at T3.

Figure 5.12: The real-time ship trajectory prediction.

(k) Route prediction at T6.

(l) Trajectory prediction at T6.
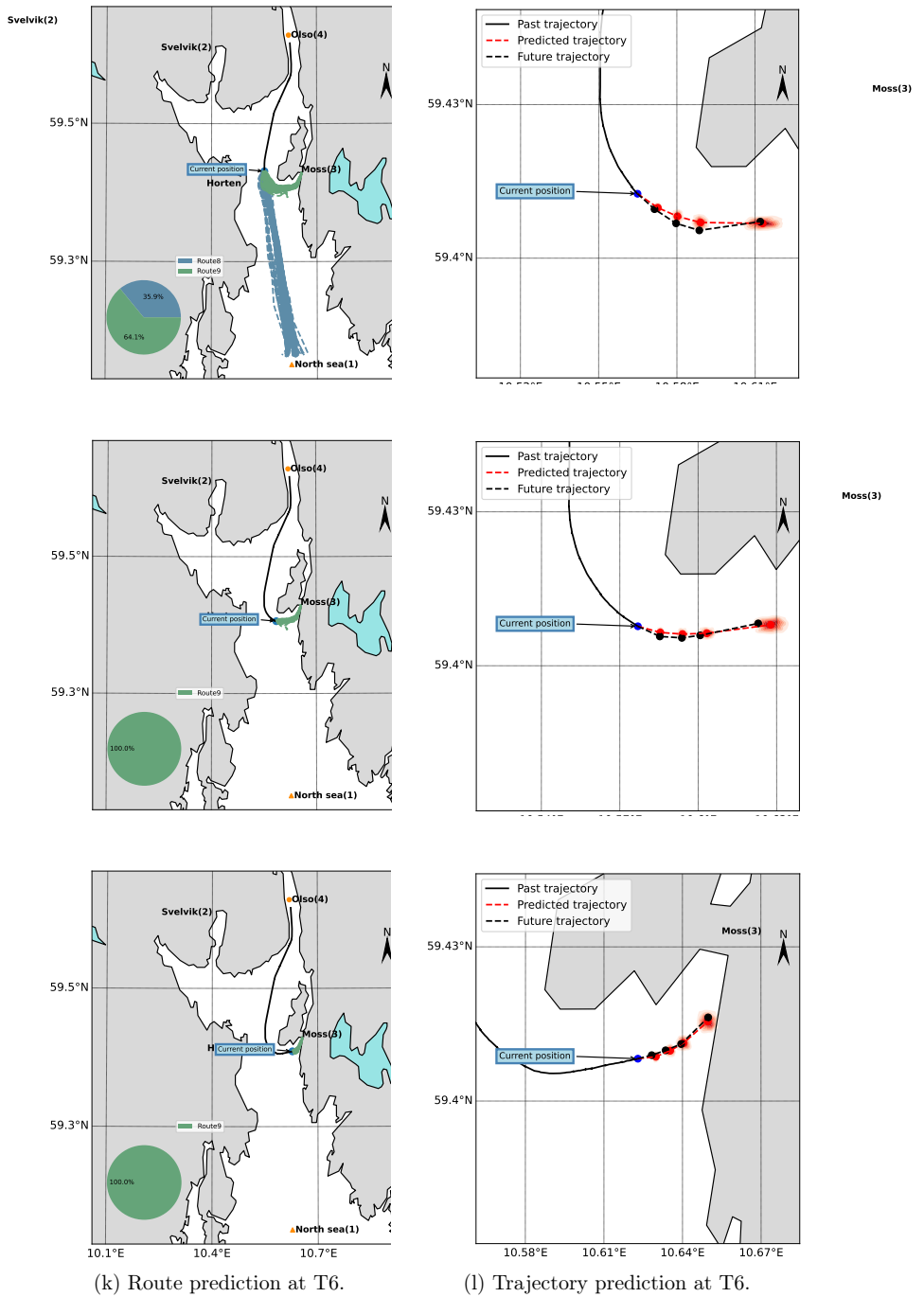
Figure 5.12: The real-time ship trajectory prediction.

uncertainty at different time horizons. The blue point is the current position of this ship. When the largest probability changes to route9 in the following time, the trajectory is predicted by the predictor trained by trajectories in route9. From this figure, it can be observed that the predicted trajectory is identical to the ground truth, which means the proposed hybrid model did have a better performance on the short-term ship trajectory prediction.

5.4   Chapter summary

This chapter mainly introduces AIS data analysis and modelling. Three cases studies are given as follows:

- Descriptive and diagnostic analytics are applied to analyze the interaction between COVID-19 and ship behaviours by AIS data analysis. The important features are extracted by statistical analysis methods and then used for short-term and long-term impact analysis via DTW. The results show that the proposed method can identify the abnormal ship behaviour influenced by COVID-19. They can also be used for the support of economic policy-making and port management if extreme quarantine measures are required in the next health crisis.

- To ensure safe ship navigation, descriptive and diagnostic analytics are applied to cluster ship trajectories using one-year AIS data of container ships navigating in a regional area. The result shows that the proposed method can identify routes correctly. Based on the clustered trajectories, DTW is used to estimate the possibility of a ship tracking certain routes. The analysis result could be used to support decision-makers for collision avoidance.

- Ship motion is prone to the impacts of various uncertain factors. A hybrid model is proposed for UQ of the predicted ship positions via the knowledge fusion of historical trajectory information and online prediction of the Gaussian Process model. Through experiment comparison, the proposed method achieves a better performance on trajectory prediction. In addition, an uncertainty area is given to describe the probabilistic ship movement.

   AIS data contains information-rich ship static and dynamic states. Mining valuable knowledge from AIS data benefits onboard support of various marine operations.

6

Conclusion and future work

This dissertation explores the application of data analysis in providing onboard support for demanding marine operations. It presents research findings based on three case studies of marine data mining for situation awareness and decision support. The main objective of this work is to provide support for various marine operations through data analysis and modeling. To achieve this, various analysis technologies were employed to analyze marine data. A major contribution of this study is the use of prescriptive analytics, SA and UA, to improve the reliability of predictive models and reduce decision-making risks. In summary, this dissertation highlights the potential of data analysis for improving marine operations, and its findings can serve as a valuable resource for researchers and practitioners in the field.

6.1 Summary of contributions

The rich historical information contained in marine data can aid decision-makers in taking timely actions to fit future development and challenges, as stated in RO1. Descriptive and diagnostic analytics are effective tools for extracting potential patterns and knowledge from large marine datasets. For example, statistical analysis and clustering methods can be used to discover potential patterns in the AIS data and structural response data. However, they can not answer the question of what is likely to happen in the future, leading to the need for RO2. ML and DL are used to create a surrogate model between system inputs and outputs to carry out, for example, ship position prediction. However, data-driven models are susceptible to uncertainties and are black-box and not explainable, leading to RO3, RO4, and RO5. To identify the most influential factors to expected outcomes, SA is introduced to improve the model's reliability and aid operators in making well-informed decisions in RO3 through a case study of thrusters' significance analysis. For the problem of UQ, RO4 developed a probabilistic model to quantify the uncertainty of observations of interest such as ship positions. RO5 aims to use SA to reduce the computational cost of modal parameters UQ, which can implement online support for structural health monitoring.

The main contributions of this dissertation are as follows:

- Descriptive analytics and diagnostic analytics are utilized in marine data analysis to uncover potential patterns and support decision-making. Two case studies demonstrate the effectiveness of these methods.

- ML technologies are employed to construct a surrogate model from marine data, which enables SA to analyze the importance of each factor to a specific operation.

- A hybrid model is proposed to enhance the confidence level of predicted results

through UQ for online decision support.

- A framework based on UA and SA-based framework is proposed to implement rapid UQ of observations of interest for onboard support of SHM.

6.2 Summary of publications

The summary of publications is as follows:

Paper I combines SA with SVM to analyze thrusters' importance to the ship's DP capability, taking into account the influence of thruster failures and environmental factors. SVM is used to construct a surrogate model between thrusters and ship DP capability while SA is used to analyze the sensitivity of thrusters. The results of the analysis can be used to support decision-making for improving the ship's DP capability on board. The proposed method is validated through simulations, and the results demonstrate that it can accurately identify the sensitivity of thrusters.

Paper II uses a clustering method and DTW to analyze the impact of COVID-19 on ship behaviours in the port areas using AIS data. Clustering is to identify quays in Oslo port and DTW is to quantify the impact of COVID-19. Through long-term impact analysis, the mined knowledge can be applied to port management if extreme measures are required in the next health crisis.

Paper III applies the clustering method to identify the physical modes of offshore structures. After SSI-cov, stable poles can be obtained by a stable criterion. A clustering method is then used to cluster poles to render physical modes. The proposed method is able to reduce the many user intervenes and enables efficient automatic parameter identification. The results show that clustering can render more successful identifications than the slack value-based method among different ice speeds.

Paper IV presents a UA framework based on SA to implement online UQ of modal parameters. SA is applied to pick up the input variables that contribute the most to the identified modal parameters. Next, the important variables are left free to vary over their range of existence to obtain the modal parameters' uncertainties. The framework is applied to full-scale tests from Norströmsgrund lighthouse and model tests carried out in HSVA. The comparison with traditional automated modal parameter identification shows its superiority in terms of the accuracy of modal parameters.

Paper V combines clustering with DTW to make a probabilistic ship route prediction. Clustering is utilized to group ship trajectories to different routes. Next, DTW is used to compare the similarity between trajectories and the identified routes to estimate the possible route that a ship is following. The proposed method is demonstrated through several case studies, showcasing its effectiveness for online probabilistic route prediction. The analysis results can support decision-makers in making collision avoidance decisions.

Paper VI extends the work presented in Paper V by using AIS data to construct a GP model for ship trajectory prediction. A hybrid model is then developed by combining historical trajectory information with the GP model to improve prediction accuracy. The proposed hybrid model outperforms the GP model alone, achieving more accurate and less uncertain ship trajectory predictions.

6.3 Future work

This thesis has mainly focused on three aspects of onboard support of marine operations: thrusters' importance analysis, structural health monitoring, and AIS data analysis and modelling. Different kinds of analysis approaches are combined to meet the requirement of different data types and applications. The main focus revolved around UA and SA. The below bullet points provide suggestions for how the presented research may be extended.

- To analyze the importance of thrusters under various influences, DP operations were simulated on the Zeus simulator with different scenarios of thruster failures and environmental factors. The simulation provided the flexibility to change input parameters conveniently with human interventions, such as modifying the wind's direction and speed to predefined values. However, the simulation data has limitations in replicating real-world scenarios. Therefore, in future work, the proposed method will be applied to real DP operations to validate the results obtained from the simulations.

- A hybrid model has been proposed to quantify the uncertainty of ship movements by integrating historical knowledge with the predicted position obtained from a GP model. The training dataset was obtained from one year of AIS data for container ships. However, for some routes, the dataset only contained a few trajectories, which is not sufficient to train a GP model accurately. Therefore, additional trajectories need to be extracted to construct a more precise model for ship trajectory prediction. Furthermore, a marine traffic monitoring system will be developed by integrating the knowledge discovered from AIS data and the constructed predictors. This system is expected to provide online predictions of ship positions and early warnings for collision avoidance.

- The black-box nature of data-driven models may increase the risk of overconfidence in their predictions and limit their application in certain engineering fields. SA is an efficient tool for understanding the learned input-output relationship and can help open up this black box by investigating the impacts of parameters on outputs. SA has been widely used for improving various neural networks, such as the proposed bilateral SA method that measures the relationship between neurons and layers to improve deep neural networks [84], and the use of SA to optimize the internal structure of convolutional neural networks by reducing inputs in its first layer [85]. In future work, SA will be applied to improve data-driven models for onboard support of marine operations.

- Causality analysis involves determining the cause-and-effect relationships between different factors and understanding how changes in one variable can lead to changes in another. The combination of SA and causality analysis can tell decision-makers what causes it to happen and how much the influential factors contribute to marine operations. Future research might explore it.

References

[1] L. Guidi, A. F. Guerra, C. Canchaya, E. Curry, F. Foglini, J. O. Irisson, K. Malde, C. T. Marshall, M. Obst, R. P. Ribeiro, *et al.*, "Big data in marine science," 2020.

[2] H. Zhang, G. Li, L. I. Hatledal, Y. Chu, A. L. Ellefsen, P. Han, P. Major, R. Skulstad, T. Wang, and H. P. Hildre, "A digital twin of the research vessel gunnerus for lifecycle services: Outlining key technologies," *IEEE Robotics Automation Magazine*, pp. 2–15, 2022.

[3] A. Coraddu, L. Oneto, F. Baldi, F. Cipollini, M. Atlar, and S. Savio, "Data-driven ship digital twin for estimating the speed loss caused by the marine fouling," *Ocean Engineering*, vol. 186, p. 106063, 2019.

[4] B. Schleich, N. Anwer, L. Mathieu, and S. Wartzack, "Shaping the digital twin for design and production engineering," *CIRP annals*, vol. 66, no. 1, pp. 141–144, 2017.

[5] G. Li, H. Zhang, B. Kawan, H. Wang, O. L. Osen, and A. Styve, "Analysis and modeling of sensor data for ship motion prediction," in *Oceans*, pp. 1–7, 2016.

[6] H. Jiang, S. Duan, L. Huang, Y. Han, H. Yang, and Q. Ma, "Scale effects in ar model real-time ship motion prediction," *Ocean Engineering*, vol. 203, p. 107202, 2020.

[7] G. Li, H. P. Hildre, and H. Zhang, "Toward time-optimal trajectory planning for autonomous ship maneuvering in close-range encounters," *IEEE Journal of Oceanic Engineering*, vol. 45, no. 4, pp. 1219–1234, 2019.

[8] 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*, vol. 21, pp. 251–260, 2016.

[9] P. Han, A. L. Ellefsen, G. Li, F. T. Holmeset, and H. Zhang, "Fault detection with lstm-based variational autoencoder for maritime components," *IEEE Sensors Journal*, vol. 21, no. 19, pp. 21903–21912, 2021.

[10] P. Han, G. Li, R. Skulstad, S. Skjong, and H. Zhang, "A deep learning approach to detect and isolate thruster failures for dynamically positioned vessels using motion data," *IEEE Transactions on Instrumentation and Measurement*, 2020.

[11] 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)*, pp. 6644–6650, IEEE, 2019.

[12] 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.

[13] D. Huang, D. Zhao, L. Wei, Z. Wang, and Y. Du, "Modeling and analysis in marine big data: Advances and challenges," *Mathematical Problems in Engineering*, vol. 2015, 2015.

[14] I. Lytra, M.-E. Vidal, F. Orlandi, and J. Attard, "A big data architecture for managing oceans of data and maritime applications," in *2017 International Conference on Engineering, Technology and Innovation (ICE/ITMC)*, pp. 1216–1226, IEEE, 2017.

[15] H.-N. Dai, H. Wang, G. Xu, J. Wan, and M. Imran, "Big data analytics for manufacturing internet of things: opportunities, challenges and enabling technologies," *Enterprise Information Systems*, vol. 14, no. 9-10, pp. 1279–1303, 2020.

[16] D. Solomatine, L. M. See, and R. Abrahart, "Data-driven modelling: concepts, approaches and experiences," *Practical hydroinformatics: Computational intelligence and technological developments in water applications*, pp. 17–30, 2008.

[17] R. Skulstad, G. Li, T. I. Fossen, B. Vik, and H. Zhang, "A hybrid approach to motion prediction for ship docking—integration of a neural network model into the ship dynamic model," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–11, 2020.

[18] P. Han, G. Li, S. Skjong, B. Wu, and H. Zhang, "Data-driven sea state estimation for vessels using multi-domain features from motion responses," in *2021 IEEE international conference on robotics and automation (ICRA)*, pp. 2120–2126, IEEE, 2021.

[19] T. Uyanık, Ç. Karatuğ, and Y. Arslanoğlu, "Machine learning approach to ship fuel consumption: A case of container vessel," *Transportation Research Part D: Transport and Environment*, vol. 84, p. 102389, 2020.

[20] K. C. A. Khanzode and R. D. Sarode, "Advantages and disadvantages of artificial intelligence and machine learning: A literature review," *International Journal of Library & Information Science (IJLIS)*, vol. 9, no. 1, p. 3, 2020.

[21] D. G. Cacuci, M. Ionescu-Bujor, and I. M. Navon, *Sensitivity and uncertainty analysis, volume II: applications to large-scale systems*, vol. 2. CRC press, 2005.

[22] H. Herodotou, S. Aslam, H. Holm, and S. Theodossiou, "Big maritime data management," in *Maritime Informatics*, pp. 313–334, Springer, 2020.

[23] M. Chen, S. Mao, and Y. Liu, "Big data: A survey," *Mobile networks and applications*, vol. 19, pp. 171–209, 2014.

[24] J. Yu, D. Wang, and M. Zheng, "Uncertainty quantification: Can we trust artificial intelligence in drug discovery?," *Iscience*, p. 104814, 2022.

[25] A. Der Kiureghian and O. Ditlevsen, "Aleatory or epistemic? does it matter?," *Structural safety*, vol. 31, no. 2, pp. 105–112, 2009.

[26] S. Razavi, A. Jakeman, A. Saltelli, C. Prieur, B. Iooss, E. Borgonovo, E. Plischke, S. L. Piano, T. Iwanaga, W. Becker, *et al.*, "The future of sensitivity analysis: An essential discipline for systems modeling and policy support," *Environmental Modelling & Software*, vol. 137, p. 104954, 2021.

[27] S. Tarantola, D. Gatelli, S. Kucherenko, W. Mauntz, *et al.*, "Estimating the approximation error when fixing unessential factors in global sensitivity analysis," *Reliability engineering & system safety*, vol. 92, no. 7, pp. 957–960, 2007.

[28] M. Mesgarpour and I. Dickinson, "Enhancing the value of commercial vehicle telematics data through analytics and optimisation techniques," *Archives of Transport System Telematics*, vol. 7, no. 3, pp. 27–30, 2014.

[29] R. Zhen, Y. Jin, Q. Hu, Z. Shao, and N. Nikitakos, "Maritime anomaly detection within coastal waters based on vessel trajectory clustering and naïve bayes classifier," *The Journal of Navigation*, vol. 70, no. 3, pp. 648–670, 2017.

[30] K. Sheng, Z. Liu, D. Zhou, A. He, and C. Feng, "Research on ship classification based on trajectory features," *The Journal of Navigation*, vol. 71, no. 1, pp. 100–116, 2018.

[31] L. Zhang, Q. Meng, and T. F. Fwa, "Big ais data based spatial-temporal analyses of ship traffic in singapore port waters," *Transportation Research Part E: Logistics and Transportation Review*, vol. 129, pp. 287–304, 2019.

[32] H. Tang, Y. Yin, and H. Shen, "A model for vessel trajectory prediction based on long short-term memory neural network," *Journal of Marine Engineering & Technology*, vol. 21, no. 3, pp. 136–145, 2022.

[33] H. Rong, A. Teixeira, and C. G. Soares, "Ship trajectory uncertainty prediction based on a gaussian process model," *Ocean Engineering*, vol. 182, pp. 499–511, 2019.

[34] M. Döhler and L. Mevel, "Efficient multi-order uncertainty computation for stochastic subspace identification," *Mechanical Systems and Signal Processing*, vol. 38, no. 2, pp. 346–366, 2013.

[35] T. S. Nord, Ø. W. Petersen, and H. Hendrikse, "Stochastic subspace identification of modal parameters during ice–structure interaction," *Philosophical Transactions of the Royal Society A*, vol. 377, no. 2155, p. 20190030, 2019.

[36] L. Wang, J.-m. Yang, and S.-w. Xu, "Dynamic positioning capability analysis for marine vessels based on a dpcap polar plot program," *China Ocean Engineering*, vol. 32, no. 1, pp. 90–98, 2018.

[37] F. Mauro and R. Nabergoj, "A probabilistic approach for dynamic positioning capability and operability predictions," *Ocean Engineering*, vol. 262, p. 112250, 2022.

[38] V. Cohen-Addad, V. Kanade, F. Mallmann-Trenn, and C. Mathieu, "Hierarchical clustering: Objective functions and algorithms," *Journal of the ACM (JACM)*, vol. 66, no. 4, pp. 1–42, 2019.

[39] J.-X. Mao, H. Wang, Y.-G. Fu, and B. F. Spencer Jr, "Automated modal identification using principal component and cluster analysis: Application to a long-span cable-stayed bridge," *Structural Control and Health Monitoring*, vol. 26, no. 10, p. e2430, 2019.

[40] G. Pamuji and H. Rongtao, "A comparison study of dbscan and k-means clustering in jakarta rainfall based on the tropical rainfall measuring mission (trmm) 1998-2007," in *IOP Conference Series: Materials Science and Engineering*, vol. 879, p. 012057, IOP Publishing, 2020.

[41] E. Schubert, J. Sander, M. Ester, H. P. Kriegel, and X. Xu, "Dbscan revisited, revisited: why and how you should (still) use dbscan," *ACM Transactions on Database Systems (TODS)*, vol. 42, no. 3, pp. 1–21, 2017.

[42] C. Cassisi, P. Montalto, M. Aliotta, A. Cannata, and A. Pulvirenti, "Similarity measures and dimensionality reduction techniques for time series data mining," *Advances in data mining knowledge discovery and applications'(InTech, Rijeka, Croatia, 2012,*, pp. 71–96, 2012.

[43] M. O. Stitson, A. Gammerman, V. Vapnik, V. Vovk, C. Watkins, and J. Weston, "Support vector regression with anova decomposition kernels," in *Advances in kernel methods*, pp. 285–291, 1999.

[44] Mattera, Davide, Haykin, and Simon, "Support vector machines for dynamic reconstruction of a chaotic system," 2008.

[45] N. Cressie, "The origins of kriging," *Mathematical geology*, vol. 22, pp. 239–252, 1990.

[46] C. K. Williams and C. E. Rasmussen, *Gaussian processes for machine learning*, vol. 2. MIT press Cambridge, MA, 2006.

[47] H. Liu, Y.-S. Ong, X. Shen, and J. Cai, "When gaussian process meets big data: A review of scalable gps," *IEEE transactions on neural networks and learning systems*, vol. 31, no. 11, pp. 4405–4423, 2020.

[48] C. Zhao, J. Lv, and S. Du, "Geometrical deviation modeling and monitoring of 3d surface based on multi-output gaussian process," *Measurement*, vol. 199, p. 111569, 2022.

[49] H. Liu, J. Cai, and Y.-S. Ong, "Remarks on multi-output gaussian process regression," *Knowledge-Based Systems*, vol. 144, pp. 102–121, 2018.

[50] E. V. Bonilla, K. Chai, and C. Williams, "Multi-task gaussian process prediction," *Advances in neural information processing systems*, vol. 20, 2007.

[51] J. Hensman, N. Fusi, and N. D. Lawrence, "Gaussian processes for big data," *arXiv preprint arXiv:1309.6835*, 2013.

[52] H. B. Moss, S. W. Ober, and V. Picheny, "Inducing point allocation for sparse gaussian processes in high-throughput bayesian optimisation," *arXiv preprint arXiv:2301.10123*, 2023.

[53] V. Todorov, I. Dimov, T. Ostromsky, S. Apostolov, R. Georgieva, Y. Dimitrov, and Z. Zlatev, "Advanced stochastic approaches for sobol'sensitivity indices evaluation," *Neural Computing and Applications*, pp. 1–16, 2020.

[54] S. Tarantola and T. A. Mara, "Variance-based sensitivity indices of computer models with dependent inputs: The fourier amplitude sensitivity test," *International Journal for Uncertainty Quantification*, vol. 7, no. 6, 2017.

[55] I. Kovacs, A. Iosub, M. Țopa, A. Buzo, and G. Pelz, "A gradient-based sensitivity analysis method for complex systems," in *2019 IEEE 25th International Symposium for Design and Technology in Electronic Packaging (SIITME)*, pp. 333–338, IEEE, 2019.

[56] F. K. Zadeh, J. Nossent, F. Sarrazin, F. Pianosi, A. V. Griensven, T. Wagener, and W. Bauwens, "Comparison of variance-based and moment-independent global sensitivity analysis approaches by application to the swat model," *Environmental Modelling & Software*, vol. 91, no. C, pp. 210–222, 2017.

[57] F. Pianosi, F. Sarrazin, and T. Wagener, *A MATLAB toolbox for global sensitivity analysis.* Elsevier Science Publishers B. V., 2015.

[58] F. Pianosi and T. Wagener, "A simple and efficient method for global sensitivity analysis based oncumulative distribution functions," *Environmental Modelling and Software*, vol. 67, pp. 1–11, 2015.

[59] A. Saltelli, *Global sensitivity analysis : the primer.* John Wiley, 2008.

[60] F. Pianosi and T. Wagener, "A simple and efficient method for global sensitivity analysis based on cumulative distribution functions," *Environmental Modelling & Software*, vol. 67, pp. 1–11, 2015.

[61] G. Ziemer, T. Stange, and Q. Hisette, "Hsva model ice–a status report," in *International Conference on Offshore Mechanics and Arctic Engineering*, vol. 85918, p. V006T07A010, American Society of Mechanical Engineers, 2022.

[62] T. Stange, G. Ziemer, and R. U. F. von Bock und Polach, "Development of an experimental setup to investigate the impact of higher structural modes on dynamic ice-structure interaction," in *25th IAHR International Symposium on Ice 2020*, 2020.

[63] T. S. Nord, "Force and response estimation on bottom-founded structures prone to ice-induced vibrations," 2015.

[64] D. Yang, L. Wu, S. Wang, H. Jia, and K. X. Li, "How big data enriches maritime research–a critical review of automatic identification system (ais) data applications," *Transport Reviews*, vol. 39, no. 6, pp. 755–773, 2019.

[65] D. AS, "Assessment of station keeping capability of dynamic positioning vessels," 2016.

[66] L. Pivano, D. Nguyen, and Ø. Smøgeli, "Full-scale validation of a vessel's station-keeping capability with dyncap," in *ASME 2017 36th International Conference on Ocean, Offshore and Arctic Engineering*, pp. V009T12A054–V009T12A054, American Society of Mechanical Engineers, 2017.

[67] S. Xu, X. Wang, L. Wang, and X. Li, "Investigation of the positioning performances for dp vessels considering thruster failure modes by a novel synthesized criterion," *Journal of Marine Science and Technology*, pp. 1–15, 2017.

[68] S. Kuhlmann and M. Tressl, "Comparison of exponential-logarithmic and logarithmic-exponential series," *Mathematical Logic Quarterly*, vol. 58, no. 6, pp. 434–448, 2012.

[69] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, 2011. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[70] M. a. Jefferies, "Dynamic response of" molikpaq" to ice-structure interaction.," *Proceedings of the 7th OMAE, Houston, February 7-12, 1988*, vol. 4, pp. 201–220, 1988.

[71] K. Blenkarn, "Measurement and analysis of ice forces on cook inlet structures," in *Offshore Technology Conference*, OnePetro, 1970.

[72] J. Zeng and Y. Hoon Kim, "A two-stage framework for automated operational modal identification," *Structure and Infrastructure Engineering*, pp. 1–20, 2021.

[73] C. Ye and X. Zhao, "Automated operational modal analysis based on dbscan clustering," in *2020 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS)*, pp. 864–869, IEEE, 2020.

[74] F. Magalhaes, A. Cunha, and E. Caetano, "Online automatic identification of the modal parameters of a long span arch bridge," *Mechanical Systems and Signal Processing*, vol. 23, no. 2, pp. 316–329, 2009.

[75] E. Reynders, J. Houbrechts, and G. De Roeck, "Fully automated (operational) modal analysis," *Mechanical Systems and Signal Processing*, vol. 29, pp. 228–250, 2012.

[76] R. S. Pappa, G. H. James III, and D. C. Zimmerman, "Autonomous modal identification of the space shuttle tail rudder," *Journal of Spacecraft and Rockets*, vol. 35, no. 2, pp. 163–169, 1998.

[77] F. Pianosi, F. Sarrazin, and T. Wagener, "A matlab toolbox for global sensitivity analysis," *Environmental Modelling & Software*, vol. 70, pp. 80–85, 2015.

[78] C. Wang, T. S. Nord, and G. Li, "Automated modal parameters identification during ice-structure interactions," in *International Conference on Offshore Mechanics and Arctic Engineering*, vol. 85864, p. V002T02A020, American Society of Mechanical Engineers, 2022.

[79] M. Själander, M. Jahre, G. Tufte, and N. Reissmann, "EPIC: An energy-efficient, high-performance GPGPU computing research infrastructure," 2019.

[80] D. Depellegrin, M. Bastianini, A. Fadini, and S. Menegon, "The effects of covid-19 induced lockdown measures on maritime settings of a coastal region," *Science of the Total Environment*, vol. 740, p. 140123, 2020.

[81] L. M. Millefiori, D. Zissis, L. Cazzanti, and G. Arcieri, "A distributed approach to estimating sea port operational regions from lots of ais data," in *2016 IEEE International Conference on Big Data (Big Data)*, pp. 1627–1632, IEEE, 2016.

[82] C. Wang, G. Li, P. Han, O. Osen, and H. Zhang, "Impacts of covid-19 on ship behaviours in port area: An ais data-based pattern recognition approach," *IEEE Transactions on Intelligent Transportation Systems*, 2022.

[83] J. Gardner, G. Pleiss, K. Q. Weinberger, D. Bindel, and A. G. Wilson, "Gpytorch: Blackbox matrix-matrix gaussian process inference with gpu acceleration," *Advances in neural information processing systems*, vol. 31, 2018.

[84] J. Wang, H. Zhang, K. Zhang, and N. R. Pal, "Bilateral sensitivity analysis for understandable neural networks and its application to reservoir engineering," 2020.

[85] E. Jeczmionek and P. A. Kowalski, "Input reduction of convolutional neural networks with global sensitivity analysis as a data-centric approach," *Neurocomputing*, vol. 506, pp. 196–205, 2022.

Appendix

A

Paper I

A sensitivity quantification approach to significance analysis of thrusters in dynamic positioning operations

Chunlin Wang [a], Guoyuan Li [a,*], Robert Skulstad [a], Xu Cheng [b], Ottar Osen [c], Houxiang Zhang [a]

[a] *Department of Ocean Operations and Civil Engineering, Norwegian University of Science and Technology, 6009, Ålesund, Norway*
[b] *Department of Manufacturing and Civil Engineering, Norwegian University of Science and Technology, 2815, Gjøvik, Norway*
[c] *Department of ICT and Natural Sciences, Norwegian University of Science and Technology, 6009, Ålesund, Norway*

ARTICLE INFO

ABSTRACT

The safety of offshore operations is highly dependent on the dynamic positioning (DP) capability of a vessel. Meanwhile, DP capability comes down to the ability of the thrust generated by thrusters to counteract environmental forces. Therefore, it is significant to investigate which thrusters are important to the position-keeping ability of vessels. However, complex environmental factors make the investigation of thrusters' importance more complicated. Hence, this paper proposes a new method to identify the influence of each thruster on vessel's station-keeping capability in different sea states. The station-keeping capability is quantified by a defined synthesized positioning ability criterion comprised by vessel position, heading angle, and consumed power. Through the comparison of different machine learning approaches, support vector machine (SVM) is used for building a surrogate model between DP capability and thrusters. In order to determine the most sensitive thruster in the whole process of vessel operation, an improved sensitivity analysis (SA) called 'PAWN' is employed along with statistical analysis to evaluate the significance of thrusters from different perspectives. Seventeen cases are investigated with respect to different thruster failures in various sea states. The results show the proposed method is able to identify the significance of each thruster in different scenarios.

1. Introduction

As the exploration and exploitation of marine resources such as oil and gas, renewable energy and other minerals, marine operations are becoming more and more frequent in recent years. Due to the influence of environmental disturbances, it represents significant safety and integrity challenges that shall threaten the offshore operations. For the sake of safe offshore operations, vessels with dynamic positioning (DP) system are playing a critical role. They can automatically maintain the desired position. In order to ensure that a loss of position shall not occur even after a worst-case failure in all components, DP 2 and DP 3 are designed with redundant power systems in which 20% of electrically generated power shall be reserved (AS, 2016). The high position-keeping ability of DP 2 and DP 3 enables them to work in various offshore operations. Their wide applications have drawn great attention from stakeholders. Many researchers devoted to optimizing control parameters, improving controller performance, and detecting thruster failure (Lee et al., 2020; Brodtkorb et al., 2018; Han et al., 2020). However, few of them investigated the interior relation between

thrusters and the vessel's DP capability. Hence, it is of great potential to analyze the interaction among thrusters and environmental factors for on-board support of the vessel's DP capabilities improvement.

In order to test the operational safety of DP vessels, a digital twin is introduced and widely used in the service of designing and evaluating system performance, safety, and structural integrity. It is a digital model that integrates data from varying sources, and can simulate all operations in the real asset while saving time and money. The digital twin has been successfully applied in a simulation of DP operations as well as the assessment of DP capability (Skulstad et al., 2019). As all DP vessels carry a risk of loss of position, which has detrimental effects on personnel, the environment and equipment (Chen Nygärdet al., 2016), they have a high requirement of DP capability. For the assessment of DP capability of a vessel, thruster's failures are also seen as the first concern in most of assessing guidelines (Karlsen et al., 2016). It makes sense to use digital simulation platform for investigating whether vessels can provide sufficient forces using the rest of thrusters to counteract against environmental loads when a certain thruster failure occurs such as a tunnel thruster failure or a main thruster failure.

---

* Corresponding author.
 E-mail address: guoyuan.li@ntnu.no (G. Li).

To date, there have been many attempts to analyze thruster failure in marine operations. Xu et al. developed a novel synthesized criterion to analyze the positioning performance of DP vessels. Various thruster failures were considered in the research (Xu et al., 2017). Benetazzo et al. utilized a parity space approach and a Luenberger observer to gain the residuals. Next, the cumulative sum algorithm was applied on these residuals to detect and isolate thruster failures (Benetazzo et al., 2015). Sheng et al. developed a program to investigate the DP capability of semi-submersible vessels under the case of thruster failure (Sheng wen et al., 2016). This research contributed to demonstrating the safety of the DP system and provided adequate guidance to the thrust system's design. Han et al. used a deep Convolutional Neural Network method to detect the potential thruster failure (Han et al., 2020). This data-driven method had a good performance to detect and isolate thruster failure without any vessel-dependent models. However, the relation between DP capability and thruster failures is not investigated further for papers as mentioned above. Xu et al. proposed a method using sensitivity analysis (SA) to investigate the influence of thrusters on positioning capability (Xu et al., 2015). However, the paper adopted local SA which can not reflect the characteristics of vessel sea-keeping ability in whole input space. Cheng et al. used global SA method to analyze thrusters' importance to ship heading (Cheng et al., 2019). Nevertheless different thruster failure cases were not considered in their study. In a word, there are few researches to carry out a comprehensive analysis of how much contribution thrusters make to DP capability in the case of various thruster failures and different sea states.

This paper proposes a novel methodology to analyze the significance of each thruster on DP capability. It could not only provide onboard support for improving DP capability, but also give guidance for power system design as well as thrusters' maintenance with the help of statistical analysis and SA. The predominant contributions are as follows: 1) positioning capability is quantified by a designed synthesized criterion made up of ship position, heading angle, and consumed power; 2) machine learning (ML) and a modified PAWN are combined to quantify the significance of each thruster; 3) this method is applied to analyze the importance of each thruster during DP operation in different failure conditions and environmental load scenarios.

The rest of this paper is structured as follows: the next section describes related works on DP capability assessment and SA. Section 3 details the procedure of obtaining significance of thrusters from data generation, data preprocessing, an optimal ML selection to significance analysis. Section 4 compares the performance of ML based on the benchmark function, and tests the ability of the proposed method to analyze the importance of thrusters using professional simulator in a variety of scenarios. Section 5 is conclusion.

2. Related work

2.1. 1Dynamic positioning capability assessment

Some offshore operations, like oil production, pipe laying, and drilling, deeply rely on DP capability to maintain vessel position or heading within an accepted criterion. Traditionally, dynamic positioning capability (DPCap) analysis is performed based on industrial standards, such as 'IMCA M140', 'DNV GL ERN', and 'ABS skp' (Pivano et al., 2017). DPCap studies test whether the vessel has favorable actuator capacity to counteract environmental loads while keeping a constant position (Wang et al., 2018a). However, they have limited ability to provide other relevant and desired information. A significant shortcoming of the quasi-static DPCap analysis is the inability to consider the transient conditions during a failure and recovery after the failure (Pivano et al., 2014).

These deficiencies call for the development of next-level DP capability analysis. Dynamic capability (DynCap) was proposed to determine the station-keeping capability of a vessel using systematic time-domain simulations. It employs a complete six-degree of freedom (DOF) vessel

model. This model includes dynamic environmental loads, a complete propulsion system with thrust losses and so on (Pivano et al., 2014). One of the advantages of the DynCap analysis, compared to traditional DPCap, is that the limiting environment can be computed by applying a set of user-defined acceptance criteria. The position and heading excursion are set to allow a wide or narrow footprint. The 'DNVGL-ST-0111' standard introduced detailed requirements, principles and acceptance criteria (AS, 2016). It also provides complete analysis methods for the three DP capability levels.

Many researchers have been working on DP capability analysis for decades. Pivano et al. performed full-scale trials using the DynCap method to validate a vessel's station-keeping capability (Pivano et al., 2017). Different comparisons were made by statistics of time-domain data with various environmental loads. Xu et al. investigated positioning performances for DP vessels considering thruster failure modes by a synthesized criterion (Xu et al., 2017). The criterion is used to quantify the positioning ability by integrating positioning accuracy and consumed power. However, these criteria cannot fully represent the DP capability from the perspective of statistics.

In this study, positioning capability refers to how well the DP vessel is positioned, instead of the extremity of the environmental conditions the vessel can counteract, as underlined by (Xu et al., 2015). Based on prior studies and our SA method (Wang et al., 2018b), positioning capability is quantified by time-series ship parameters such as ship position, heading, and consumed power. Some aforementioned statistics of time-domain data to analyze the DP capability of offshore vessels were accepted and adopted.

2.2. Sensitivity analysis

SA is a powerful tool to identify how much the variation of model output can be apportioned to inputs (Pianosi and Wagener, 2015). SA, in general, is made up of variance-based and density-based methods.

Variance-based methods includes Sobol (Todorov et al., 2020), the Fourier Amplitude Sensitivity Test (FAST) (Tarantola, Mara), and the Extend-FAST (EFAST) (Kovacs et al., 2019) and so on. A well-known advantage of variance-based methods is their ability to quantify the individual parameter contribution and the contribution resulting from parameter interactions (Zadeh et al., 2017). However, variance-based methods do not completely represent the output uncertainty when the model output is highly skewed (Pianosi et al., 2015).

To overcome this drawback, a new method called moment-independent global SA method—also known as density-based method, was proposed, which includes an Entropy-based sensitivity measure (Yun et al., 2019) and the δ- sensitivity method (Plischkeabc, 2013). However, optimal bandwidth selection has a high computational cost. Hence, the development of these methods has been limited. Francesca et al. came up with a novel SA method called 'PAWN' that characterizes the output distribution by its cumulative distribution function (CDF) instead of probability distribution function (PDF) (Pianosi and Wagener, 2015). One advantage of PAWN is that it hugely reduces computational cost because there is no need to compute unknown parameters for the approximation of empirical CDF. Another advantage is that sensitivity indices can be easily obtained, by considering either entire range of variation of the model output or a sub-range.

SA is widely applied for maritime applications with different purposes. Li et al. applied a derivative-based SA method to simplify a neural network (NN) model so as to predict ship motion (Li et al., 2017). Zhang et al. adopted a sum of square derivatives to choose inputs for the nonlinear auto aggressive model in order to create a compact ship motion model (Zhang and Liu, 2014). Mizythras et al. proposed an SA to determine parameters that have impacts on vessel propulsion and maneuverability (Mizythras et al., 2016).

In this study, based on our previous experience (Wang et al., 2018b), the PAWN method is adopted to conduct an SA of thrusters. In addition, we make some modifications and improvements according to features of
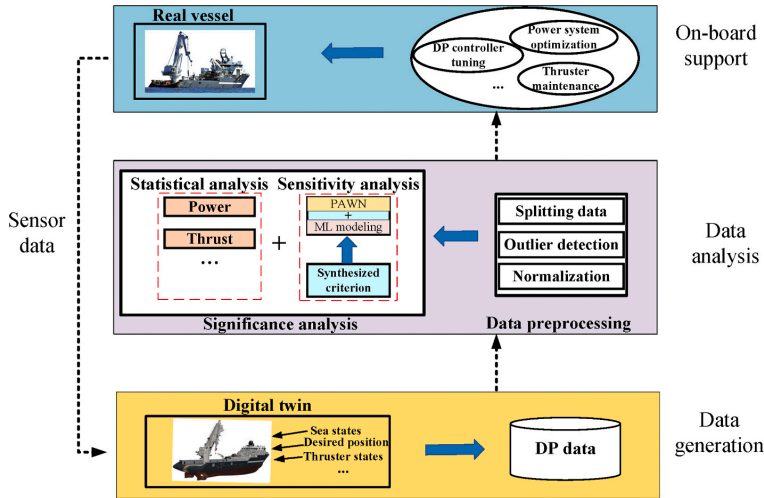
Fig. 1. The system structure of significance analysis of thrusters in DP operations.
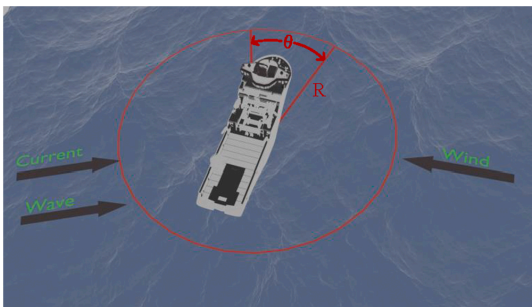
Fig. 2. DP operations of a vessel at sea.

DP data.

3. System structure

This section outlines the procedure of significance analysis of thrusters in DP operations. The workflow consists of three parts as shown in Fig. 1. The first part generates raw simulated DP data by DP simulator which is considered as a digital twin of a real vessel. Users are able to change inputs to the simulator, such as sea states, desired position, and thruster states, to simulate different scenarios to obtain several data sets. After the behavior of vessel changes over time, new raw sensor data are generated and come into the digital platform for further modeling and simulation. The second part is data analysis that is made of data preprocessing and significance analysis. Outcomes of analysis are used to offer on-board support of real vessel's operations as well as system optimization.

3.1. DP data generation

In the study, the DP data are generated from a professional simulator in the Offshore Simulator Centre — the world's most advanced provider of simulators for demanding marine operations[1]. Fig. 2 illustrates the simulator conducting DP operation under the impact of environmental disturbances. Its position is limited within a red circle whose diameter is
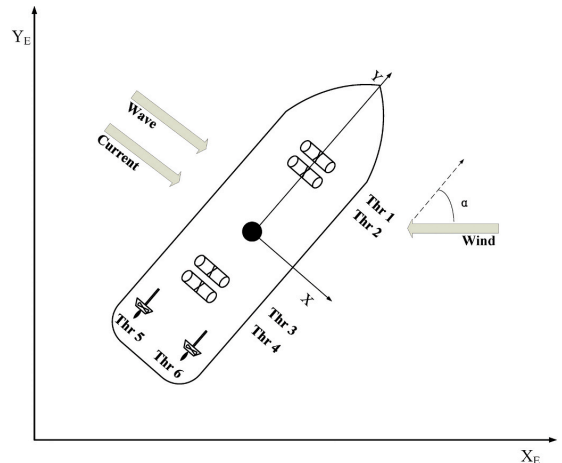
Fig. 3. The thruster configuration of DP vessel.

Table 1
Sea states.

Beaufort	Wind velocity	Wave height	Wave period	Current speed
Description	[m/s]	[m]	[s]	[m/s]
Fresh Breeze	7.90	1.30	6.50	0.75
Strong breeze	13.80	3.10	8.50	0.75

denoted as R. The limit of heading is restricted by red sector whose angle is represented as θ. Fig. 3 shows the environmental effects on the ship. Wind with an attack angle of α can be changed in the simulation. Current and wave coming from other directions are fixed in the study. In Fig. 3, the Earth-fixed reference frame is denoted as (X_E, Y_E). The body-fixed reference frame (X,Y) is fixed on the body of the vessel. Its origin is the vessel's center of gravity. The DP vessel is equipped with six

thrusters including four tunnel thrusters (Thruster 1–4) and two main thrusters (Thruster 5 and 6). In the simulator, sea state, thruster state, and the desired position are all adjustable.

In this paper, two different sea states are investigated as shown in Table 1. The desired position is set to (0, 0). Thruster states involve various thruster failure modes. Based on our experiment design, after the corresponding thrusters are shut off, the rest of thrusters are used for actuating vessel to generate several groups of time-domain DP data. For each sea state, experiments are performed on different thruster failure modes. Then ship position and heading are obtained after each experiment, and the other ship state parameters such as thruster arguments are obtained as shown in Table 3. These time series data are raw DP data. They will be processed in the following step.

3.2. Data preprocessing

Data preprocessing makes it possible to ensure efficiency and accuracy for computation of the computed PAWN sensitivity indices. It requires three substeps that are splitting data, denoising, and normalization. This experiment was set as a ship that was intact at the beginning but in failure mode by the end. The whole experiment produced a lot of time-series DP data related to various combinations of thruster failure modes and sea states. These data are full of anomalies resulting from noise, which would threaten the accuracy of SA. In this paper, Isolation Forest (iForest) was applied for data cleaning. The iForest is an algorithm that uses a tree structure to isolate instances (Schneider et al., 2016). It can (i) achieve a low linear time-complexity and a small memory-requirement, and (ii) deal with the effects of swamping and masking effectively. iForest detection is a two-stage process. The first stage uses the given training data to build an isolation tree. The second one computes an average path length of each instance through isolation trees.

Let $X = [x_1, x_2, ..., x_m] \subseteq R^{m \times d}$ be a sample set of m instances with d-variate distribution. Firstly, iForest is constructed by the proposed algorithm in (Liu et al., 2012). Secondly, path length h(x) of each instance is computed by counting the number of edges from the root node to a leaf node in an iTree. Next, Eq. (1) is used to gain c(ψ) that is the average of h(x) given ψ.

$$c(\psi) = \begin{cases} 2H(\psi - 1) - 2(\psi - 1)/m & \psi > 2, \\ 1 & \psi = 2, \\ 0 & otherwise. \end{cases} \quad (1)$$

where ψ is the subsampling size during the stage of building an iForest; H(I) is the harmonic number which can be estimated by Euler's constant (ln(i)+0.57721). Finally, Eq. (2) is used to calculate the score of every instance:

$$s(x, \psi) = 2^{-\frac{E(h(x))}{c(\psi)}} \quad (2)$$

where E(h(x)) is the expectation of h(x) from the collection of iTrees. If s is close to 1, then the instance is seen as an anomaly and removed from the data set.

After data cleaning, these data need to be normalized in the range of [0, 1] by Eq. (3) for the purpose of formulating a synthesized criterion.

$$\widetilde{x}_i = \frac{\widehat{x}_i - min(\widehat{X})}{max(\widehat{X}) - min(\widehat{X})} \quad i = 1 \dots l \quad (3)$$

Where $\widehat{X} = [\widehat{x_1}, \widehat{x_2}, ..., \widehat{x_l}] \subseteq \mathbb{R}^{l \times d}$. Therein, l is smaller than m because some abnormal instances are removed. After the procedure of data preprocessing, the processed data will be used to create a synthesized criterion to construct a surrogate model.

3.3. Significance analysis

Significance analysis is the last step to identify the significance of thrusters. It is comprised of statistical analysis and SA. These two methods can analyze the significance of thrusters from different respects. Meanwhile, the integration of both methods can provide guidance for power allocation, DP system optimization. Statistical analysis focuses on statistical features of DP data by virtue of mean, maximum value, variance and PDF (Pivano et al., 2017). As a supplementary instruction for SA, it is able to show the variation of each of data attributes intuitively. SA is capable of quantifying the contribution of each thruster to DP capability. It is comprised of three portions: 1) proposing a synthesized criteria to quantify DP capability; 2) selecting an optimal ML method to build a surrogate model; 3) using PAWN to compute sensitivity indices.

3.3.1. Synthesized criterion

To investigate the significance of every thruster on positioning capability in different sea states and failure modes, a synthesized criterion that quantifies the positioning performance needs to be defined. This criterion is used to evaluate how well the ship is positioned. According to the DP capability level in 'DNVGL-ST-0111' standard, assessment of station-keeping capability is mainly based on statistics of the position deviation and heading deviation. Therefore, position and heading should be integrated into the synthesized criterion. In addition, for ensuring the safety and accuracy of DP operations, the DP vessel has a higher power requirement than other conventional vessels (Xu et al., 2017). Therefore, power consumption is also taken into consideration in this criterion. As a result, we create a synthesized criterion by Eq. (4) to lump the above-mentioned ship parameters together, with extra modification to make it adapt to the SA method.

$$\begin{cases} V = \omega_1 \times D + \omega_2 \times A + \omega_3 \times P \\ \omega_1 + \omega_2 + \omega_3 = 1 \\ Cri = -\ln(V) \quad V > 0 \end{cases} \quad (4)$$

where ω_1, ω_2, and ω_3 are weighting factors within [0, 1]; D is position deviation computed by the distance between current and original position; A denotes the heading angle variation; P represents total power consumed by thrusters; Cri is the synthesized criterion computed by the inverse of the monotone increasing function 'ln'. The larger V is, the worse the positioning capability (Cri). Compared to the exponential function in the interval [0, 1], the minus of 'ln' function can amplify the value of V to better reflect the distinction of positioning capability (Kuhlmann and Tressl, 2012). Cri will be used as the model output when ML trains a surrogate model between thrusters' parameters and DP capability.

3.3.2. Sensitivity analysis

A modified PAWN is adopted as an SA method to quantify the influence of thrusters to positioning capability. Compared with traditional method, it is able to overcome the issue of being hard to define three parameters, i.e., the number of unconditional input samples (N_u), the number of conditional input samples (N_c), and the number of conditional points (n) (Pianosi and Wagener, 2018).

Let $< \widetilde{X}, Y >$ be a generic sample where $\widetilde{X}$ is the processed input samples; Y denotes the value of quantifying DP capability. It is handled by splitting the range of input factor $\widetilde{x}_i$ into n equal subintervals $\{I_k\}$. The PAWN indices approximation is shown as follows:

$$\begin{cases} \widehat{S}_i = \max_{k=1,...,l} KS(I_k) \\ KS(I_k) = \max_y \left| F_y(y) - F_{y|\widetilde{x}_i}(y | \widetilde{x}_i \in I_k) \right| \end{cases} \quad (5)$$

where $\widehat{S}_i$ is sensitivity index; KS is Kolmogorov-Smirnov statistic; $F_y(y)$ is
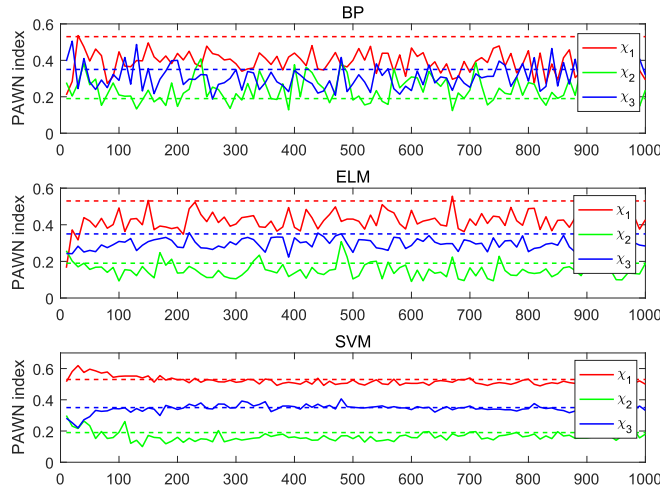
Fig. 4. SA results computed by PAWN based on different ML methods.

unconditional CDF where $y \subseteq Y$ and $F_{y|\widetilde{x_i}}(y|\widetilde{x_i} \in I_k)$ is conditional CDF where $\widetilde{x_i}$ is fixed. Using Eq. (5) to compute the sensitivity index ensures there is no need to specify N_c. It coincides with the number of points in I_k as approximately N/n, where N is the size of the generic sample. As for the unconditional sample N_u, a better option is to use subsample of Y as the conditional ones i.e., $N_u = N_c$.

The process of SA executed by PAWN combined with ML is shown in Algorithm 1. In this algorithm, 'LIBSVM' is used as an SVM tool to train the surrogate model (Chang and Lin, 2011). The model training parameters like 's', 't', 'bestc', 'bestg', 'p', 'v', and the introduction of functions like 'SVMcgForRegress', 'libsvmtrain', and 'libsvmpredict' can be found in (Chang and Lin, 2011). This algorithm mainly includes three parts. The first part is modelling (line 2–6). The thrust of all thrusters is the model input, and the positioning capability as defined by Cri above is the model output. ML is employed to construct a surrogate model between the model input and output. The second part is resampling (line 6–7). 'Unconditional_sampling' is used to generate unconditional samples; 'PAWN_sampling' is used to gain conditional samples. The last part is sensitivity index computation (line 9–10). The 'PAWN' indices of all thrusters are computed by 'PAWN_index'. Its function is shown in line 11–17. Line 12–13 is to calculate the unconditional output and conditional output. Line 14–16 is to compute the 'PAWN' index using Eq. (5). Detailed computing process could be found in (Pianosi and Wagener, 2018). The introduction of parameters and functions regarding PAWN method can be found in (Pianosi et al., 2015).


```
Algorithm 1: SA algorithm
   Input: Thrust, Cri s, t, p, v
   Output: SA_index
1  for i = 1 : num do
2      X ← Thrust
3      Y ← Cri
4      [bestc, bestg] ← SVMcgForRegress(X, Y)
5      cmd ← [s, t, bestc, bestg, p, v]
6      model ← libsvmtrain(X, Y, cmd)
7      U ← Unconditional_sampling
8      C ← PAWN_sampling
9      index(i) ← PAWN_index(U, C, model)
10  SA_index ← index/num
11  Function PAWN_index(Xu, XX, model):
12      Yu ← libsvmpredict(Xu, model)
13      YY ← libsvmpredict(XX, model)
14      [YF, Fu, Fc] ← PAWN_cdf(Yu, YY)
15      KS ← PAWN_ks(YF, Fu, Fc)
16      index ← max(KS)
17      return index
```

4. Case study

4.1. An optimal ML selection based on Ishigami function

In order to find an optimal modeling method, first of all, three prevalent ML methods, such as back propagation (BP), regularized extreme learning machine (RELM), and SVM, are introduced into training models (Wang et al., 2018b; Cui and Jing, 2019; Shao and Er, 2016). Next, PAWN combined with these three models is used to compute sensitivity indices of three parameters of Ishigami function. Finally, SA results are compared with a benchmark to identify the

Table 2
Parameters of the offshore vessel.

Items	Values
Length between perpendiculars [m]	82.7
Breadth [m]	23.0
Draught [m]	7.5
Tunnel thruster propulsion [KN]	≤173.0
Main thruster propulsion [KN]	≤1350.0

Table 3
The variables of DP data.

	Inputs	Unit
Ship status	east position	[m]
	west position	[m]
	heading	[deg]
Thruster 1	rpm	[RPM]
	thrust	[KN]
	consumed power	[KW]
Thruster 2	rpm	[RPM]
	thrust	[KN]
	consumed power	[KW]
Thruster 3	rpm	[RPM]
	thrust	[KN]
	consumed power	[KW]
Thruster 4	rpm	[RPM]
	thrust	[KN]
	consumed power	[KW]
Thruster 5	rpm	[RPM]
	thrust	[KN]
	consumed power	[KW]
Thruster 6	rpm	[RPM]
	thrust	[KN]
	consumed power	[KW]

Table 4
Environment and thruster failures setting for significance analysis.

Sea states	Attack angle [deg]	Thruster failure
Strong breeze	45	011111
		101111
		110111
		111011
		111101
		111110
		110110
Strong breeze	90	101111
		110110
Strong breeze	135	101111
		110110
Fresh breeze	45	101111
		110110

detailed reasoning process has been interpreted in (Fossen, 2011).

The attack angles α is set as $[45°, 90°, 135°]$ for different scenarios. The direction of current and wave is fixed for simplifying the experiment. The limits of ship position and heading are set as R = 3m and θ = 6°, respectively. Table 4 lists four different combinations of sea states and attack angle. They are 'strong breeze 45°', 'strong breeze 90°', 'strong breeze 135°', and 'fresh breeze 45°'. For 'strong breeze 45°', there are seven different thruster failure modes represented by binary string: '011111', '101111', '110111', '111011', '111101', '111110', '110110'. Here, '0' denotes the thruster is malfunctioning; '1' denotes the thruster is working normally. For example, '101111' indicates the second thruster is malfunctioning while the others are working normally. The required parameters of ship states are listed in the Table 3. The sampling frequency is set as 20HZ.

The synthesized criterion involves specifying three weighting factors ω_1, ω_2, and ω_3. In this study, we set $\omega_1 = 0.5$, $\omega_2 = 0.4$, and $\omega_3 = 0.1$ based on the following reasons. On the one hand, since DP vessels are designed with the redundant power system, in general, 20% of power will be reserved to avoid loss-of-position occurrence. That indicates the power is sufficient to keep a vessel's position and heading during DP operations. Therefore, power utilization was considered the least important factor in the criterion. On the other hand, ship position is seen as the most significant factor because the loss of position brings a more considerable detrimental impact on DP operations than heading. For PAWN, n is set to 10 based on the samples of data as well as experience as described in other papers (Wang et al., 2018b; Pianosi and Wagener, 2018).

In this paper, the experiment investigates the significance of thrusters under circumstances of different thruster failures in two sea states. Using the proposed method for timely computation of thrusters' sensitivity is studied as well.

4.3. Significance analysis in different thruster failure modes at two sea states

This section mainly analyzes and compares SA results in different environmental factors and thruster conditions. Table 5 lists the SA results of thruster failures at the strong breeze and fresh breeze sea states. It is found that thruster 5 is more significant than the rest of thrusters in most cases. Its contribution accounts for around 30% ~40%. Especially, when thruster 6 fails to work, the significance of thruster 5 exceeds 35% because thruster 5 as the only main propeller must generate much more thrust to counteract the influence of environmental disturbances. When one thruster failure occurs, the significance of thrusters that play a complementary role will have a significant increase as shown in Table 5. For example, the PAWN index of thruster 6 increases from 8% to 30% when thruster 5 fails in 'strong breeze 45°'. The same happens to thruster 1 and 2. For the case of '101111' in 'strong breeze 45°', for instance, the significance of thruster 1 rises by 13% up to 26.42%. For dual thruster failure '110110' in all sea states, at least two of tunnel

optimal ML method for analyzing the significance of thrusters.

In the course of determining an optimal ML method, Ishigami function is selected as a mathematical model, because Ishigami is a widely-used benchmark model that is applied to test the validity of sensitivity analysis method (Pianosi and Wagener, 2015). It is shown in Eq. (6).

$$y = sin(\chi_1) + asin(\chi_2)^2 + b\chi_3^4 sin(\chi_1) \qquad (6)$$

where a and b are random constants that can influence the sensitivity index of χ_i, $i \in \{1, 2, 3\}$. χ_i follows a uniform distribution over $[-\pi, \pi]$. Here, we set a = 2 and b = 1. Fig. 4 displays SA results as well as benchmark value. The dotted line is the benchmark value of sensitivity indices of the three parameters χ_i in Eq. (6). The corresponding sensitivity indices are $S_1 = 0.53$, $S_2 = 0.19$, and $S_3 = 0.35$, respectively. It is evident that both BP and RELM cannot figure sensitivity index out correctly; whereas PAWN combined with SVM has a better approximation to the benchmark. Therefore, SVM is selected as modelling method in the follow-up sensitivity analysis of thrusters in different scenarios.

4.2. Experimental design

This significance analysis of thrusters is conducted to determine the variation of positioning capability apportioned to each thruster. The specifications of the vessel are listed in Table 2. This vessel is actuated by six thrusters shown in Fig. 3. The actuator forces relate to the control forces and moments by $\tau = \mathbf{T}(\xi)\mathbf{f}$, where $\xi = [\xi_1, ..., \xi_p] \in \mathbb{R}^p$ is a vector of azimuth angles and $\mathbf{T}(\xi)$ is the thrust configuration matrix (Fossen, 2011). In this paper, ξ is fixed. In order to obtain the demanded thrust for each thruster, an unconstrained least-squares (LS) optimization problem is constructed. Through using Lagrange Multipliers to solve LS optimization problem, we can obtain $\mathbf{f} = \mathbf{T}^\dagger \boldsymbol{\tau}$, where $\mathbf{T}^\dagger = \mathbf{W}^{-1}\mathbf{T}^\top(\mathbf{T}\mathbf{W}^{-1}\mathbf{T}^\top)^{-1}$ is recognized as the generalized inverse (GI) matrix. Here, $\mathbf{W}$ is a positive definite matrix weighting the control forces. The

Table 5
SA results of thruster failures in strong breeze and fresh breeze.

Sea states	Direction (deg)	Thruster failure	PAWN index					
			Thr1	Thr2	Thr3	Thr4	Thr5	Thr6
Strong breeze	45	111111	0.1342	0.1040	0.1576	0.2246	**0.2976**	0.0817
		011111	0	**0.3701**	0.1448	0.1480	0.2284	0.1087
		101111	0.2642	0	0.0604	0.1069	**0.3222**	0.2459
		110111	0.1992	0.2080	0	0.0850	**0.3058**	0.2019
		111011	0.1415	0.2483	0.0853	0	**0.3472**	0.1775
		111101	0.2629	0.1839	0.1433	0.1098	0	**0.3000**
		111110	0.1674	0.1225	0.1435	0.1456	**0.4209**	0
		110110	0.2106	0.2026	0	0.2050	**0.3818**	0
	90	111111	**0.2877**	0.1211	0.0829	0.1485	0.1313	0.2283
		101111	0.2723	0	0.1103	0.1100	**0.2985**	0.2089
		110110	0.0737	0.3337	0	0.2392	**0.3534**	0
	135	111111	0.1832	0.1638	0.1224	0.1888	**0.2544**	0.0873
		101111	0.0987	0	**0.3491**	0.3273	0.1268	0.0980
		110110	0.2285	**0.4016**	0	0.0997	0.2702	0
Fresh breeze	45	111111	0.1373	0.0729	0.1317	0.0771	**0.3460**	0.2350
		101111	0.2591	0	0.1007	0.0901	**0.3401**	0.2099
		110110	0.1826	0.2113	0	0.2282	**0.3780**	0

Fig. 5. The SA result and average thrust of 6 thrusters for '111111' in 'strong breeze 45°'.

Fig. 6. The PDF of consumed power and thrust. generated by thrusters 5 and 6 for '111111' with strong breeze and α = 45°.

thrusters' significance go up to over 20% compared with one thruster failure. That possibly results from the drastic variation of the ship heading. It is reflected from the above analysis that the significance of thrusters depends on the conjunction of sea states, wind direction as well as thruster failures.

Next, significance analysis of thrusters is carried out in detail from the respect of statistics and SA. In order to illustrate how to do analysis by SA coupled with statistical analysis, we will use '111111' in the case of strong breeze with an attack angle 45° as an example. For the case of '111111' in 'strong breeze 45°', an average of thrust and SA results are shown in Fig. 5. The left y-axis represents the PAWN index of each thruster while the right one denotes mean value of thrust. These two analysis methods are able to show the importance of thrusters from their own perspective. In addition, there are interior connections between these two methods. The results of SA show that the order of importance of thrusters is quite as similar as that of statistical analysis. The PAWN index shows that thruster 5 has the most influential effect on positioning capability, at 29.76%. The second-largest effect is thruster 4, accounting for roughly 22.46%. Thrusters 3, 1, 2, and 6 follow in that order. Thruster 6 makes only an 8.17% contribution to the station-keeping ability of DP vessel despite its similarities to thruster 5, which makes the largest contribution. However these two methods show some distinctions, such as inconsistency of SA results with statistical analysis for
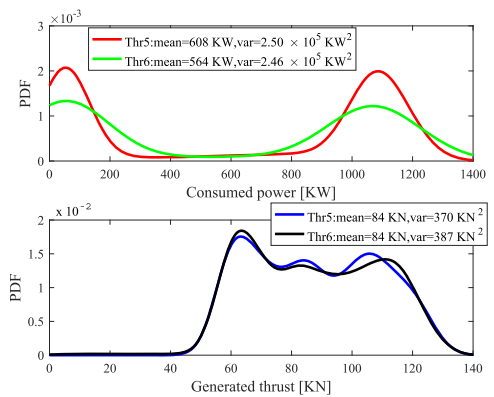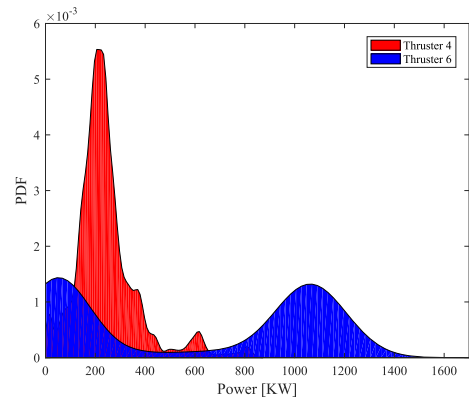

Fig. 7. The power consumed by thrusters 4 and 6 in 'strongbreeze 45° 111111'.
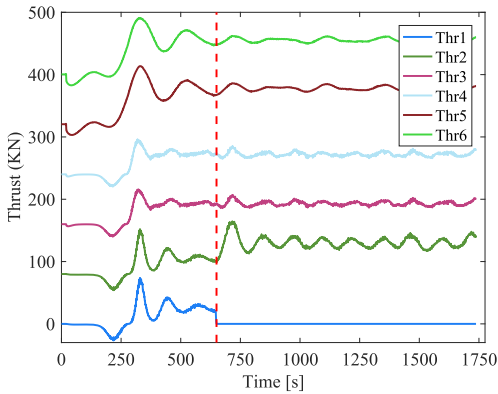
Fig. 8. Time-domain variation of thrust from '111111' to '011111'.

thruster 6.

From the perspective of statistics, thruster 6 has as much thrust as thruster 5 as shown in Fig. 6. The mean and variance of thrust generated by thruster 5 are the same as those generated by thruster 6. The two thrusters also consume the same amount of power and have similar statistical features. But observing results obtained by the proposed SA method in Fig. 5, in which all SA indices are drawn as blue bars, shows that thruster 6 is far less significant than thruster 5. It is even less than thruster 2. It reveals that SA results do not entirely conform with results obtained by statistical analysis. Both methods did give us insights that thruster 6 consumed amounts of power but generated too much useless force in this case.

To obtain more insights from Fig. 5, thrusters 4 and 6 are for detailed investigation. Fig. 7 displays the PDF of power consumed by thrusters 4 and 6, respectively. The power consumed mostly appears in the interval [100 KW, 600 KW], which is far less than the power consumed by thruster 6 as shown in the blue area. Moreover, the mean of thrust generated by thruster 4 is far less than that generated by thruster 6. Based on Figs. 5 and 7, we can find that thruster 6 consumed more power and generated more thrust but less contribution than thruster 4.

Through SA and statistical analysis, it is definitely found that some thrusters have fewer influences on DP capability, although they consumed more power. That results in a waste of power. Therefore, significance results could be used to provide guidance to improve the power allocation algorithm. For example, sensitivity indices as

weighting factors are added into the algorithm. In this case, thruster 6 with high power consumption but a little contribution to DP capability will be reallocated less power by the power system. Instead, more power should be redistributed to thruster 4, which could improve DP capability with low power consumption.

4.4. Real-time computation of thrusters' sensitivity

Although the existing method is efficient to analyze the thrusters' significance in (Xu et al., 2015), it is not competent in the real-time computation of thrusters' sensitivity. This section is to verify the feasibility of the proposed method in estimating thrusters' sensitivity online.

A simulation experiment is carried out when thruster state changes from '111111' to '011111' in 'strong breeze 45°'. The thrust generated by thrusters is shown in Fig. 8. Red dotted line represents the point at which thruster 1 fails to work. In order to visualize each curve clearly, multiple shifts of 80 KN along the y-axis direction is performed for thruster 2–6. In fact, the value of the thrust of all thrusters starts from 0.

Fig. 9 shows the variation of sensitivity indices of thrusters over time. The horizontal axis denotes sensitivity index is computed at a window time of 25s that comprises 500 sample points. Evidently, the proposed method is able to gain the contribution of each thruster to the DP capability in the process of vessel counteracting against environmental forces. Especially, when thruster 1 shuts down at 650s depicted by a red circle, the importance of thruster 1 becomes 0 thereafter. On the other
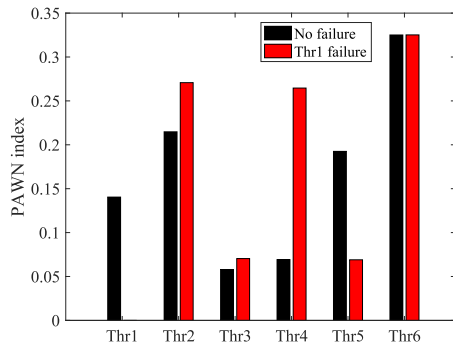

Fig. 10. The instant variation of the significance of thrusters before and after thruster 1 failure.
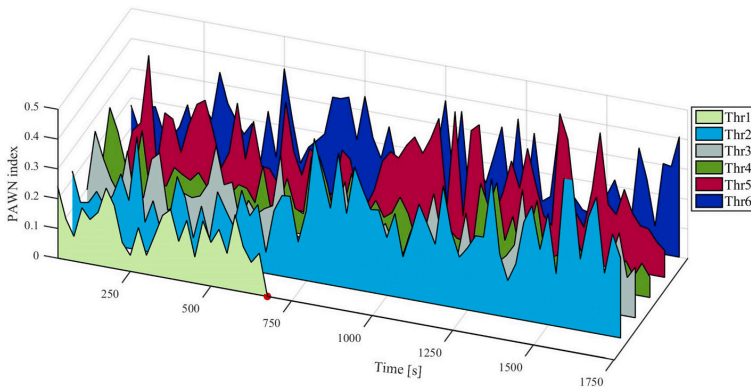

Fig. 9. Real-time computation of the significance of thrusters.

hand, thruster 2 plays a more and more important role since this point. This is because thruster 1 and 2 are bow thrusters, as shown in Fig. 3, the malfunction of thruster 1 leads to the rise of thruster 2 importance in the long term. In addition, the importance of other thrusters rises to some extent as well after thruster 1 fails to work.

At the point of 650s, the detailed information can be found in the Fig. 10. From this figure, the importance of thruster 2 and 4 grow rapidly compared with other thrusters. Therefore, the instant change of the indices could provide the operator evidence to improve the power-consuming of thruster 2 and 4 to promote the DP capability quickly after the failure of thruster 1.

To sum up, the proposed method is capable of finding the contribution of all thrusters in a real-time manner.

4.5. Discussion

For the case of '111111' in 'strong breeze 45°' in Fig. 6, the discrepancy in terms of power and thrust between thruster 5 and 6 possibly results from the fact that the rudder angle of main thrusters is fixed. As shown in Fig. 3, in order to resist the wind whose attack angle is 45°, thruster 5 must bear much more load than thruster 6. Therefore, the power and thrust of thruster 5 vary more drastically compared with those of thruster 6. It can be shown from the above analysis that thruster's importance is affected by a synthesized factor, including the configuration of thrusters, the attack angle of sea states, and the thrust allocation algorithm.

In Fig. 4, the result of BP and ELM is not as ideal as that of SVM. This situation mainly results from the limited training sample on account of online significance analysis. Considering the requirement of on-board support, therefore, SVM is used for the real-time estimation of sensitivity indices. Since the sensitivity index computed by SVM can converge to a stable value after 500 training samples, we chose a window time of 25s corresponding to 500 training samples under the sampling frequency of 20HZ in Section 4.4.

6. Conclusion

This paper proposes a method that mainly focuses on studying the significance of thrusters based on a synthesized positioning capability criterion in different thruster failure conditions. In order to quantify the DP capability, a synthesized assessment criterion is proposed by integrating ship position, heading and power. Next, the Ishigami function is used as a benchmark to determine an optimal modelling method. Through the comparison with ANN and ELM, SVM is selected to construct a surrogate model between thrusters and DP capability. Finally, different thruster failure cases in two sea states are designed to elaborate on how statistical features and SA are combined to quantify and analyze the significance of thrusters.

The purpose of significance analysis results is as follows: 1) they can provide onboard support to control power system to allocate more power to the most significant thruster when thruster fails to work, which contributes to efficiently improving DP capability; 2) they also can be used to provide guidance to optimize power allocation. By observing statistics of power, and sensitivity results, thrusters that consumed more power but made much less contribution to positioning capability should be reallocated less power. This is able to be accomplished by, for example, adding sensitivity indices as weighting factors into the allocation algorithm. That is helpful to improve vessel's DP capability with less power consumption.

For future work, efforts will be put on investigating the impact of azimuth thrusters and the thrust allocation logic on the significance of thrusters in DP operations.

CRediT authorship contribution statement

Chunlin Wang: Conceptualization, Methodology, Software, Formal analysis, Investigation, Visualization, Writing - original draft. **Guoyuan Li:** Conceptualization, Resources, Writing - review & editing, Data curation, Supervision. **Robert Skulstad:** Software, Investigation, Validation. **Xu Cheng:** Investigation, Software. **Ottar Osen:** Resources, Writing - review & editing. **Houxiang Zhang:** Writing - review & editing, Supervision, Project administration, Funding acquisition.

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.

References

AS, D., 2016. Assessment of Station Keeping Capability of Dynamic Positioning Vessels.

Benetazzo, F., Ippoliti, G., Longhi, S., Raspa, P., 2015. Advanced control for fault-tolerant dynamic positioning of an offshore supply vessel. Ocean Eng. 106, 472–484.

Brodtkorb, A.H., Værnø, S.A., Teel, A.R., Sørensen, A.J., Skjetne, R., 2018. Hybrid controller concept for dynamic positioning of marine vessels with experimental results. Automatica 93, 489–497.

Chang, C.-C., Lin, C.-J., 2011. LIBSVM: a library for support vector machines. ACM Transactions on Intelligent Systems and Technology 2 (27), 1–27, 27, software available at: http://www.csie.ntu.edu.tw/~cjlin/libsvm.

Chen, H., Nygård, B., et al., 2016. Quantified risk analysis of dp operations principles and challenges. In: SPE International Conference and Exhibition on Health, Safety, Security, Environment, and Social Responsibility. Society of Petroleum Engineers.

Cheng, X., Li, G., Skulstad, R., Chen, S., Hildre, H.P., Zhang, H., 2019. A neural-network-based sensitivity analysis approach for data-driven modeling of shipmotion. IEEE J. Ocean. Eng. 45 (2), 451–461.

Cui, K., Jing, X., 2019. Research on prediction model of geotechnical parameters based on bp neural network. Neural Comput. Appl. 31 (12), 8205–8215.

Fossen, T.I., 2011. Handbook of Marine Craft Hydrodynamics and Motion Control. John Wiley & Sons.

Han, P., Li, G., Skulstad, R., Skjong, S., Zhang, H., 2020. A deep learning approach to detect and isolate thruster failures for dynamically positioned vessels using motion data. IEEE Trans. Instr. Meas. 70, 1–11.

Karlsen, A., Pivano, L., Ruth, E., 2016. Dnv gl dp capability-a new standard for assessment of the station-keeping capability of dp vessels. In: Proceedings of Marine Technology Society (MTS) DP Conference, pp. 1–15. Houston (TX), USA.

Kovacs, I., Iosub, A., Topa, M., Buzo, A., Pelz, G., 2019. A gradient-based sensitivity analysis method for complex systems. In: 2019 IEEE 25th International Symposium for Design and Technology in Electronic Packaging (SIITME). IEEE, pp. 333–338.

Kuhlmann, S., Tressl, M., 2012. Comparison of exponential-logarithmic and logarithmic-exponential series. Math. Logic Q. 58 (6), 434–448.

Lee, D., Lee, S.J., Yim, S.C., 2020. Reinforcement learning-based adaptive pid controller for dps. Ocean Eng. 216, 108053.

Li, G., Kawan, B., Wang, H., Zhang, H., 2017. Neural-network-based modelling and analysis for time series prediction of ship motion. Ship Technol. Res. 64 (1), 30–39.

Liu, F.T., Ting, K.M., Zhou, Z.H., 2012. Isolation-based anomaly detection. ACM Trans. Knowl. Discov. Data 6 (1), 1–39.

Mizythras, P., Boulougouris, E., Priftis, A., Incecik, A., Turan, O., Reddy, D., 2016. Sensitivity analysis of the tool for assessing safe manoeuvrability of ships in adverse sea conditions. In: International Conference on Shipping in Changing Climates 2016, pp. 1–13.

Pianosi, F., Wagener, T., 2015. A simple and efficient method for global sensitivity analysis based oncumulative distribution functions. Environ. Model. Software 67, 1–11.

Pianosi, F., Wagener, T., 2018. Distribution-based sensitivity analysis from a generic input-output sample. Environ. Model. Software 108, 197–207.

Pianosi, F., Sarrazin, F., Wagener, T., 2015. A MATLAB Toolbox for Global Sensitivity Analysis. Elsevier Science Publishers B. V.

Pivano, L., Smøgeli, Ø., Muddusetti, S., Ramsey, J., 2014. Better analysis better data-better decisions-better operational risk management= delivery of incident free operations: enabled by dyncap. In: Dyn. Position. Conf.

Pivano, L., Nguyen, D., Smøgeli, Ø., 2017. Full-scale validation of a vessel's station-keeping capability with dyncap. In: ASME 2017 36th International Conference on Ocean, Offshore and Arctic Engineering. American Society of Mechanical Engineers. V009T12A054– V009T12A054.

Plischkeabc, E., 2013. Global sensitivity measures from given data. Eur. J. Oper. Res. 226 (3), 536–550.

Schneider, M., Ertel, W., Ramos, F., 2016. Expected similarity estimation for large-scale batch and streaming anomaly detection. Mach. Learn. 105 (3), 1–29.

Shao, Z., Er, M.J., 2016. An online sequential learning algorithm for regularized extreme learning machine. Neurocomputing 173, 788.

Sheng wen, X.U., Wang, X.F., Wang, L., 2016. A dynamic positioning capability analysis for a semi-submersible considering thruster failure mode. J. Ship Mech. 5, 558–565.

Skulstad, R., Li, G., Fossen, T.I., Vik, B., Zhang, H., 2019. Dead reckoning of dynamically positioned ships: using an efficient recurrent neural network. IEEE Robot. Autom. Mag. 26 (3), 39–51.

Tarantola, S., Mara, T.A., 2017. Variance-based sensitivity indices of computermodelswith dependent inputs: the fourier amplitude sensitivity test. Int. J. Uncertain. Quantification 7 (6).

Todorov, V., Dimov, I., Ostromsky, T., Apostolov, S., Georgieva, R., Dimitrov, Y., Zlatev, Z., 2020. Advanced stochastic approaches for sobol'sensitivity indices evaluation. Neural Comput. Appl. 1–16.

Wang, L., Yang, J.-m., Xu, S.-w., 2018a. Dynamic positioning capability analysis for marine vessels based on a dpcap polar plot program. China Ocean Eng. 32 (1), 90–98.

Wang, C., Cheng, X., Chen, S., Li, G., Zhang, H., 2018b. A svm-based sensitivity analysis approach for data-driven modeling of ship motion. In: 2018 IEEE International Conference on Mechatronics and Automation (ICMA). IEEE, pp. 803–808.

Xu, S., Wang, X., Wang, L., Meng, S., Li, B., 2015. A thrust sensitivity analysis based on a synthesized positioning capability criterion in dpcap/dyncap analysis for marine vessels. Ocean Eng. 108, 164–172.

Xu, S., Wang, X., Wang, L., Li, X., 2017. Investigation of the positioning performances for dp vessels considering thruster failure modes by a novel synthesized criterion. J. Mar. Sci. Technol. 1–15.

Yun, W., Lu, Z., Jiang, X., 2019. An efficient method for moment-independent global sensitivity analysis by dimensional reduction technique and principle of maximum entropy. Reliab. Eng. Syst. Saf. 187, 174–182.

Zadeh, F.K., Nossent, J., Sarrazin, F., Pianosi, F., Griensven, A.V., Wagener, T., Bauwens, W., 2017. Comparison of variance-based and moment-independent global sensitivity analysis approaches by application to the swat model. Environ. Model. Software 91 (C), 210–222.

Zhang, W., Liu, Z., 2014. Real-time ship motion prediction based on time delay wavelet neural network. J. Appl. Math. 2014.

This paper is not included due to copyright restrictions

C

Paper III

OMAE2022-81075

AUTOMATED MODAL PARAMETERS IDENTIFICATION DURING ICE-STRUCTURE INTERACTIONS

Chunlin Wang
Department of Ocean Operations
and Civil Engineering
Norwegian University of Science
and Technology
6009 Ålesund, Norway
Email: chunliw@ntnu.no

Torodd Skjerve Nord
Department of Ocean Operations
and Civil Engineering
Norwegian University of Science
and Technology
6009 Ålesund, Norway
Email: torodd.nord@ntnu.no

Guoyuan Li
Department of Ocean Operations
and Civil Engineering
Norwegian University of Science
and Technology
6009 Ålesund, Norway
Email: guoyuan.li@ntnu.no

ABSTRACT

Offshore structures are prone to damage caused by ice-induced vibrations. It is presently unknown to what extent different ice conditions change the properties of the structure, such as natural frequency, damping ratio, and mode shape. Understanding the dynamic interaction between ice and structures are important for the operational ability of offshore structures. In this study, the covariance-driven stochastic subspace identification algorithm (SSI-cov) is introduced to identify modal parameters of a scale-model structure during ice-structure interactions. In order to reduce the number of user interactions and inherent bias to the identified modal parameters, we therefore introduce an automated parameter identification approach. First, SSI-cov is used to obtain poles that describe the information: damping ratio, mode shape, etc. After that, a stable criterion is used to pick up stable poles. Finally, Hierarchical clustering is used to cluster poles to identify the natural frequency. The proposed method is able to reduce the many user-intervenes and enables efficient automatic parameter identification. The results show that Hierarchical clustering can render more successful identifications than the slack value-based method among different ice speeds. The results also show changes in the system frequencies for different ice conditions.

Keywords: Ice-structure interaction, SSI-cov, Automated parameter identification, Hierarchical clustering.

NOMENCLATURE

μ continuous time eigenvalue.
λ eigenvalue.
ϕ eigenvector.
f frequency.
ξ damping coefficient.
φ mode shape.
ω frequency in radian.
σ normalized standard deviation.
S_f variance of frequency.
S_ξ variance of damping.
S_{MAC} variance of MAC.

INTRODUCTION

The action of drifting ice may induce vibrations in offshore structures, posing a threat to the structural integrity. It is important to understand the system characteristics during ice-structure interaction for the operational ability of offshore structures. The presence of ice surrounding a structure may alter the system properties, such as natural frequencies, damping ratios, and mode shapes. Identifying the modal parameters under different ice conditions may therefore give insight into how the ice actions affect modal properties, and in turn provide uncertainty bounds to each parameter. Nord et al. further showed that for some ambient interaction types it was difficult to identify the system proper-

ties, and showed that there is a need for a controlled environment assessment of when to expect a successful modal parameter identification [1].

In order to avoid bias from the analyst to the identified parameters, the system identification should ideally be performed without too many user interactions. Unfortunately, traditional methods involve many user interactions, which results in large computational cost [2] and bias to the results. Therefore, it is necessary to develop an analysis method for automatic modal parameters identification.

In what follows, it is assumed that for a limited time window of ice-structure interaction, the process can be described by a linear time-invariant system. To obtain the structural properties, a covariance-driven stochastic subspace identification (SSI) algorithm is applied to estimate modal parameters. All identified modal parameters are afflicted with statistical uncertainty because of the finite number of data samples, undefined measurement noises, non-stationary excitation, etc. [3]. Hence, a covariance-driven SSI (SSI-cov) algorithm was proposed to estimate the frequencies, damping ratios and their uncertainties [4].

After SSI-cov analysis, poles at different system orders are obtained. A pole is considered stable if the deviances in frequency, damping and normalized standard deviation of the frequency fulfill the predefined stability criterion. After that, a stabilization diagram is constructed by stable poles via taking frequency as abscissa and system order as ordinate [5]. Physical modes should then show up as vertical lines in the diagram.

To date, there are many suggested methods to automatically determine the modal parameters. Magalhaes et al. applied hierarchical clustering to identify the modes successfully based on the data from concrete arch bridge [6]. Verboven et al. [7] and Vanlanduit et al. [8] employed fuzzy C-means clustering to classify the modes into two categories (physical and spurious). Reynders et al. introduced how to use hierarchical clustering to identify the physical modes based on single-mode validation criteria [2]. It does not require any user-specified parameter values. The validation example shows the hierarchical clustering has better robustness to identify modal parameters than the traditional identification approach. Inspired by this research, hierarchical clustering is used to identify the parameters of the ice-structure interaction model.

This study proposes a workflow of modal parameters identification which is made up of three parts: data preprocessing, SSI-cov analysis and physical mode identification. This analysis procedure could identify modal parameters with few users intervenes and achieve a better performance of parameters identification than the slack value-based identification method. The main contributions are shown as follows: 1) several validation experiments are carried out to choose proper parameters for the selection of stable poles in order to improve the accuracy of identified frequencies; 2) Hierarchical clustering is compared with slack value-based identification approach to estimate the param-

eters of ice-structure interaction model.

The rest of this paper is structured as follows: the next section describes the procedure on modal parameters identification, including data preprocessing, SSI-cov analysis, and physical mode identification. Case study compares two cases regarding optimal parameters selection and makes a comparison between the slack value and hierarchical clustering. Discussion and Conclusion are given finally.

Modal parameters identification procedure

This section introduces the main procedure of mode analysis. As shown in Fig. 1, The procedure includes three parts: data preprocessing, SSI-cov analysis, physical mode identification. Data preprocessing is to process the collected sensor data. Next, the processed data is analyzed by SSI-cov algorithm. Finally, physical modes could be clustered by the proposed algorithm.

Ice-structure interaction model testing

The ice-structure interaction model tests were carried out in the Hamburg Ship Model Basin's (HSVA) large ice model basin [1]. The setup consists of a flexible foundation with adjustable mass and stiffness to mimic certain dynamic characteristics of the structure and a rigid model. The flexible foundation was designed to have one or two natural frequencies in ice drift direction (21.36 and 29.53 rad/s). A cylindrical model (red) with a 500 mm diameter was used for the tests considered for the presented study. This model was equipped with tactile sensors to monitor local ice loads. Additionally, global loads were recorded by a 6-component load scale connecting the compliant basis and the model, and lasers and accelerometers monitored the ice-induced vibrations of the structure in x- and y-direction (loading direction and perpendicular in-plane motion). The setup is described in detail in [9]. The setup was instrumented by three Triax accelerometers to measure the structural response over different ice velocities as shown in Fig. 1. The data was obtained under different structural and ice related properties. These include the SDOF (one natural frequency) and MDOF (two natural frequencies) setup, both tested in two different ice thicknesses with constant compressive strength, and in two different ice types: standard model ice, and an alternative model ice type developed for crushing failure. Hence, eight ice-structure property combinations have been investigated. The full data set is described by Stange et al. [9]. Run 32010 investigated in the presented analysis was conducted in 41 mm thick standard model ice. HSVA's standard model ice is frozen from a 0.7% sodium chloride solution using a spraying technique which creates a fine grained fresh water ice top layer. Subsequently, the ice grows in the natural way with primarily columnar structure. During growth, air
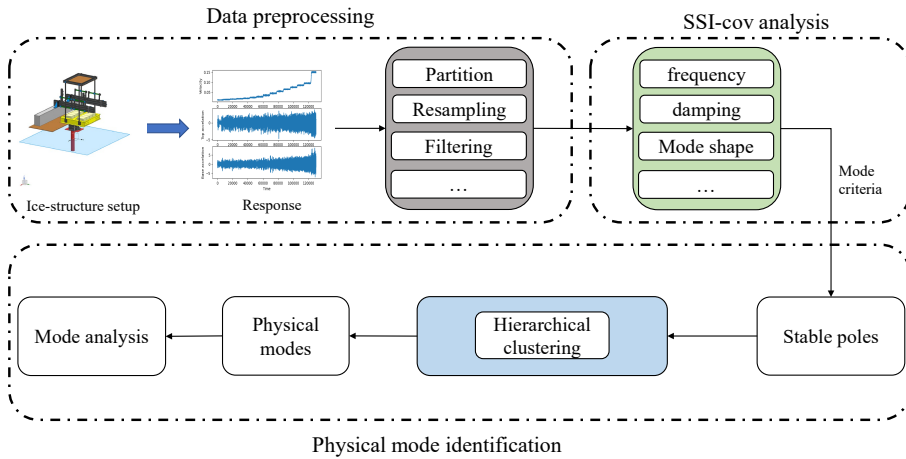
----

[1] https://www.hsva.de/

Data preprocessing

SSI-cov analysis

Physical mode identification

FIGURE 1: The procedure of automated modal parameters identification.

is embedded into the growing ice sheet to adjust its density, increase its brittleness, and give the ice a white appearance. When the target thickness has been reached, cooling is switched off, and heat is released into the ice tank room. Consequently, the ice is weakened until the target strength is reached. Detailed information on HSVA's standard model ice, as well as the alternative ice type mentioned above, is given by Ziemer et al. [10].

All eight test runs contain several different ice drift speeds from 4 to 150 mm/s. Therefore, for test data analysis the measurements are subdivided into segments with constant velocity first. Second, segments are grouped for different ice failure types (intermittent crushing (IC), frequency lock-in (FLI), continuous crushing (CC)). The ice failure types strongly affect the dynamic response of the structure: In IC, the ice load is sawtooth-shaped with irregular loading periods which are much longer than the natural periods of the structure. The model, therefore, follows the ice load in a quasi-static manner. IC occurs at low ice speed. When the speed increases, the interaction changes to FLI. This failure mode is characterized by quasi-synchronized local ice failures that cause large oscillation amplitudes in a frequency close to the natural frequency of the structure. As the ice drift speed increases further, the failure mode changes to CC and creates an irregular, broadband excitation. After subdividing the data, it is resampled with 100Hz. Finally, a high pass filter whose cutoff is 0.2 HZ is used to remove the noise from the data. The

processed data is used as the input of SSI-cov algorithm.

Covariance-driven stochastic subspace identification algorithm

The linear time-invariant system is described by a discrete time stat-space model

$$\begin{cases} x_{k+1} = Ax_k + w_k \\ y_k = Cx_k + v_k \end{cases} \quad (1)$$

where w_k and v_k are the process and output noise, respectively. In order to identify matrices A and C from which the modal frequencies, damping and mode shapes can be obtained, the eigenvalues and eigenvectors of the system in Eqn.1 is calculated by the following equations

$$\begin{cases} (A - \lambda_i I)\phi_i = 0 \\ \varphi_i = C\phi_i \end{cases} \quad (2)$$

from which the μ_i, f_i, and ξ_i can be obtained:

$$\mu_i = \frac{ln\lambda_i}{T}, f_i = \frac{|\mu_i|}{2\pi}, \xi_i = -100\frac{\Re(\mu_i)}{|\mu_i|} \quad (3)$$

where T is the sampling period. SSI is a prevalent method to estimate the matrices A and C. The algorithm uses the output data to build a subspace matrix $H_{p+1,q} \in \mathbb{R}^{(p+1)r \times qr_0}$. Therein, r is the number of sensors, r_0 is the number of reference sensors, and p and q are the parameters chosen such that $pr \geq qr_0 \geq n$, where n is the model order. The subspace matrix $H_{p+1,q}$ can be truncated at a user-defined model order n via singular value decomposition (SVD)

$$H_{p+1,q} = \begin{bmatrix} U_1 & U_0 \end{bmatrix} \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_0 \end{bmatrix} \begin{bmatrix} V_1^T \\ V_0^T \end{bmatrix} \quad (4)$$

and

$$O_{p+1} = U_1 \Sigma_1^{1/2} \quad (5)$$

The C matrix can be directly extracted from the first block of r rows of the observability matrix O_{p+1}, while the A matrix can be obtained from a least-squares solution of

$$O_{p+1}^{\uparrow} A = O_{p+1}^{\downarrow} \quad (6)$$

where $O_{p+1}^{\uparrow} = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{p-1} \end{bmatrix}, O_{p+1}^{\downarrow} = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{p} \end{bmatrix}$

The principle of SSI-cov is to propagate the covariance of the subspace matrix, Σ_H, to the modal parameters through first-order perturbations. The covariance of the modal parameters are obtained as

$$cov\left(\begin{bmatrix} f_i \\ \xi_i \end{bmatrix}, \begin{bmatrix} f_j \\ \xi_j \end{bmatrix}\right) = \begin{bmatrix} J_{f_i,A} & 0_{1,rn} \\ J_{\xi_i,A} & 0_{1,rn} \end{bmatrix} \Sigma_{AC} \begin{bmatrix} J_{f_i,A} & 0_{1,rn} \\ J_{\xi_i,A} & 0_{1,rn} \end{bmatrix}^T$$
$$cov\left(\begin{bmatrix} \Re(\phi_i) \\ \Im(\phi_i) \end{bmatrix}, \begin{bmatrix} \Re(\phi_j) \\ \Im(\phi_j) \end{bmatrix}\right) = \begin{bmatrix} \Re(J_{\phi_i,A,C}) \\ \Im(J_{\phi_i,A,C}) \end{bmatrix} \Sigma_{AC} \begin{bmatrix} \Re(J_{\phi_i,A,C}) \\ \Im(J_{\phi_i,A,C}) \end{bmatrix}^T \quad (7)$$

The detailed computational process can be referred to [4] After SSI-cov analysis, the modal parameters are derived. Next, the mode stability criterion is employed to pick stable poles. The selected stable poles are further analyzed to obtain physical modes in the following step.

Physical mode identification Once poles that are stable/unstable are identified, one must group poles with similar modal characteristics. This is commonly performed in a stabilization diagram, which shows the frequency of the poles on the horizontal axis and the order of the system on the vertical axis. A
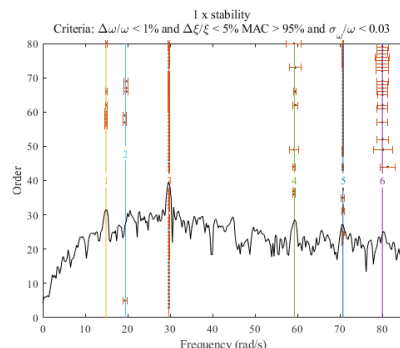
(a) Two accelerations
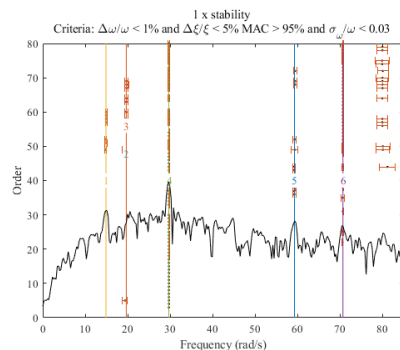
(b) Three accelerations

FIGURE 2: Mode frequency under the different number of structural response signals.

physical mode appears as a straight vertical line of poles, and the line with the corresponding lowest frequency is the first eigenfrequency, the column with the corresponding second lowest frequency is the second natural frequency, and so on. Poles that are not stacked on a vertical line are usually what is referred to as spurious poles/modes, i.e. modes without physical interpretation. Once one has determined which poles that should be counted as part of one column, it is common to compute the average value of these poles, from which we find the corresponding natural frequency, damping and mode shape. The major challenge lies in the process of choosing the poles that should be counted as part of the column of poles (mode), due to the fact that some lie at a slightly different frequency, have different damping values or mode shape, and different corresponding uncertainties. Therefore different techniques have emerged to handle the physical mode selection, where clustering algorithms have been suggested

TABLE 1: THE EXPERIMENT SETTINGS FOR ICE-STRUCTURE INTERACTION ANALYSIS.

Test	Model	Ice type	Ice speed	Ice thickness	Flexural strength
32010	9500MDOF	Model ice	4-150mm/s	41mm	56kPa

(a) rb=100

(b) rb=150

(c) rb=200

FIGURE 3: The comparison of identified modes for different choices of blockrows.

(a) nb=10

(b) nb=20

(c) nb=30

FIGURE 4: The comparison of identified modes among different number of blocks.

as an efficient technique to determine the physical modes. One of the popular methods is Hierarchical clustering [2, 11]. Hierarchical clustering is a recursive partitioning of a dataset into successively smaller clusters. The input is a weighted graph whose edge weights represent pairwise similarities or dissimilarities between data points. Hierarchical clustering is represented by a rooted tree where each leaf represents a data point and each internal node represents a cluster containing its descendant leaves. The tree is constructed based on the distance information between different data points [12]. It is suitable for the data set with arbitrary shapes and attributes of arbitrary type. And the hierarchical

relationship among clusters is easily detected, and relatively high scalability in general [13].

Let $Q = q_1, q_2, ..., q_n$ be a set of objects. The dendrogram is constructed by the following steps [11]:

1) Compute the proximity matrix containing the distance between each pair of objects (q_i, q_j).
2) Group the objects into a hierarchical cluster tree using the distance information.
3) Choose the cut off value to partition the hierarchical tree into clusters.

Copyright © 2022 by ASME

TABLE 2: The identified frequencies by slack value and hierarchical clustering based on the data in '32010' under the ice failures of IC, FLI, and CC.

Data file	Frequency	Method	ice velocity (mm/s)														
			4	6	8	10	12	14	16	18	20	28	45	65	80	95	150
			IC	IC	FLI	FLI	FLI	FLI	FLI	FLI	FLI	FLI	FLI	CC	CC	CC	CC
32010	First	Slack value	70.80	14.86	**20.25**	29.76	**21.02**	**21.07**	30.71	**21.06**	**21.19**	**21.27**	**21.57**	**21.38**	**21.50**	**21.38**	**21.61**
		Hierarchical	70.80	14.86	**20.24**	20.99	**21.02**	**21.07**	21.03	**21.06**	**21.19**	**21.27**	**21.57**	21.43	21.51	**21.50**	**21.61**
	Second	Slack value	119.20	**29.80**	61.01	41.85	**30.80**	28.39	31.75	29.05	28.55	29.79	30.29	30.13	30.14	21.41	26.83
		Hierarchical	120.38	**29.80**	24.90	29.76	**30.80**	28.38	31.32	29.17	28.55	29.79	30.30	30.08	30.14	25.96	26.83
	Third	Slack value	122.92	36.88	80.66	70.58	42.01	**31.30**	31.79	29.22	**31.61**	42.57	59.68	61.41	59.33	25.94	**30.39**
		Hierarchical	152.26	36.88	**29.95**	41.86	42.01	**31.26**	42.02	42.14	**31.61**	42.57	53.21	61.35	49.06	**30.08**	**30.39**

TABLE 3: The identified damping by slack value and hierarchical clustering based on the data in '32010' under the ice failures of IC, FLI, and CC.

Data file	Damping (%)	Method	ice velocity (mm/s)														
			4	6	8	10	12	14	16	18	20	28	45	65	80	95	150
			IC	IC	FLI	FLI	FLI	FLI	FLI	FLI	FLI	FLI	FLI	CC	CC	CC	CC
32010	First	Slack value	NULL	NULL	0.11	NULL	0.08	0.01	NULL	0.08	0.03	0.02	0.64	2.44	1.67	1.64	2.47
		Hierarchical	NULL	NULL	0.13	0.05	0.08	0.01	0.08	0.08	0.03	0.02	0.64	2.35	1.73	1.51	2.47
	Second	Slack value	NULL	0.54	0.65	2.57	3.23	1.82	2.69	3.69	1.89	2.53	0.92	0.79	1.17	1.02	0.37
		Hierarchical	NULL	0.54	1.02	2.57	3.23	1.57	3.26	3.58	1.89	2.51	0.91	0.82	1.17	1.09	0.38

In this study, eigenfrequency difference and MAC are used as distance measures in [14]. Its form is shown in Eq. (8)

$$d(k,l) = |f_k - f_l| + (1 - MAC(\phi_k, \phi_l)) \qquad (8)$$

where f_k is the eigenfrequency of mode k; MAC is computed by Eq. (9)

$$MAC(\phi_k, \phi_l) = \frac{|\phi_k^T \phi_l|^2}{||\phi_k||_2^2 ||\phi_l||_2^2}, ||\phi_k||_2^2, ||\phi_l||_2^2 \neq 0 \qquad (9)$$

where ϕ_k is the mode shape of mode k.

Through continuous iterations of evaluating the paired distance, the data points that are smaller than the cutoff value are partitioned into the same cluster. Finally, hierarchical clustering yields a set of similar mode sets from the cleared stabilization diagram. After that, the identified physical modes are evaluated and analyzed further based on natural frequencies and damping ratios.

Case study

This section mainly introduces two parts of the experiments. The first part is to pick up the optimal parameters for SSI-cov analysis. Next, hierarchical clustering is compared with a slack value-based approach in [1] to examine the efficiency of the automated modal analysis.

Parameters selection

The SSI-cov algorithm involves user interaction to choose a couple of parameters that need to be selected. For example, there are three accelerometers to measure the acceleration of the structure. However, not all measured signals contribute to accurate parameters identification. In addition, the number of blocks (nb) of output data matrices, as well as the number of blockrows (rb), have influences to some extent. Other parameters such as sampling frequency, system orders, could affect the identified results, which are not our main concern in this study. Their settings can be referred to [1].

In order to select proper parameters for modal parameters identification, Test 32010 is used as a case and its corresponding

Copyright © 2022 by ASME

settings are shown in Table 1. The test had a stepwise increasing ice speed, and therefore different regimes of ice-structure interaction would take place during the total time span. The data was cut to begin when the ice speed reached 4 mm/s. Thereafter the data was analyzed in time windows that each consisted of 2000 data points. The choice of data points was selected to have the sufficient number of data points for the SSI-cov algorithm to render consistent results, and few enough data points for the interaction regime to significantly change. Next, tolerance deviances to frequency, damping, and MAC-values, as well as the normalized standard deviation of the frequency ($\hat{\sigma}_{\omega_i}/\omega_i$) are leveraged to pick up stable poles. Based on [1], a pole at order n was considered stable if the deviances in frequency, damping ratio, and normalized standard deviation of the frequency between a pole at order n and n-1 were less than 0.01, 0.05 and 0.05, respectively, and corresponding MAC-values exceeded 0.95. After that, S_f, S_ξ and S_{MAC} are chosen to be 0.02, 0.3 and 0.5 respectively to select eigenmodes. Figure. 2(a) and Figure. 2(b) show the identified modes under the case of different accelerations. Therein, the black curve is the power spectrum whose peaks represent the possible physical modes. The stable poles with variance are plotted following their order. The formed straight lines represent the identified frequencies by stable poles. In Fig. 2(b), the identified second and third frequencies are overlapped while Fig. 2(a) presents a better identification result. Therefore, this study uses two accelerations as input data for SSI-cov analysis.

Figure. 3(a), Figure. 3(b), and Figure. 3(c) display the identified eigenfrequencies and their estimated standard deviations in the cases of 'rb = 100', 'rb = 150', 'rb = 200' separately. When rb equals to 150 or 200, there are more spurious modes that have larger standard deviations, as shown in Fig. 3(b) and Fig. 3(c). Compared with them, 'rb = 100' could obtain more accurate results which are in line with the position of the peaks of the power spectrum, as shown in Fig. 3(a). Hence, the study prefers 100 as blockrows. Figure. 4(a), Figure. 4(b), and Figure. 4(c) display the identified eigenfrequencies and their estimated standard deviations in the cases of 'nb = 10', 'nb = 20', 'nb = 30', separately. Through the comparison among these three figures, it is easy to find that 'nb = 20' could obtain better results as frequency (20) disappeared in other stability diagrams. For this reason, the number of blocks is selected as 20 in this study.

Comparison between slack value and hierarchical clustering

This section compares the slack value-based parameters identification approach with hierarchical clustering. The data whose ice velocities are 8 mm/s and 95mm/s in datafile '32010' is chosen as two cases to compare these two methods. The benchmark values of the first two eigenfrequencies are 21.352 and 29.516 rad/s separately, which are estimated when the structure was moving in the open water [9]. Considering the uncertain
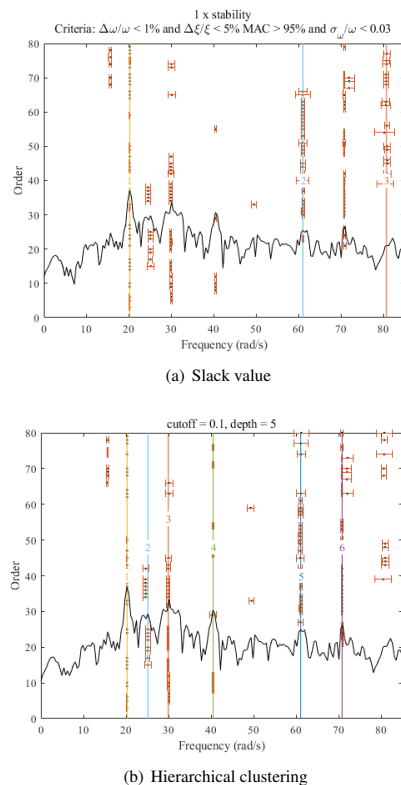

(a) Slack value

(b) Hierarchical clustering

FIGURE 5: The comparison of slack value and hierarchical clustering when ice velocity is 8 mm/s.

factors, the benchmark values are expanded by 10% deviation to an interval: [19.22, 23.49] for the first frequency, [26.56, 32.47] for the second frequency. The cutoff and depth of the Hierarchical clustering algorithm are chosen as 0.1 and 5 separately based on data characteristics.

The focus of this study falls on the first three natural frequencies that represent the most concerning modes. The ice velocities range from 4-150 mm/s. The identified frequencies are shown in Table. 2. The bold numbers represent the successful identifications of natural frequencies by two methods. For IC, the identified first frequencies by slack value and Hierarchical clustering are 70.8, 14.86, separately for different ice speeds. Results show that both methods fail to identify the first frequency. For FLI and CC, the first two natural frequencies identified by Hierarchical clustering are around 21 rad/s and 29 rad/s among different ice speeds. From this table, Hierarchical clustering renders more

Copyright © 2022 by ASME

bold numbers than the slack value-based approach. Table. 3 shows the identified damping of the first two modes. 'NULL' means the corresponding damping can not be obtained due to the failure of parameters identification. It can be seen that both methods achieve similar results. For 'FLI', the damping of the first mode is quite lower than that of the second mode, whereas it is opposite for other cases. This trend probably results from the increase of ice velocity.

Figure. 5(a) and Figure. 5(b) show the identified frequencies by these two methods in the case of ice velocity being 8 mm/s. Based on the referenced values of the first two natural frequencies, it is easily found that hierarchical clustering can identify these two frequencies correctly while slack value can not. As shown in Fig. 5(a), some missing modes are supposed to be identified in the stabilization diagram. For example, at the position of frequency 30 rad/s, there is supposed to be a mode that appears on the peak of the power spectrum. Based on aforementioned analyses, it is concluded that hierarchical clustering outperforms the slack value-based approach as a whole.

Discussion

The section above introduced the Hierarchical clustering approach to identify the modal parameters of the structure when encountering ice-structure interaction. As shown in Table. 2, it often fails to identify the correct modal parameters for certain cases like IC failure and FLI at low ice velocities. For IC failure, likely the ice-structure interaction system is too time-variant and too nonlinear for the current method to identify the structural parameters inherently hidden in the measured signals. This phenomenon does probably depend on the severity of the ice-actions compared to the mass and stiffness of the structure. However, given that for a certain structure ice-actions are rare and operational parameters are to be extracted automatically as part of a structural health monitoring system at daily or hourly intervals, our results show that hierarchical clustering did have a better performance of parameters identification than the traditional slack-value method.

Another limitation lies in that input parameters impact results. For example, the change of nb and rb turns out to be different identification results. In other words, the uncertainty of input parameters would affect the accuracy of parameters identification. For the convenience of analysis, in this study, the limited numbers are compared based on previous research to obtain a relatively accurate result.

Conclusion

This study introduced a Hierarchical clustering method to automatically identify the parameters of the ice-structure interaction model. The proposed analysis workflow is shown in Fig. 1, including data preprocessing, SSI-cov analysis, modal

parameters identification. In order to verify the superiority of the proposed method, the slack value-based parameter identification method is leveraged to make a comparison based on data file 32010. First of all, parameters such as rb and nb are selected based on contrast tests. Next, hierarchical clustering and slack value are compared under the difference ice velocities from 4 - 150 mm/s. The results show hierarchical clustering outperforms slack value in terms of the accuracy of parameters identification for the ice failures of 'FLI' and 'CC'.

Accurate parameter identification is pivotal to the operational ability of offshore structures. As the second limitation in the Discussion, however, it is hard to obtain an accurate result due to parameters uncertainty. Therefore, it is necessary to quantify the uncertainty of input parameters from the perspective of statistics to implement a more accurate estimation.

REFERENCES

[1] Nord, T. S., Petersen, Ø. W., and Hendrikse, H., 2019. "Stochastic subspace identification of modal parameters during ice–structure interaction". *Philosophical Transactions of the Royal Society A,* **377**(2155), p. 20190030.

[2] Reynders, E., Houbrechts, J., and De Roeck, G., 2012. "Fully automated (operational) modal analysis". *Mechanical Systems and Signal Processing,* **29**, pp. 228–250.

[3] Döhler, M., Lam, X.-B., and Mevel, L., 2013. "Uncertainty quantification for modal parameters from stochastic subspace identification on multi-setup measurements". *Mechanical Systems and Signal Processing,* **36**(2), pp. 562–581.

[4] Döhler, M., and Mevel, L., 2013. "Efficient multi-order uncertainty computation for stochastic subspace identification". *Mechanical Systems and Signal Processing,* **38**(2), pp. 346–366.

[5] Ye, C., and Zhao, X., 2020. "Automated operational modal analysis based on dbscan clustering". In 2020 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS), IEEE, pp. 864–869.

[6] Magalhaes, F., Cunha, A., and Caetano, E., 2009. "Online automatic identification of the modal parameters of a long

span arch bridge". *Mechanical Systems and Signal Processing, 23*(2), pp. 316–329.

[7] Verboven, P., Parloo, E., Guillaume, P., and Van Overmeire, M., 2002. "Autonomous structural health monitoring—part i: modal parameter estimation and tracking". *Mechanical Systems and Signal Processing, 16*(4), pp. 637–657.

[8] Vanlanduit, S., Verboven, P., Guillaume, P., and Schoukens, J., 2003. "An automatic frequency domain modal parameter estimation algorithm". *Journal of Sound and Vibration, 265*(3), pp. 647–661.

[9] Stange, T., Ziemer, G., and von Bock und Polach, R. U. F., 2020. "Development of an experimental setup to investigate the impact of higher structural modes on dynamic ice-structure interaction". In 25th IAHR International Symposium on Ice 2020.

[10] Ziemer, G., Stange, T., and Hisette, Q., 2022. "Hsva model ice - a status report". In Proceedings of the ASME 2022 41st International Conference on Ocean, Offshore and Arctic Engineering OMAE2022.

[11] Mao, J.-X., Wang, H., Fu, Y.-G., and Spencer Jr, B. F., 2019. "Automated modal identification using principal component and cluster analysis: Application to a long-span cable-stayed bridge". *Structural Control and Health Monitoring, 26*(10), p. e2430.

[12] Cohen-Addad, V., Kanade, V., Mallmann-Trenn, F., and Mathieu, C., 2019. "Hierarchical clustering: Objective functions and algorithms". *Journal of the ACM (JACM), 66*(4), pp. 1–42.

[13] Xu, D., and Tian, Y., 2015. "A comprehensive survey of clustering algorithms". *Annals of Data Science, 2*(2), pp. 165–193.

[14] Pappa, R. S., James III, G. H., and Zimmerman, D. C., 1998. "Autonomous modal identification of the space shuttle tail rudder". *Journal of Spacecraft and Rockets, 35*(2), pp. 163–169.

Paper IV

Towards uncertainty and sensitivity analysis for modal parameters identification during ice–structure interaction

Chunlin Wang [a,*], Torodd Skjerve Nord [a], Gesa Ziemer [b], Guoyuan Li [a]

[a] *Department of Ocean Operations and Civil Engineering, Norwegian University of Science and Technology, 6009 Ålesund, Norway*
[b] *Institute for Maritime Energy Systems, DLR, Geesthacht, Germany*

A R T I C L E   I N F O

A B S T R A C T

The integrity of offshore structures is prone to the threat of drifting sea ice. By exciting the natural frequencies of the ice–structure interaction system, severe vibrations may occur. For a better understanding of how dynamic ice–structure interaction affects the system properties, a stochastic subspace system identification technique, SSI-cov is introduced to identify modal parameters during dynamic ice–structure interactions. Due to the uncertain input variables of SSI-cov, the identified modal parameters suffer from the quantitative judgment of the analyst. To address this problem, this study proposes an uncertainty analysis (UA) framework to obtain estimates of modal parameters. This framework constitutes both a sensitivity analysis (SA) and a UA. First, SA is applied to pick up the input variables that contribute the most to the identified modal parameters. Next, the important variables are left free to vary over their range of existence to obtain the modal parameters' uncertainties. The framework is applied to full-scale tests from Norströmsgrund lighthouse and model tests carried out in Hamburg Ship Model Basin (HSVA). The comparison with traditional automated modal parameter identification shows its superiority in terms of the accuracy of modal parameters.

1. Introduction

The integrity of a structure located in ice-infested waters is prone to the threats of ice forces and ice-induced vibrations. Substantial efforts were made to understand both the physical processes leading to ice-induced vibrations (IIVs), based on observations, measurement campaigns on lighthouses, bridge piers, and oil platforms for decades, but also model scale tests in ice basins (Nord et al., 2019). Drifting ice may cause various modes of ice–structure interaction, in which some lead to more severe vibrations than others, such as for instance crushing failure mode leading to frequency lock-in vibrations (FLI). This mode is particularly violent with respect to ice forces and structural responses, some examples in full scale with threats of the structural integrity (Jefferies, 1988; Blenkarn, 1970). IIVs are often associated with crushing failure, and IIVs are also shown to pose a threat to the structural integrity as well (Blenkarn, 1970).

While much attention naturally has been made to come up with phenomenological models to predict IIVs, a very limited number of studies were made to seek fundamental system parameters of the underlying mechanical system by means of system identification. For structures exposed to wind and waves, system identification has provided unique insight into how the system properties change for different properties of the wind or wave characteristics. For ice–structure interaction systems,

some work on model-scale experiments was conducted by Singh et al. (1990), while on Full-scale, Nord et al. applied SSI-cov to identify modal parameters on the Norströmsgrund lighthouse under a variety of ice–structure interaction modes (Nord et al., 2019). Level ice acting on a vertically-sided structure may fail to generate different failure modes depending on indentation velocity, aspect ratio, and ice properties (Timco, 1991). In Fig. 1, crushing failure Fig. 1(a) represents the deformation and failure of ice at high indentation speeds and low aspect ratios. It results from the non-simultaneous occurrence of high-pressure zones across the ice–structure interface (Nord et al., 2017). The crushing failure mode is mostly associated with IIVs and the main ice failure mode in our experiment (Hendrikse and Nord, 2019). Other failure modes, such as bending failure Fig. 1(b), splitting failure Fig. 1(c), buckling failure Fig. 1(d), and pushing floes Fig. 1(e) are not considered in this paper and can be referred to Kärnä and Jochmann (2003). The understanding of how ice failure modes change the modal parameters of the structure is beneficial to the structural safe design. However, it is still challenging in the research field for safe design and accurate predictions in fatigue assessments.

To this end, Operational Modal Analysis (OMA) was proposed to estimate the modal parameters from measurements of the vibration response only (Rainieri and Fabbrocino, 2014). It can be implemented

---

* Corresponding author.
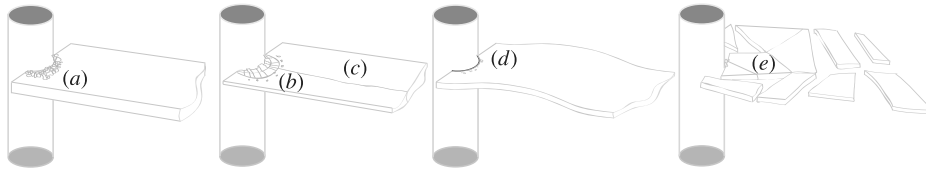 *E-mail address:* chunliw@ntnu.no (C. Wang).

Fig. 1. The different types of failure modes in ice–structure interaction (Nord et al., 2017).

efficiently, economically, and safely, and does not interfere with the normal use of the structure. The identified modal parameters are representative of the actual behavior of the structure since it takes advantage of natural excitation instead of artificial excitation. The most popular OMA method is the Stochastic Subspace Identification (SSI) method since it offers high accuracy in the identification of closely-spaced modes and is especially suited to be automated (He et al., 2021). SSI is to identify physical modes by picking up stable poles in a stability diagram. Since the estimated modal parameters are afflicted with statistical uncertainty, an efficient multi-order uncertainty computation method, called covariance-based SSI (SSI-cov), was proposed to estimate the variance of the identified modal parameters (Döhler and Mevel, 2013). Due to its low computational cost and high identification accuracy, it has been widely applied to analyze the parameters of structures in recent years.

Although SSI-cov can effectively remove modes with large variances to obtain physical modes, its accuracy of modal parameter identification cannot be guaranteed due to various assumptions and uncertain input parameters. For example, ice forces usually cannot be represented as Gaussian white noise and therefore the input violates the random white noise assumption of SSI-cov. In addition, the structure is described as a linear time-invariant system. Such assumptions could result in errors in the modal parameters identification. Besides that, SSI-cov contains many uncertain input parameters such as slack values, stability criterion, system orders, etc. The parameters' uncertainty is propagated to the identified parameters and results in the poor estimation of their values. To reduce the uncertainty of slack values, clustering technologies were introduced for automated OMA (Reynders et al., 2012). Clustering approaches can implement parameter identification in an automatic way and avoid the artificial selection of slack values. Nevertheless, clustering algorithms will bring additional uncertainties as they contain several uncertain parameters and algorithm structure uncertainty (known as model uncertainty). As a consequence, a robust outlier detection was proposed to reduce statistical uncertainty caused by the clustering algorithm (Zeng and Hoon Kim, 2021). However, it cannot remove the algorithm uncertainty completely. In addition, clustering technologies did not consider the uncertainty from other input parameters of SSI-cov. Therefore, it is significant to investigate how the identified parameters' values vary under the different sources of input variables' uncertainty, which could ensure more convincing parameter estimation for engineers.

To improve the accuracy of SSI-cov identification, this research proposed a modal parameters analysis framework based on UA and sensitivity analysis (SA). SA aims to pick up those factors that account for the most contributions to the model output. That is beneficial to lower the burden of UA caused by large amounts of variables. The main contributions are as follows: (1) SA is introduced to remove unimportant input parameters of SSI-cov; (2) Sihouette and robust outlier detection are used to improve the robustness of clustering algorithm for automated OMA; (3) UA is applied on several ice–structure interaction datasets for precise parameters identification. The analysis results could provide support for structural health monitoring (SHM) of offshore structures.

The whole paper is structured as follows. Section 2 presents related work regarding automated OMA approaches and UA in other fields. Section 3 gives an introduction on the proposed analysis framework and corresponding algorithms. In the next section, several case studies are shown to verify the feasibility of the proposed method.

2. Related work

Safe design of the deployed offshore structures requires understanding the ice-induced vibrations as a result of level ice interacting with a vertically-sided structure (Kärnä et al., 2013). To obtain knowledge of the dynamic behavior of the structure, OMA is introduced to analyze structure modes, each one characterized by a set of parameters (natural frequency, damping ratio, model shape). As uncertainty is intrinsic in various identification algorithms, it is of importance to evaluate the uncertainty in the identified parameters. This section mainly introduces the development of OMA and uncertainty evaluation and explores how to apply UA for better OMA of offshore structures.

2.1. Operational modal analysis

Before the wide application of OMA, experimental modal analysis (EMA) played a pivotal role in parameter identification. However, EMA techniques cannot be applied to analyze massive structures and a system in operational condition because the power of the device is insufficient to excite the structure to attain the required magnitude (Lauwagie et al., 2006). For this reason, OMA was developed to estimate modal parameters based on the data collected when the structure is under the operational conditions (Zahid et al., 2020). The identified parameters can be representative of the actual behavior of the structure.

SHM requires real-time parameters analysis to perform fault detection. Unfortunately, much user intervention on modal parameters analysis is an obstacle in a real application (Zeng and Hoon Kim, 2021). In recent years, growing attention has been put on automated OMA. Many system identification algorithms have been developed to identify system modes. Yang et al. developed an automated OMA based on an eigensystem realization algorithm and a two-stage clustering strategy. It can estimate modal parameters effectively in real-time (Yang et al., 2019). Magalhaes et al. applied the poly-Least Squares Complex Frequency Domain method to perform online parameters identification (Magalhaes et al., 2009). Au et al. combined the Bayesian method with a fast Fourier transform of ambient data (Au, 2011). It can meet the requirement of real-time parameters analysis as Bayesian modal identification can be performed in a few seconds. Reynders et al. introduced uncertainty bounds on the modal parameters estimated with SSI (Reynders et al., 2008). The proposed method can estimate the variance of modes from a single measurement record. Due to the time-consuming variance estimation of parameters, Döhler et al. proposed an efficient multi-order uncertainty computation approach (SSI-cov) (Döhler and Mevel, 2013). Among these approaches, SSI-cov is the most popular method since it offers high accuracy in the identification of closely-spaced modes and is especially suitable to be automated (He et al., 2021). Therefore, it has been widely used for model parameters identification of ice–structure interaction in Nord et al. (2019), flexible spacecraft in Xie et al. (2016), large-scale bridge in Pan et al. (2021), and so on.

After SSI analysis, poles at different orders are obtained and form a stabilization diagram. Next, the slack value is used to pick up the stable poles that are representative of the physical modes. The stable poles form multiple vertical lines in the diagram that are regarded as physical modes (Nord et al., 2019). Due to the limitation of many

parameters selection, several strategies have been proposed for automatic physical mode identification. Hierarchical clustering was proposed based on eigenfrequency difference and the modal assurance criteria (MAC) value as distance measures (Pappa et al., 1998). Afterwards, Goethals et al. incorporated the eigenfrequency and damping ratio difference into the Hierarchical clustering algorithm. Closely-spaced modes are grouped into the same cluster, they are then separated by the MAC value. In addition to Hierarchical clustering, K-means, and fuzzy C-means are employed to find physical modes as well (Mao et al., 2019; He et al., 2021). However, clustering results are sensitive to the user-defined number of clusters. To address this limitation, Density-based spatial clustering of applications with noise (DBSCAN) was proposed to cluster physical modes (Ye and Zhao, 2020; Li et al., 2020). In recent work, Kvaale et al. proposed hierarchical DBSCAN (HDBSCAN) for automated parameter identification since it requires fewer parameters (Kvåle and Øiseth, 2021).

Although the aforementioned clustering techniques achieve better performance on mode identification, they cannot remove estimating errors caused by parameter uncertainty. Therefore, it is necessary to make an uncertainty evaluation during automated OMA to obtain more convincing results.

2.2. Uncertainty analysis

In order to obtain accurate model parameters, SSI-cov uses stabilization criteria to remove part of bias of the model and the modes, and slack values to remove artificial modes with large variances (Reynders et al., 2008). However, it cannot remove or quantify other uncertainties such as the number of data samples and parameters' values. In addition, stabilization criteria and slack values bring more uncertain parameters.

To reduce the impact of slack values on stable poles selection, Hierarchical clustering is used to replace slack values-based method for clustering physical modes (Wang et al., 2022). However, uncertainties induced by other parameters are not considered. Therefore, this study introduces UA to quantify uncertainties caused by uncertain input parameters.

UA aims to provide confidence that model-based decisions are robust to underlying uncertainties. Uncertainties are mainly divided into two categories: aleatory uncertainty and epistemic uncertainty. Aleatory uncertainty arises from physical phenomena that are random by nature. It is extremely hard to quantify and eliminate aleatory uncertainty during modeling calibration. In contrast to aleatory uncertainty, epistemic uncertainty concern the model parameters and model discrepancy, also known as parametric uncertainty and structural uncertainty. Epistemic uncertainty can be reduced when better knowledge of the model structure and more accurate data become available (Xiao and Cinnella, 2019). Therefore, the main concern of this study is to investigate the epistemic uncertainty of quantities of interest (QoI).

Uncertainty quantification (UQ) aims to quantify the uncertainty of QoI that is induced by propagating uncertain random variables to the model output. There are many technologies that are developed for UQ, for instance, SA, Monte Carlo simulation, response surface approaches, evaluation of classical statistical confidence bounds, Dempster-Shafer theory, and Bayesian inference (Zhang et al., 2020). They are widely applied to various domains such as Deep Learning, medical health, environmental science, material science, modal parameter identification, etc. For example, Dusenberry et al. analyzed RNN model uncertainty for electronic health records (Dusenberry et al., 2020). Various Bayesian RNNs were introduced to place priors on different subsets of the parameters to determine the level of the model. Hu et al. presented the Markov Chain Monte Carlo (MCMC) algorithm to quantify the parameter uncertainty that arises from the modeling of macro-fiber composite materials (Hu et al., 2014). The study extended a data-driven deterministic estimation technique, presented in Hu et al. (2012) that was used to obtain the unknown model parameters, to investigate the parameter uncertainty that fits into non-Gaussian distributions. In order

to improve the accuracy of modal parameters identification, SSI-cov was proposed to estimate the variance of modal parameters (Reynders et al., 2008). Due to the high computational cost of SSI-cov, Döhler et al. delivered an efficient multi-order uncertainty computation for SSI-cov (Döhler and Mevel, 2013). This improved SSI-cov is both computationally and memory efficient. However, parametric uncertainty still exists as SSI-cov only estimates the structural uncertainties from SSI algorithm.

It is challenging for UQ since a large number of simulations are required. Specifically, SSI-cov needs to conduct singular-value decomposition to obtain system matrices, which is extremely time-consuming. Therefore, in this study, SA is introduced to reduce the burden of UQ. SA is to study how the variation of QoI is apportioned to the model inputs. After SA, those important variables can be picked up for the following UQ. The detailed description regarding SA can be found in review papers (Zhang et al., 2020; Saltelli et al., 2019; Razavi et al., 2021).

3. Method

This section is to introduce the proposed UA framework that is applied to identify modal parameters. Compared with traditional OMA, it can estimate the uncertainty bound of parameters induced by uncertain input variables. As shown in Fig. 2, the framework is mainly divided into three parts. The first part is to do a model evaluation based on the input parameters sampled from a hypothetical distribution. The second one is automatic parameter identification using the clustering method. The identified modal parameters are used for the following sensitivity analysis and uncertainty analysis. The workflow is first conducting SA to select the most important input variables, second doing UA based on the selected uncertain variables.

3.1. Model evaluation

Model evaluation is to make simulations based on experimental settings. It is the core and most time-consuming part of UA and SA. UQ involves determining the probability distributions of each input variable. Generally, the probability distribution function (PDF) is inferred based on experimental data or thumb rules. The PDF indicates the whole range of parameters' uncertainties. An accurate PDF estimation is extremely hard because of the limited data samples.

In general, all parameters are assumed to be independent and subject to certain distributions (uniform or normal). In order to represent the PDF, two hyperparameters (μ and σ) are introduced to control the uncertainty across all parameters (Pathmanathan et al., 2019). The PDF is defined as shown in Eqs. (1) and (2):

$$p \sim N(\mu, \sigma^2) \tag{1}$$

where 'N' denotes normal distribution.

$$p \sim U(\mu - \sigma, \mu + \sigma) \tag{2}$$

where 'U' represents uniform distribution.

The assumption of PDF has a significant influence on the identified results. If PDF is assumed a normal distribution, some data are sampled from the tail of the normal distribution. These data might exceed the reasonable range of the parameter's value, leading to an error in SSI-cov analysis. To avoid this problem, the uniform function is chosen as the distribution function of parameters since it can define the minimum and maximum value of the parameter's range. After the PDF is determined, the random samples are obtained by Latin-hypercube sampling (LHS). Compared with Monte Carlo sampling, LHS can sample data across the whole space of parameters, and hence it is effective in SA with small samples (Helton and Davis, 2003). The random samples and the measured structural response are fed into the SSI-cov to compute modal parameters. The simulations will be repeated by as many times as the
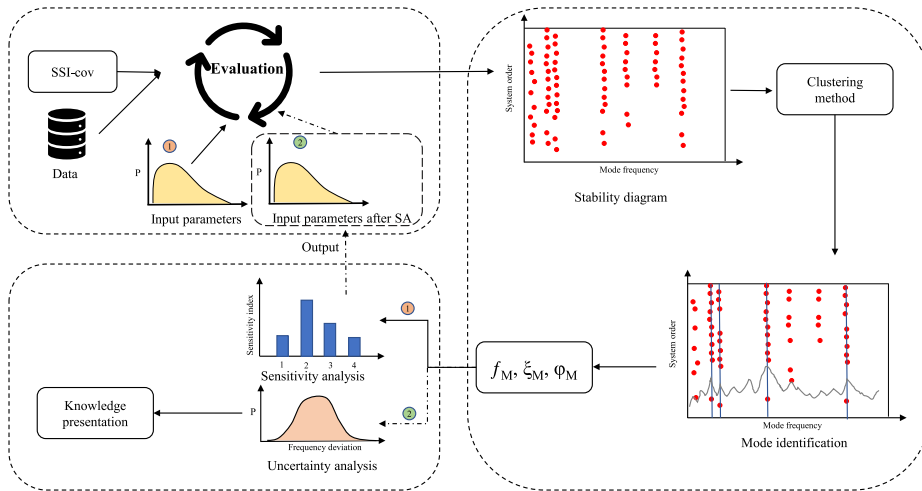
Fig. 2. The framework of uncertainty analysis and sensitivity analysis on modal parameters identification.
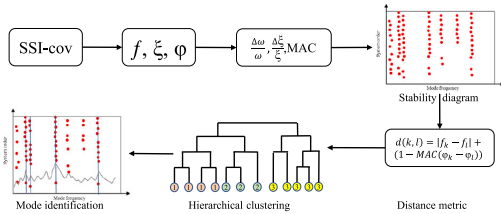

Fig. 3. The schematic diagram of mode identification using Hierarchical clustering.

size of the samples. Finally, the sensitivity of inputs on QoI is computed based on simulated results.

The model evaluation will be carried out two times. The first time is for SA to pick up the influential input factors and the second time is for UA when the selected factors are applied for SSI-cov to identify modal parameters under different ice failures and ice velocities.

3.2. Modal parameters identification

After model evaluation, the obtained parameters (eigenvalues and eigenvectors) are used to compute modal parameters ($f, \xi, and \varphi$). Next, modal parameter identification is carried out to obtain physical modes. The schematic diagram of modal parameter identification is shown in Fig. 3.

First of all, stable poles are picked up based on the stability criterion to form a stability diagram. The stability criterion involves the tolerance deviances to frequency ($\frac{\delta\omega}{\omega}$), damping ($\frac{\delta\xi}{\xi}$), and MAC-values, as well as the normalized standard deviation of the frequency ($\hat{\sigma}_{\omega_i}/\omega_i$). The pole is taken as stable if its stability parameters fulfill the pre-defined criterion.

After stable poles are obtained, a clustering algorithm is introduced to identify modes from these stable poles. In this study, a Hierarchical clustering algorithm is applied to identify modal parameters automatically. Compared with the slack value-based method, it has the advantage of higher accuracy and less human intervention (Wang et al., 2022). In addition, it has fewer parameters to determine, which can reduce the uncertainties of parameters. The distance metric in the Hierarchical clustering is selected as $d(k,l) = |f_k - f_l| + (1 - MAC(\varphi_k - \varphi_l))$. $d(k,l)$ represents the distance between mode 'k' and mode 'l'.

In order to eliminate uncertainties in the Hierarchical clustering algorithm, two strategies are employed to improve the robustness of the algorithm. The first method is to use the Silhouette value to evaluate the clustering results. It measures how similar a point is to points in its own cluster when compared to points in other clusters. The higher it is, the better the samples are clustered. The principle is shown in Eq. (3).

$$s(i) = \frac{b(i) - a(i)}{max(b(i), a(i))} \tag{3}$$

where b(i) denotes the average distance of point 'i' with all points in the closest cluster to its cluster; a(i) is the average distance of point i with other points in the same clusters; s(i) is the silhouette coefficient that ranges from [−1,1].

Due to the variability of modal estimates, outlier detection is used for penalizing undesirable modes in the final clusters to reduce identification uncertainties (Zeng and Hoon Kim, 2021). Eq. (4) defines the robust distance (RD).

$$RD(x) = d(x, \hat{\mu}_{MCD}, \hat{\mu}_{MCD}) \tag{4}$$

where MCD is the minimum covariance determinant that is used for outlying values detection. x represents frequency in this study; $\hat{\mu}_{MCD}$ denotes the MCD estimates of location; $\hat{\mu}_{MCD}$ is the covariance of MCD.

After that, a hierarchical tree could be created as shown in Fig. 3. The color of the leaves in the tree represents different clusters. If a cluster contains a pre-defined number of poles, the poles in this cluster render a physical mode as shown by the straight line in the stability diagram. After clustering, the stability diagram could show physical modes.

3.3. Uncertainty analysis and sensitivity analysis

Too many uncertain parameters lead to the high cost of UA. Therefore, SA is used to pick up the most sensitive parameters that contribute to the variation of the modal parameters. Global sensitivity analysis (GSA) is a powerful tool to estimate the total sensitivity of all factors. In this study, two popular GSA approaches are chosen to conduct SA. They are the variance-based Sobol method (Sobol, 2001) and cumulative distribution function (CDF)-based PAWN method (Pianosi and Wagener, 2015).

A generic model is described as follows.

$$Y = f(X_1, X_2, \ldots, X_M) \tag{5}$$

Table 1
The main test runs and the corresponding parameters settings.

Run	Model	Ice type	Ice drift velocity	Ice thickness	Flexural strength
32010	9500 MDOF	Model ice	4–150 mm/s	41 mm	56 kPa
25010	9200 MDOF	ICMI	4–150 mm/s	23 mm	86 kPa

where Y is the model output of interest; $X=(X_1, X_2, \ldots, X_M) \in \mathbb{R}^{M \times 1}$ is the model input which contains M factors; $f(X)$ can represent abstract models (data-driven models, mathematical model, or defined function) and mechanical models (robots). In this study, $f(X)$ corresponds to SSI-cov and Hierarchical clustering methods.

Sobol's method is based on the total variance decomposition.

$$V(Y) = \sum_{i=1}^{M} V_i + \sum_{i=1}^{M-1} \sum_{j=i+1}^{M} V_{ij} + \cdots + V_{1,\ldots,M} \qquad (6)$$

where $V(Y)$ is the variance of model output Y; V_i is the variance contribution of X_i to the model output; V_{ij} is the variance from the interaction between X_i and X_j; $V_{1,\ldots,M}$ represents the variance induced by the interaction between M parameters. The V_i is addressed as the first-order or main effect of X_i on Y. Therefore the first-order sensitivity index (S_i) of X_i is computed by Eq. (7). The total sensitivity index (S_{Ti}) is obtained by Eq. (8).

$$S_i = \frac{V_i}{V(Y)} \qquad (7)$$

$$S_{Ti} = \frac{V_i + V_{ij} + \cdots + V_{1,\ldots,M}}{V(Y)} = 1 - \frac{V_{\sim i}}{V(Y)} \qquad (8)$$

where $V_{\sim i}$ represents the total variance contribution of remaining parameters to Y given X_i. The detailed description can be referred to Saltelli et al. (2008).

'PAWN' is a CDF-based GSA method. The main principle is to estimate the difference between unconditional CDF and conditional CDF using Kolmogorov–Smirnov (KS) test.

$$\begin{cases} \hat{S}_i = \max_{k=1,\ldots,M} KS(I_k) \\ KS(I_k) = \max_y |F_y(y) - F_{y|\bar{x}_i}(y|\bar{x}_i \in I_k)| \end{cases} \qquad (9)$$

where KS is Kolmogorov–Smirnov statistic; $F_y(y)$ is unconditional CDF where $y \subseteq Y$ and $F_{y|\bar{x}_i}(y|\bar{x}_i \in I_k)$ is conditional CDF where $\bar{x}_i$ is fixed. The detailed information can be found in Pianosi and Wagener (2015).

The flowchart of the two SA methods is shown in Fig. 4. First of all, some variables are defined manually. 'N' is the size of Sobol's sample. 'M' is the number of parameters in modal parameters identification. 'Unif' assumes the distribution function is uniform. 'LHS' is a sampling strategy. 'NU' is the size of unconditional samples while 'NC' is the size of conditional samples. 'n' is the conditional point from which conditional samples are sampled. Next, different sampling methods are applied to generate different data samples. 'XA' is the sampled data that is used to estimate 'YA' through model evaluation while 'XB' is used to compute 'YB'. Likewise, 'XU' is to estimate 'YU' and 'XX' is to obtain 'YC'. After that, 'YA' is used to estimate V_i by the Sobol method while 'YB' is employed to estimate $V_{\sim i}$. 'YU' is to obtain $F_y(y)$ while is to get $F_{y|\bar{x}_i}$. Followed by Sobol and PAWN analysis, two sets of SA results can be obtained. They are compared to obtain convincing sensitivities of all parameters. After important parameters are picked, LHS sampling is used to generate a sample, and Model evaluation is conducted to carry out UQ of identified modal parameters. The two approaches are integrated into a MATLAB toolbox in Pianosi et al. (2015).

4. Case study

4.1. Data collection

Ice–structure interaction data are collected from model-scale experiments and full-scale experiments. All model-scale tests are carried out
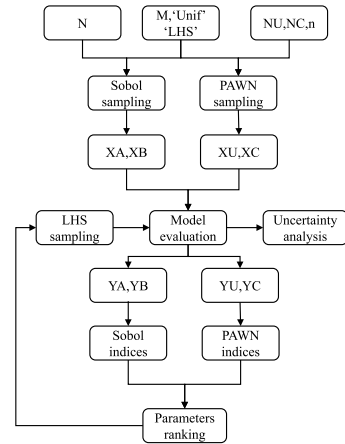

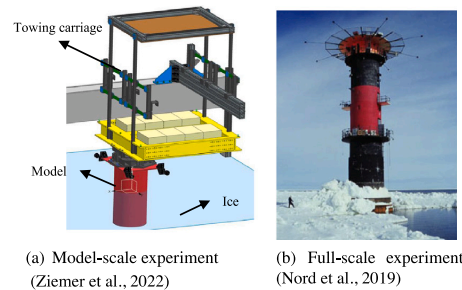**Fig. 4.** The flowchart of Sobol and PAWN methods for sensitivity analysis.

(a) Model-scale experiment (Ziemer et al., 2022)

(b) Full-scale experiment (Nord et al., 2019)

Fig. 5. The experiment setup for model-scale and full-scale tests.

in the Hamburg Ship Model Basin's (HSVA) large ice model basin.[1] The experiment setup was designed with a flexible foundation with adjustable mass and stiffness to mimic certain dynamic characteristics of the structure and a rigid model. Three Triax accelerometers are used to monitor the ice-induced vibrations of the structures in x- and y-direction (loading direction and perpendicular in-plane motion). The setup is shown in Fig. 5(a) (Stange et al., 2020). The data were collected under different structural and ice-related properties. The main tests and corresponding properties are shown in Table 1.

Different runs are designed with different global stiffness of the setup. Hydrodynamic added masses for models '9200' and '9500' are 16 kg and 19 kg separately. 'MDOF' represents the structure as multi-degree-freedom vibration which has two dominant frequencies ($f_1 = 2.81$ HZ and $f_2 = 3.77$ HZ). Ice types are HSVA's standard model ice and an improved crushing model ice (ICMI) (Ziemer et al., 2022). Model ice is generated by exposing the water surface to cooled air. The current model ice were not always ideal for crushing failure type of dynamic ice–structure interaction tests. Therefore, an alternative wave ice was proposed by simulating sea wave effects during water freezing. The first three test runs contain ice drift velocity from 4–150 mm/s whilst the ice velocity of '460101' starts from 14 to 150 mm/s. Based on different ice velocities, the measurements are grouped into different corresponding ice failure types (intermittent crushing (IC), frequency lock-in (FLI), and continuous crushing (CC)). The full data set is described by Stange et al. (2020).

---

[1] https://www.hsva.de/.

The full-scale experiments are conducted on the Norströmsgrund lighthouse which is located in the Gulf of Bothnia. As shown in Fig. 10, It is a gravity-based concrete structure with a wall thickness varying between 0.2 m at the top and 1.4 m at the mean water level (Nord et al., 2019). The diameter of the structure at the mean water level were 7.5 m at an elevation of +14.2 m . Nine panels were installed across the outer surface at the mean water level to measure the ice forces. Four accelerometers were equipped at different positions of the lighthouse. The detailed description can be found in Nord (2015), Nord et al. (2019).

4.2. Experiment setting

For sensitivity analysis, first of all, the sampling strategy and distribution function are selected as 'LHS' and uniform distribution. $\sigma = 40\% \times \mu$, where μ represents the predefined values of seven input parameters (Wang et al., 2022). Next, due to the time-consuming SSI-cov computation, it is not practical to set up a larger 'N'. Therefore, the Sobol index needs evaluation when 'N' is chosen as 2000. 'M' means 7 parameters, including the number of blocks (NB), block rows (BR), sampling frequency (SF), system order (SO), and stability criteria (the deviance in frequency (SC-I), the deviance in damping ratio (SC-II), and normalized standard deviation of the frequency (SC-III)). 'NU = 150', 'NC = 100', and 'n = 10' are referred to Pianosi and Wagener (2015). In Eq. (3), s is 0.5. As the first two modes (f_M, ξ_M, φ_M) are the modes in the model-scaled experiment that were designed to be easily excited by the ice force. Hence they are used for the output of interest in SA.

For UA, the sample size of uncertain parameters is 400. The benchmark values of the first and second natural frequencies are 21.35 and 29.52 (rad/s). SA is carried out using MATLAB toolbox (Pianosi et al., 2015). All MATLAB programs are run on the high-performance computer at the Norwegian University of Science and Technology (Själander et al., 2019).

4.3. Sensitivity analysis

Due to the desperately high time cost of model evaluations, several random cases are chosen for SA. For data file 32010, SA is applied for four different cases in which ice velocities are 14, 16, 18, and 20 (mm/s) separately. For data file 25010, ice velocities are from 18 to 26 (mm/s) with the step of 2 (mm/s). For full-scale measurements, ice velocities vary from 30 to 200 (mm/s).

Fig. 6 shows an example of how to use SA to pick up important factors. This figure plots the importance of seven factors to the first natural frequency when ice velocity is 20 mm/s. The blue squares are assumed indices with equivalent importance around 0.143. The PAWN index of SF is the largest at around 0.42. It is followed by BR at around 0.25. These two factors are to determine data size and input matrix for SSI-cov analysis. Hence, they have a very large influence on the identified frequencies. PAWN index ranks three around 0.21. That means the first frequency is impacted by the number of SO. However, when SO attains a certain value, the first frequency does not change anymore. Three stability criteria and NB account for a small contribution (from 0.1 to 0.17) to the variation of the first natural frequency. This is possibly because NB does not affect the eigenvalues, but only affects the results through the stabilization criterion. They can be removed from the set of uncertain input parameters. In other words, they are fixed at constant values in the following UA.

Fig. 7 presents statistics of parameters' PAWN indices among different velocities 14, 16, 18, 20 mm/s while parameters' Sobol indices are shown in Fig. 8. In both figures, the box represents the quantiles of parameters' sensitivity on the first and second natural frequencies. The PAWN indices of BR and SF are larger than the remaining overall. Therefore, BR and SF are chosen as uncertain factors for UA on data file 32010. In addition, for different cases, PAWN indices of different parameters have small variations. Nevertheless, in Fig. 8, Sobol indices
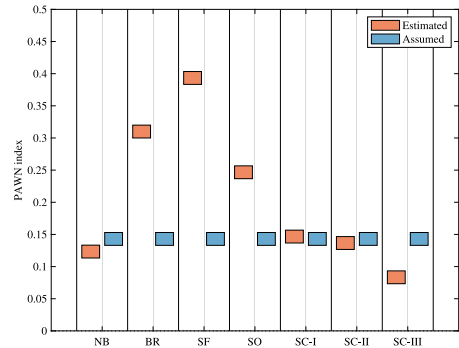

Fig. 6. The PAWN indices of seven input parameters on the first natural frequency when ice velocity is 20 mm/s.
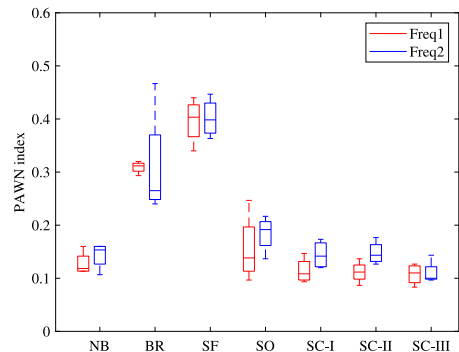

Fig. 7. The statistics of PAWN indices of seven parameters for different ice velocities in data file 32010.

of parameters vary drastically and are hard to rank. Therefore, the PAWN method has better performance regarding SA of parameters for ice–structure interaction analysis.

Fig. 9 shows the PAWN indices of BR and SF are the largest, varying from 0.25 to 0.55. The PAWN indices of the remaining parameters are below 0.2. Hence, BR and SF are chosen as uncertain factors for UA on data file 25010.

Fig. 10 shows the top two sensitive parameters are BR and SF whose PAWN indices are around 0.4 and 0.35 separately. Their indices are a bit smaller than the result in data files 32010 and 25010. This is possibly caused by the larger noise from the full-scale experiment in the real world.

To sum up, the most influential parameters are chosen as BR and SF for the following UA on data files 32010, 25010, and the full-scale experiment.

4.4. Uncertainty analysis

Table 2 lists the statistics of the identified first two natural frequencies. The statistics include mean value (Mean), standard deviation (Std), upper bound of 95% confidence interval (Upper), and lower bound of 95% confidence interval (Lower). The bold number means the correct identification. For the identified first natural frequency, the proposed method fails to identify it when ice failure is IC. Hence, it is not discussed in the following analysis. After the ice failure mode changes to FLI, the first natural frequencies are 19.84, 29.76, and 23.53 rad/s when ice velocities are 8, 10, and 20 mm/s, respectively. The
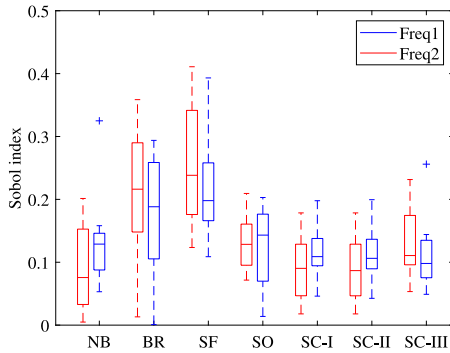
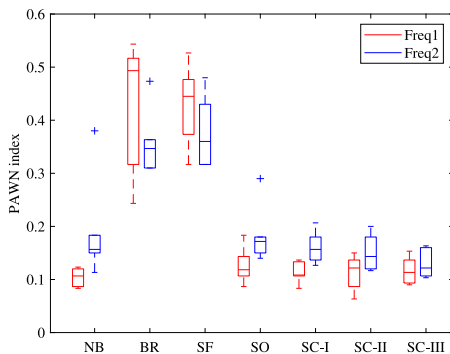**Fig. 8.** The statistics of Sobol indices of seven parameters for different ice velocities in data file 32010.

Fig. 9. The statistics of PAWN indices of seven parameters for different ice velocities in data file 25010.
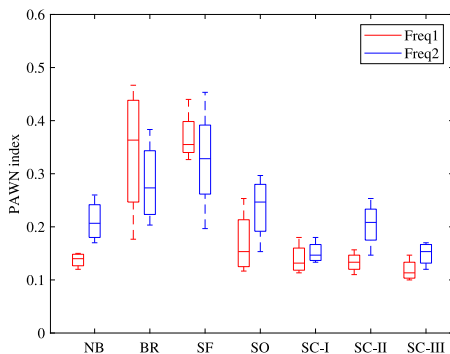

Fig. 10. The statistics of PAWN indices of seven parameters for different ice velocities in the full-scale test.

frequencies in the remaining cases are around 21.5 which is close to the benchmark value. In contrast to the first natural frequency, the second frequency identification becomes more intractable. 5 out of 15 identified second frequencies are over 30. One is less than 29, around 26.66 rad/s. Other frequencies can be identified correctly.

As to uncertainties of the identified frequencies, it is represented by the 95% confidence interval which is computed by '$Mean \pm 1.96 \times Std$'. The uncertainty explains the possible values of the identified

parameters. In Table 2, the lower bound of the first frequency is −7.38 when ice velocity is 4 (mm/s). This value is obtained by statistical analysis. However, it cannot be less than zero in practice. When ice velocities are 12, 18. 28, and 65 (mm/s), the standard deviation of the identified first frequency is less than 0.5 while the uncertainties for other cases are quite large. Compared with the standard deviation of the first natural frequency, the identified second frequencies have a larger standard deviation. Almost all of them are over 2.00. As a consequence, it has larger uncertain intervals for different ice velocities. That means input parameters cause different uncertainties to modal parameters.

Table 3 shows the corresponding damping ratios of modal parameters. For the damping ratio corresponding to the first natural frequency, it is quite small (less than 0.01) for ice failure FLI. When ice failure changes to CC, it becomes larger (more than 0.018). The standard deviation is very large. It is even larger than the mean value. For example, when ice velocities are 10, 14, 16, 20, and 28 (mm/s), the large standard deviations result in unreasonable significance intervals that exceed the range of zero. This varying trend of the first damping ratio is totally different from that of the second one. It can be seen that most FLI failures generate larger damping ratios (more than 0.017) than CC failures (less than 0.014) apart from ice velocity 150 mm/s.

To sum up, the mined knowledge is listed with the following items:

(1) The poor estimation of modal parameters happens as a result of the quasi-static response of the structure in IC.
(2) For the first natural frequency, the estimated mean value is quite close to the benchmark. The standard deviation is relatively small, which renders reasonable uncertainty intervals. With the increasing ice velocities, the varying trend is not significant.
(3) For the second frequency, the bold numbers manifest that the proposed method cannot have accurate estimations for certain cases. The standard deviation shows the estimating uncertainties are very large. FLI generates larger frequencies than CC.
(4) The damping ratio of the first mode rises with the increase of ice velocity while the damping ratio of the second mode decreases from FLI to CC.

4.5. Knowledge presentation

During UA, the first straight line in the stabilization diagram is taken as the first natural frequency while the second line represents the second natural frequency. Such analysis results in a large bias to the identified parameters. For example, the first natural frequency cannot be identified by the proposed method. There is not a straight line at the position where it is supposed to appear. As a consequence, the second natural frequency is considered the first natural frequency. At the same time, the third frequency is taken as the second one. Obviously, such an assumption causes very large uncertainties in the identified parameters as shown in Tables 2 and 3.

Next, prior knowledge is used to pick up anomalies in the results. The benchmark values of the first two natural frequencies are 21.352 and 29.516 rad/s separately, which are estimated when the structure was moving in the open water (Stange et al., 2020). Considering the difference between open water and ice, the benchmark values are expanded by 10% deviation to an interval: [19.22, 23.49] for the first frequency, [26.56, 32.47] for the second frequency. The processed UA results are shown in Fig. 11 for data file 32010, Fig. 12 for data file 25010, and Fig. 13 for the full-scale measurements.

Fig. 11(a) shows the filtered first natural frequency and corresponding damping ratio. The left Y axis is the value of the identified natural frequency. The right Y axis is the corresponding damping ratio. The green dotted line represents the identified first natural frequency in Wang et al. (2022). Black points are the mean value of identified frequencies. The length of the vertical line at each point means the estimated 95% confidence interval. Based on the varying trend of

Table 2

The uncertainty analysis results of identified frequencies based on the data in '32010' under the ice failures of IC, FLI, and CC.

Frequency	Statistics	Ice velocity (mm/s)														
		4	6	8	10	12	14	16	18	20	28	45	65	80	95	150
		IC	IC	FLI	FLI	FLI	FLI	FLI	FLI	FLI	FLI	FLI	CC	CC	CC	CC
First	Mean	45.59	17.57	19.84	29.76	21.69	21.02	21.92	21.06	23.53	21.28	21.59	21.50	21.25	21.96	21.81
	Std	28.04	5.68	1.64	2.41	0.01	1.57	4.55	0.01	3.39	0.42	0.02	0.25	1.22	2.10	1.23
	Upper	102.56	28.71	23.05	26.41	21.04	24.42	30.85	21.08	30.18	22.11	21.63	21.98	23.64	26.08	24.21
	Lower	−7.38	6.44	16.63	16.97	21.00	18.28	13.00	21.05	16.88	20.47	21.55	21.02	18.85	17.84	19.40
Second	Mean	88.00	29.73	26.66	32.00	32.49	29.68	32.43	29.01	30.38	29.71	29.32	29.93	29.42	29.76	29.57
	Std	28.84	14.48	3.44	10.59	3.88	3.62	9.07	1.60	5.12	0.66	1.82	1.92	2.43	12.02	4.47
	Upper	144.53	58.10	33.40	52.76	40.10	36.77	50.20	32.16	40.41	31.01	32.89	33.71	34.19	53.32	38.33
	Lower	31.47	1.35	19.91	11.24	24.88	22.58	14.65	25.87	20.35	28.42	25.74	26.16	24.65	6.19	20.80

Table 3

The uncertainty analysis results of identified damping ratios based on the data in '32010' under the ice failures of IC, FLI, and CC.

Damping	Statistics	Ice velocity (mm/s)														
		4	6	8	10	12	14	16	18	20	28	45	65	80	95	150
		IC	IC	FLI	FLI	FLI	FLI	FLI	FLI	FLI	FLI	FLI	CC	CC	CC	CC
First	Mean (10^{-2})	1.41	1.30	0.31	0.24	0.09	0.10	0.29	0.08	0.78	0.04	0.64	1.86	1.24	1.73	2.03
	Std (10^{-2})	0.86	0.61	0.51	0.62	0.03	0.39	0.82	0.03	1.09	0.08	0.03	0.71	0.33	0.37	0.50
	Upper(10^{-2})	3.11	2.50	1.30	1.46	0.14	0.86	1.89	0.41	2.92	0.20	0.69	3.25	1.89	2.45	3.01
	Lower (10^{-2})	−0.28	0.10	−0.68	−0.98	0.04	−0.67	−1.31	0.03	−1.36	−0.12	0.58	0.47	0.58	1.01	1.06
Second	Mean (10^{-2})	0.70	1.70	1.74	2.12	2.84	2.32	2.45	3.00	2.43	2.41	0.86	0.73	1.15	1.31	2.20
	Std (10^{-2})	0.71	1.76	0.85	0.77	1.33	1.15	1.47	1.06	0.79	0.63	0.19	0.29	0.38	0.43	2.13
	Upper (10^{-2})	2.08	5.16	3.29	3.63	5.45	4.58	5.33	5.06	3.97	3.65	1.24	1.30	1.90	2.15	6.37
	Lower (10^{-2})	−0.69	−1.75	0.08	0.61	0.24	0.06	−0.43	0.92	0.88	1.65	0.48	0.17	0.41	0.48	−1.97

the mean value, UA has a more sensible estimation than the method without UA, especially when ice velocities are larger than 45 mm/s. The superiority can also be verified in Fig. 11(b). Overall, the first natural frequency goes up with the increase in ice velocity. This trend is the same as the damping ratio as shown by the blue line. However, this trend does not appear on the identified second natural frequency as shown in Fig. 11(b). With the increase of ice velocity, the second natural frequency goes up until velocity reaches 16 mm/s. Next, it plummets when ice velocity changes to 18 mm/s. After that, it keeps an increasing trend from 20 mm/s. The damping ratio shows an overall decreasing trend when ice moves towards higher velocity. In addition, the uncertainty of the identified second frequency and damping ratio is significantly larger than that of the first frequency and corresponding damping ratio.

Fig. 12(a) shows the significant increasing trend of the first natural frequency in data file 25010. The value rises from 19.3 to 21. For the second natural frequency, its values fluctuate vary drastically as shown in Fig. 12(b). The same trend can be found in Fig. 13. In Fig. 13(a), the value of the first natural frequency starts from 18 to 22 as the increase of ice velocity. However, there is no obvious trend of the second natural frequency in Fig. 13(b). In addition, the value of the second natural frequency is far larger than that in the data files 32010 and 25010.

The mined knowledge is summarized as follows.

(1) Compared with traditional SSI-cov, UA based on prior knowledge does have superiority regarding the accuracy of modal parameter identification. In addition, the confidence interval could provide more valuable support for the SHM.
(2) The identified first natural frequency and corresponding damping ratio present a stable upward trend with the increase of ice velocity for three different experiments. No difference between model ice and wave ice was observed in relation to that trend. A small uncertainty interval demonstrates SSI-cov and Hierarchical clustering have very good robustness for the first frequency identification.
(3) For the second frequency, its trend is not significant for all three experiments. Large uncertainty means that SSI-cov and Hierarchical clustering work worse in terms of the second frequency identification.

(4) For data files 32010 and 25010, the damping ratio of the first mode is quite small, varying within [0% 3%]. However, it becomes larger in the full-scale experiment, varying within [5% 10%].

5. Discussion

This study includes two parts. The first part is SA and the second one is UA. Both SA and UA follow the same pipeline from data sampling, SSI-cov analysis, to automated modal parameters identification. Hierarchical clustering is used for automated modal parameter identification. In this study, robust distance and Silhouette value are used to improve the accuracy of Hierarchical clustering. Meanwhile, it consumes more time. It inevitably intensifies the time cost of SA. Therefore, all parameters in the Hierarchical clustering method are fixed. Although such an assumption could cause the low accuracy of modal parameters identification, in fact, SA does not require very accurate estimation because it mainly investigates how much the variation of the model is proportioned to each input factor. If time cost could be ignored, it would be better to use robust distance and Silhouette value to obtain a more accurate estimation.

Sobol and PAWN are used for SA. Through comparison of Figs. 7 and 8, it was found that PAWN outperforms Sobol with respect to sensitivity estimation of input parameters. To be exact, PAWN is superior to Sobol under the circumstance of an equivalent sample size. If Sobol's sampling size is increased up to, for example, over 10,000, Sobol could gain a very good estimation but it is desperately time-consuming. That is because the time complexity of PAWN is $O(M \times n \times Nc)$ while Sobol's time complexity is $O(M \times N)$. Hence, considering the balance of time cost and SA accuracy, this study prefers PAWN to Sobol.

PAWN is used for SA on data files 32010, 25010, and full-scale test. From Figs. 7, 9, and 10, they show the top two influential factors are BR and SF. It can be concluded that BR and SF are the largest uncertainty sources of SSI-cov. This conclusion needs verification further using more cases. In addition, SO accounts for a relatively large sensitivity to the identified frequencies to a certain extent. This is because the identified modes are affected by the variation of SO. Nevertheless,
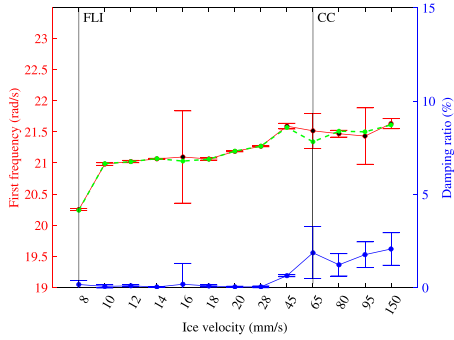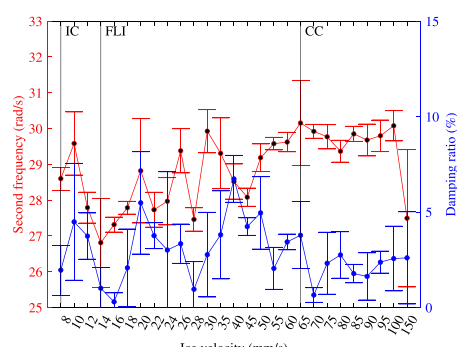
(a) First frequency and its damping ratio
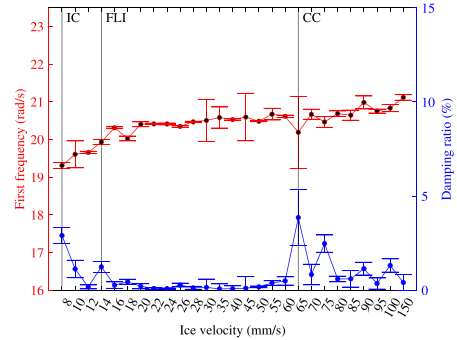
(a) First frequency and its damping ratio
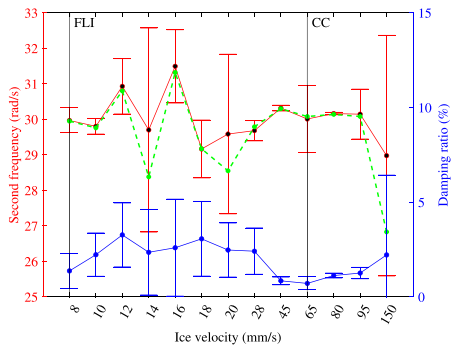
(b) Second frequency and its damping ratio

(b) Second frequency and its damping ratio

Fig. 11. The identified frequencies and corresponding damping ratios over different ice velocities in data file 32010 (Green dotted line: identified frequencies without UA; Black points: the mean value of identified frequencies; Vertical lines: 95% confidence interval.).

Fig. 12. The identified frequencies and corresponding damping ratios over different ice velocities in data file 25010.

when it attains a certain value, the modes are not changed with the increase of SO. Therefore, it is not considered in this study.

The same experiment settings are set for these three experiments. However, it does not work well for the full-scale test. This possibly results from less stable poles identified from the data. Therefore, in this study, we adjust the minimum size of a cluster to 5 for the full-scale data instead of 8 for the model-scale data. In addition, the benchmark interval is reset as [0,23.49] for the first natural frequency and [26.56, 50] for the second natural frequency. Such adjustments make the parameter identification more accurate for the full-scale test.

This study assumes all input parameters follow a uniform distribution instead of normal distribution. This is because sampled data from normal distribution could result in the failure of solving eigenvalues and eigenvectors. if a data point comes from a normal distribution, it might be beyond the defined range. For example, given $N(\mu, \sigma^2)$, the sampled data points would vary in $[\mu - 1.96\sigma, \mu + 1.96\sigma]$ rather than in $[\mu - \sigma, \mu + \sigma]$. Apparently, the former sampling space is larger than the later one. As a consequence, the sampled data points would exceed the rational range required by SSI-cov. In addition, in order to increase the uncertainty of the identified modal parameters, the σ is assumed as the 40% deviation of μ. if the deviation was increased, the sensitivity results would change a little bit. SA takes advantage of statistical methods to estimate the importance of factors and is subject to the impacts of many factors easily. The important parameters should be selected based on the integration of SA, expertise knowledge, and applications.

The second mode appeared difficult to identify accurately. One would expect that the second mode was easier than the first mode to identify, because it should be less influenced by the ice-failure process due to its smaller modal amplitude at the ice-action point, and therefore less nonlinearities for the SSI-cov to handle. However, due to the second mode's lower modal contribution of the global response, the signal to noise ratio for the second mode is also lower, thereby possibly inhibiting stable identifications.

6. Conclusion

The environmental variability of identified modal parameters of structures located in ice-infested waters are uncertain and to some extent unknown, therefore inhibiting efficient structural health monitoring. This study proposed a UA framework for automated modal parameter identification for structures exposed to drifting ice-conditions interaction by obtaining convincing parameters estimation with their uncertainty. The framework were composed by two parts: The first part applied an SA to picked up the most influential input variables for modal parameters' uncertainties, in which we presented a comparison between two methods for SA, namely PAWN and Sobol. PAWN was further selected for SA on ice–structure interaction data as it offered the most accurate estimation of parameters' importance. It was found that the most influential inputs to the modal parameters estimation were the number of block rows and sampling frequency. The input variables from PAWN were subsequently applied in a UA, which is the second part in Fig. 2. Here, the inputs were varied to obtain modal parameters
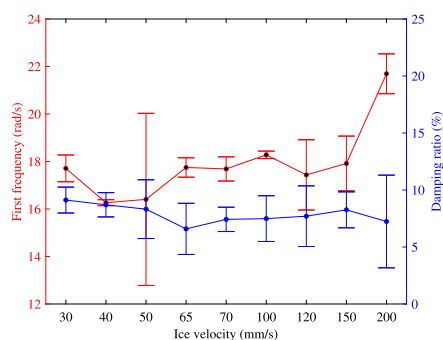
(a) First frequency and its damping ratio
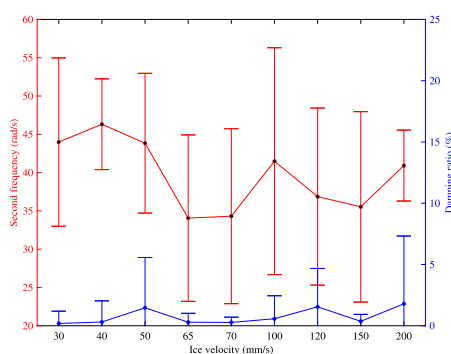
(b) Second frequency and its damping ratio

Fig. 13. The identified frequencies and corresponding damping ratios over different ice velocities in the full-scale test.

and corresponding uncertainties when applied to a covariance-driven stochastic subspace identification of modal parameters. The UA revealed significant trends in the modal parameters identified over a range of ice velocities and types of ice–structure interaction: (1) For model-scale ice–structure interaction, the first natural frequency and corresponding damping ratio arose with the increase of ice velocities (starting from frequency lock-in, and to continuous brittle crushing). Once the interaction was dominated by continuous brittle crushing, the natural frequency stabilized with further increase in ice velocity, an observation found for both model-scale and full-scale; (2) The second mode does not present a significant trend. In addition, uncertainties of the second mode (natural frequencies and damping ratios) are quite larger than those of the first mode.

The second mode appeared difficult to identify for the model-scale tests, likely due to the combination of second mode's low modal contribution of the global response and the signal to noise ratio. Future studies of environmental variability due to the presence of ice would benefit from denser sensor networks and higher sampling frequency.

CRediT authorship contribution statement

Chunlin Wang: Conceptualization, Methodology, Investigation, Writing – original draft. **Torodd Skjerve Nord:** Conceptualization, Writing – review & editing, Supervision. **Gesa Ziemer:** Review and editing, Supervision. **Guoyuan Li:** Review and editing, Supervision.

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.

Data availability

The data that has been used is confidential.

References

Au, S.K., 2011. Fast Bayesian FFT method for ambient modal identification with separated modes. J. Eng. Mech. 137 (3), 214–226.

Blenkarn, K., 1970. Measurement and analysis of ice forces on cook inlet structures. In: Offshore Technology Conference. OnePetro.

Döhler, M., Mevel, L., 2013. Efficient multi-order uncertainty computation for stochastic subspace identification. Mech. Syst. Signal Process. 38 (2), 346–366.

Dusenberry, M.W., Tran, D., Choi, E., Kemp, J., Nixon, J., Jerfel, G., Heller, K., Dai, A.M., 2020. Analyzing the role of model uncertainty for electronic health records. In: Proceedings of the ACM Conference on Health, Inference, and Learning. pp. 204–213.

He, M., Liang, P., Li, J., Zhang, Y., Liu, Y., 2021. Fully automated precise operational modal identification. Eng. Struct. 234, 111988.

Helton, J.C., Davis, F.J., 2003. Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. Reliab. Eng. Syst. Saf. 81 (1), 23–69.

Hendrikse, H., Nord, T.S., 2019. Dynamic response of an offshore structure interacting with an ice floe failing in crushing. Mar. Struct. 65, 271–290.

Hu, Z., Smith, R.C., Burch, N., Hays, M., Oates, W.S., 2014. A modeling and uncertainty quantification framework for a flexible structure with macrofiber composite actuators operating in hysteretic regimes. J. Intell. Mater. Syst. Struct. 25 (2), 204–228.

Hu, Z., Smith, R.C., Ernstberger, J., 2012. The homogenized energy model for characterizing polarization and strains in hysteretic ferroelectric materials: Implementation algorithms and data-driven parameter estimation techniques. J. Intell. Mater. Syst. Struct. 23 (16), 1869–1894.

Jefferies, M.a., 1988. Dynamic response of "Molikpaq" to ice-structure interaction. In: Proceedings of the 7th OMAE, Vol. 4. Houston, February 7–12, 1988, pp. 201–220.

Kärnä, T., Andersen, H., Gürtner, A., Metrikine, A., Sodhi, D., van het Loo, M., Kuiper, G., Gibson, R., Fenz, D., Muggeridge, K., et al., 2013. Ice-induced vibrations of offshore structures-looking beyond iso 19906. In: Proceedings of the International Conference on Port and Ocean Engineering under Arctic Conditions. POAC.

Kärnä, T., Jochmann, P., 2003. Field observations on ice failure modes. In: Proceedings of the 17th International Conference on Port and Ocean Engineering under Arctic Conditions. Trondheim, Norway, June, pp. 16–19.

Kvåle, K., Øiseth, O., 2021. Automated operational modal analysis of an end-supported pontoon bridge using covariance-driven stochastic subspace identification and a density-based hierarchical clustering algorithm. In: Bridge Maintenance, Safety, Management, Life-Cycle Sustainability and Innovations. CRC Press, pp. 3041–3048.

Lauwagie, T., Van Assche, R., Van der Straeten, J., Heylen, W., 2006. A comparison of experimental, operational, and combined experimental-operational parameter estimation techniques. In: Proceedings of the International Noise and Vibration Conference. ISMA, Citeseer, pp. 2997–3006.

Li, S., Pan, J., Luo, G., Wang, J., 2020. Automatic modal parameter identification of high arch dams: feasibility verification. Earthq. Eng. Eng. Vib. 19 (4), 953–965.

Magalhaes, F., Cunha, A., Caetano, E., 2009. Online automatic identification of the modal parameters of a long span arch bridge. Mech. Syst. Signal Process. 23 (2), 316–329.

Mao, J.X., Wang, H., Fu, Y.G., Spencer, Jr., B.F., 2019. Automated modal identification using principal component and cluster analysis: Application to a long-span cable-stayed bridge. Struct. Control Health Monit. 26 (10), e2430.

Nord, T.S., 2015. Force and Response Estimation on Bottom-Founded Structures Prone to Ice-Induced Vibrations. NTNU.

Nord, T.S., Kvåle, K.A., Petersen, Ø.W., Bjerkås, M., Lourens, E.M., 2017. Operational modal analysis on a lighthouse structure subjected to ice actions. Procedia Eng. 199, 1014–1019.

Nord, T.S., Petersen, Ø.W., Hendrikse, H., 2019. Stochastic subspace identification of modal parameters during ice–structure interaction. Phil. Trans. R. Soc. A 377 (2155), 20190030.

Pan, C., Ye, X., Mei, L., 2021. Improved automatic operational modal analysis method and application to large-scale bridges. J. Bridge Eng. 26 (8), 04021051.

Pappa, R.S., James, III, G.H., Zimmerman, D.C., 1998. Autonomous modal identification of the space shuttle tail rudder. J. Spacecr. Rockets 35 (2), 163–169.

Pathmanathan, P., Cordeiro, J.M., Gray, R.A., 2019. Comprehensive uncertainty quantification and sensitivity analysis for cardiac action potential models. Front. Physiol. 10, 721.

Pianosi, F., Sarrazin, F., Wagener, T., 2015. A Matlab toolbox for global sensitivity analysis. Environ. Model. Softw. 70, 80–85.

Pianosi, F., Wagener, T., 2015. A simple and efficient method for global sensitivity analysis based on cumulative distribution functions. Environ. Model. Softw. 67, 1–11.

Rainieri, C., Fabbrocino, G., 2014. Operational modal analysis of civil engineering structures. Springer, New York 142, 143.

Razavi, S., Jakeman, A., Saltelli, A., Prieur, C., Iooss, B., Borgonovo, E., Plischke, E., Piano, S.L., Iwanaga, T., Becker, W., et al., 2021. The future of sensitivity analysis: An essential discipline for systems modeling and policy support. Environ. Model. Softw. 137, 104954.

Reynders, E., Houbrechts, J., De Roeck, G., 2012. Fully automated (operational) modal analysis. Mech. Syst. Signal Process. 29, 228–250.

Reynders, E., Pintelon, R., De Roeck, G., 2008. Uncertainty bounds on modal parameters obtained from stochastic subspace identification. Mech. Syst. Signal Process. 22 (4), 948–969.

Saltelli, A., Aleksankina, K., Becker, W., Fennell, P., Ferretti, F., Holst, N., Li, S., Wu, Q., 2019. Why so many published sensitivity analyses are false: A systematic review of sensitivity analysis practices. Environ. Model. Softw. 114, 29–39.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., Tarantola, S., 2008. Global Sensitivity Analysis: The Primer. John Wiley & Sons.

Singh, S., Timco, G., Frederking, R., Jordaan, L., 1990. Tests of ice crushing on a flexible structure. In: Proc. 9th OMAE Conf., Vol. 4. Houston, pp. 89–94.

Själander, M., Jahre, M., Tufte, G., Reissmann, N., 2019. EPIC: An energy-efficient, high-performance GPGPU computing research infrastructure. arXiv:1912.05848.

Sobol, I.M., 2001. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. Math. Comput. Simulation 55 (1–3), 271–280.

Stange, T., Ziemer, G., von Bock und Polach, R.U.F., 2020. Development of an experimental setup to investigate the impact of higher structural modes on dynamic ice-structure interaction. In: 25th IAHR International Symposium on Ice 2020. URL: http://hdl.handle.net/11420/8438.

Timco, G., 1991. Laboratory observations of macroscopic failure modes in freshwater ice. In: Cold Regions Engineering. ASCE, pp. 605–614.

Wang, C., Nord, T.S., Li, G., 2022. Automated modal parameters identification during ice-structure interactions. In: International Conference on Offshore Mechanics and Arctic Engineering, Vol. 85864. American Society of Mechanical Engineers, V002T02A020.

Xiao, H., Cinnella, P., 2019. Quantification of model uncertainty in RANS simulations: A review. Prog. Aerosp. Sci. 108, 1–31.

Xie, Y., Liu, P., Cai, G.P., 2016. Modal parameter identification of flexible spacecraft using the covariance-driven stochastic subspace identification (SSI-COV) method. Acta Mech. Sinica 32 (4), 710–719.

Yang, X.M., Yi, T.H., Qu, C.X., Li, H.N., Liu, H., 2019. Automated eigensystem realization algorithm for operational modal identification of bridge structures. J. Aerosp. Eng. 32 (2), 04018148.

Ye, C., Zhao, X., 2020. Automated operational modal analysis based on DBSCAN clustering. In: 2020 International Conference on Intelligent Transportation, Big Data & Smart City. ICITBS, IEEE, pp. 864–869.

Zahid, F.B., Ong, Z.C., Khoo, S.Y., 2020. A review of operational modal analysis techniques for in-service modal identification. J. Braz. Soc. Mech. Sci. Eng. 42 (8), 1–18.

Zeng, J., Hoon Kim, Y., 2021. A two-stage framework for automated operational modal identification. Struct. Infract. Eng. 1–20.

Zhang, J., Yin, J., Wang, R., 2020. Basic framework and main methods of uncertainty quantification. Math. Probl. Eng. 2020.

Ziemer, G., Stange, T., Hisette, Q., 2022. HSVA model ice–A status report. In: International Conference on Offshore Mechanics and Arctic Engineering, Vol. 85918. American Society of Mechanical Engineers, V006T07A010.

This paper is not included due to copyright restrictions

$\mathcal{F}$

Paper VI

This paper is awaiting publication and ins not included

NTNU

Norwegian University of
Science and Technology