Ship Performance and Navigation Information under High Dimensional Digital Models

Lokukaluge P. Perera[1] and Brage Mo[2]

***Abstract*— Future vessels will be facilitated by modern internet of things (IoT) to collect various ship performance and navigation information. Such information is collected as large-scale data sets, so called Big Data and that should be utilized towards digitalization of the shipping industry. However, various data handling challenges are encountered by the shipping industry during the phase of digitalization, onboard as well as onshore. Data driven models, so called digital models, to support data handling frameworks of the shipping industry are proposed by this study. Such models can overcome the respective data handling challenges in shipping, where conventional mathematical models may fail to facilitate. These models can be derived from ship performance and navigation data sets by considering the high dimensional parameter space. Such high dimensional models consist of several data clusters and each data cluster may consist of a possible unique data structure. These data clusters often relate to sub-operational conditions of vessels and ship systems. The identification of the distribution of data clusters and the structure of each data cluster in relation to ship performance and navigation conditions have been done under this study for a selected vessel as the main contribution. Furthermore, the domain knowledge in shipping (i.e. vessel operational and navigation conditions) is also considered during this analysis to interpret the meaning in such digital models.**

***Index Terms*— Digitalization, Shipping industry, big data, digital models, data analytics, machine leaning and artificial intelligence.**

1 Introduction

1.1 Industrial Revolution

Fourth industrial revolution, i.e. categorized as Industry 4.0, that is facilitated by large-scale data (i.e. so called big data) with machine learning and artificial intelligence technologies will have the greatest impact on humanity. Modern industrial systems are facilitated by various internet of things (IoT), i.e. onboard and onshore, to collet such big data sets and that should be processed, appropriately to extract relevant system information. Such system level information can be used towards industrial digitalization to improve both efficiency and reliability considerations of the respective industries. Furthermore, such systems should consist of various autonomous functionalities with considerable industrial intelligence [1] to support the requirements of industrial digitalization and that have been done by machine learning (ML) and artificial intelligence (AI) technologies. However, there are various challenges have also been introduced by industrial digitalization. A considerable data handling challenges have been encountered by these industries, while transforming big data into appropriate information and such challenges can be summarized as: erroneous data conditions (i.e. data veracity), data modeling uncertainty, estimation algorithm failures, data visualization challenges and high computational power, etc. Hence, appropriate technologies to overcome such data related challenges should be developed under the respective industrial platforms. It is believed that such technologies can be developed through the same data sets, therefore the solutions can also relate to the respective industrial domain. This study proposes data drive models that can develop from the same data sets as the solution for the respective challenges. Even though the concept of data driven models has been considered under various transport system applications [2], the implementation of ML and AI technologies are in a preliminary stage, especially under the maritime transportation. Therefore, this study proposes to use ML and AI technologies in a meaningful manner to facilitate the proposed data driven models and overcome the respective data handling challenges in the shipping industry.

Data scientists are often involved with such data driven model development under various industrial applications. It is also expected that may provide the solutions to the same data handling challenges. Such models are often limited to conventional data analysis techniques, i.e. clustering and pattern recognition, of the respective industrial data sets. However, the ignorance or lack of the industrial domain knowledge within the data scientists may lead to erroneous data driven models in these industrial platforms. Therefore, such erroneous models may not support to overcome the data handling challenges that have discussed,

[1] L. P. Perera is with UiT The Arctic University of Norway, Tromso, Norway (e-mail: prasad.perera@uit.no), Corresponding author

[2] B. Mo is with SINTEF Oceans, Trondheim, Norway (e-mail: brage.mo@sintef.no).

previously. The shipping industry is also encountering the similar challenges, therefore the main contribution in this study is to present the development process of appropriate data driven models, i.e. digital models and the usage of such models to overcome the respective data handling challenges. Furthermore, appropriate methodologies, i.e. ML and AI technologies and adequate domain knowledge in shipping have also been incorporated into such digital models.

1.2 IoT & Big Data in Shipping

Future vessels will be facilitated by various onboard and onshore IoT to collect ship performance and navigation data. Such data sets can often be collected from ship navigational and automation systems and communicated as big data sets from onboard vessels to onshore centers. Furthermore, these data sets should be analyzed both onboard, i.e. data pre-processing, and onshore, i.e. data post-processing, with appropriate technologies to overcome the data handling challenges have also been encountered by the shipping industry under industrial digitalization. Shipping industrial applications are often limited to conventional mathematical models with empirical parameter relationships, therefore that may often fail to facilitate towards big data challenges [3], as mentioned before. Hence, a data handling framework with appropriate digital models, as a part of industrial digitalization in the shipping industry is proposed. Such models under a high dimensional data space can facilitate with the required solutions to extract relevant information and overcome the data handling challenges in shipping. It is also believed that the same models may present superior performance under both ship energy efficiency and system reliability application [4, 5] as compared with the existing models.

Onboard and onshore IoT collects ship performance and navigation data that intend to support the respective navigational and operational strategies. Such IoT are connected with ship navigation and automation systems to collect and exchange the data among vessels and onshore centers for the same purpose. The same network connectivity within the maritime infrastructure, i.e. IoT, creates global internetworking facilities for the shipping industry to function as an information-based industry [6]. Hence, the same maritime infrastructure creates various opportunities, i.e. efficiency, accuracy and economic benefits, for shipping in a global scale. However, appropriate integration among ship performance and navigation data, modern information technology (IT) systems and intelligent algorithms (i.e. digital models and data analytics) should be developed to enhance such industrial opportunities. Furthermore, such system integration can encounter various industrial challenges in shipping, especially in big data handling situations and even under proper maritime infrastructure. This study proposes adaquate solutions to overcome those challenges under a data-handling framework supported by intelligent algorithms (i.e. digital models and data analytics). Therefore, that eventually facilitates towards smart industrial digitalization in the shipping industry.

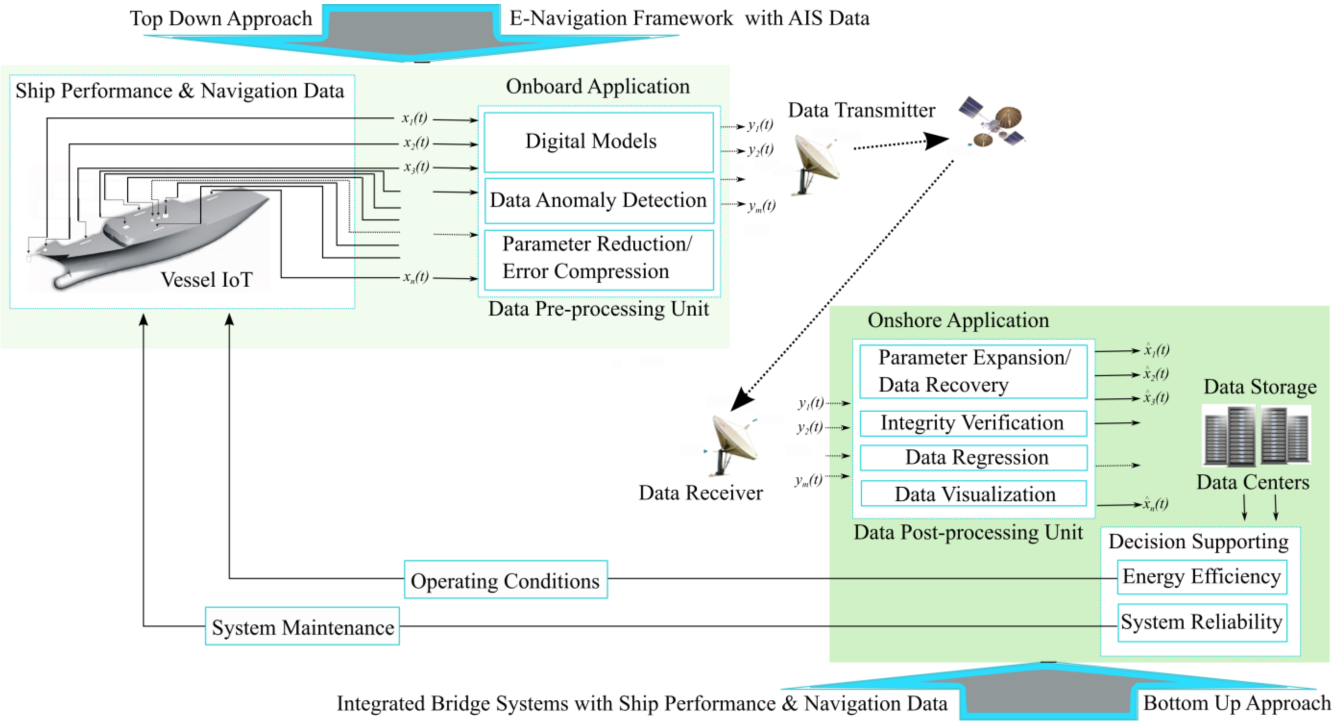

Figure 1. Data handling framework

1.3 Data Handling Framework

A modified data handling framework from the previous work [7] is presented in Figure 1. This framework consists of various data handling layers to facilitate ship performance and navigation data that are collected by onboard IoT. The same framework is supported by AIS (i.e. Automatic identification system) data collected under e-navigation in a global scale (i.e. a top-down approach) [8]. That is also supported by onboard integrated bridge systems with ship performance and navigation collected by vessel IoT (i.e. onboard IoT) in a local scale (i.e. a bottom up approach). In general, the data handling process is divided into two sections of pre and post processing units (see Figure 1). Data pre-processing unit is an onboard application consisting data driven models (i.e. digital models), anomaly detection and parameter reduction/error compression layers. The digital models as the first data-handling layer are developed by identifying various clusters within ship performance and navigation data sets, where an appropriate structure for each data cluster is also introduced. One should note that the same approach introduces a proper structure into the respective ship performance and navigation data sets and the data set its self becomes the model for the system. In the next layer, detecting and isolating data anomalies, i.e. including sensor and data acquisition (DAQ) faults and abnormal events, from ship performance and navigation data are conducted.

The respective structure derived under data driven models, i.e. digital models, is also considered for data anomaly detection layer. The significant outliers (i.e. high dimensional boundaries) beyond the respective data structure are classified as data anomalies in a high dimensional space [9, 10]. However, data anomalies can be further divided into sensor and DAQ faults and system abnormal events, therefore the domain knowledge in shipping should be used to make the respective classification. One should note that some data clusters in ship performance and navigation data sets can also relate to such abnormal events, where the same domain knowledge in shipping should be used to identify such situations. Therefore, such data clusters and the respective outliers in a high dimensional space will relate to the respective data anomalies and that knowledge can be used to improve the data quality. As the next step of this data handling process, such anomalies should be reduced and that can be done by the parameter reduction/error compression layer.

The last step in data pre-processing is denoted as the parameter reduction/error compression layer and the first step in data post-processing is denoted as the parameter expansion/data recovery layer. The main objective in these two layers to reduce and expand the number of parameters in ship performance and navigation data sets. Therefore, a set of new parameters, i.e. the reduced data set, that is a representation of original ship performance and navigation parameters can be created by the parameter reduction/error compression layer, while preserving the same amount of the respective information. The reduced data set can be expanded back to its original ship performance and navigation parameters by the parameter expansion/data recovery layer, similarly preserving the same amount of the respective information. One should note that data anomalies that are detected by the previous layer can be recovered by these layers. The parameter reduction layer compresses the data anomaly regions, i.e. error compression, and the parameter expansion layer can recover the same data anomaly regions, i.e. parameter recovery, in ship performance and navigation data sets by considering the same data driven models. The data structure that has developed under digital models is used for the parameter reduction and expansion layers. Therefore, these layers can improve the quality and reduce the quantity of the respective data sets.

The pre-processed data can communicate from onboard transmitters as much smaller improved data sets to shore based data centers. One should note that additional low-level data compression technologies can be introduced before the transmission process to further reduce the size of ship performance and navigation data sets. That reduces the respective data transfer costs from onboard vessels to onshore data centers. The modified ship performance and navigation data sets that are obtained by onshore data centers should be transferred though the data post-process unit (see Figure 1). The remaining handling layers in the data post-processing unit can be categorized as data integrity verification, regression, visualization and decision supporting layers. The same digital models, i.e. the data structure, that are developed in the data pre-process unit should also be used under these layers. Other data sources (i.e. AIS, weather data, etc.) to verify measured ship performance and navigation data can be considered under the integrity verification layer. It is expected that additional data anomaly regions (i.e. communication errors, etc.) can be identified and isolated from this handling layer and that will further improve the quality of ship performance and navigation data sets.

However, the parameter reduction and expansion layers can introduce some parameter variations into estimated ship performance and navigation data sets. The regression layer is introduced to minimize such parameter variations in ship performance and navigation data sets and that can associate with various data smoothing algorithms. The next data handling layer visualizes the respective data sets to support decision supporting applications in shipping. The same data structure developed under digital models can be used under the data visualization layer, where ship operating and navigation parameters can be visualized, appropriately. However, such data structures can be time-varying because the respective parameter relationship can vary due to vessel and ship systems performance and navigation conditions. Furthermore, the health conditions of vessels and ship systems can vary, i.e. due degradation conditions, along their remaining useful life (RUL) [11], therefore appropriate visualization techniques should be used under these analytics to observe such variations. The time-varying data structure can be used to quantify vessel operational and navigation conditions and ship system health conditions. The same information can be used for vessel energy efficiency [12] and system reliability applications [13, 14], were the respective decisions supporting features should be developed (see Figure 1). That has been done by the decision supporting layer of the data handling framework supported by the same data driven models.

1.4 Data Driven Models

Several conventional mathematical models have been presented in the recent literature to evaluate ship performance and navigation conditions from onboard data sets [15, 16]. These models are applied for an inland river ship [16], a steam-propelled merchant ship [17], a small training ship [18], and a passenger ferry [19] in the following studies. In addition, data analysis methods on the fuel usage for various ship operational and navigation conditions are presented in the study of [20]. It is noted that these studies consist of conventional mathematical models and methodologies consisting various parameter relationships, however it is noted that such models may fail to facilitate large-scale ship performance and navigation data sets. On other hand these models may not be able to utilize all ship performance and navigation parameters that are collected by onboard and onshore IoT. This is due to the reason that these conventional mathematical models and methodologies consist of limited parameter relationships. Hence, this study proposes to use high dimensional digital models to overcome such situations in the shipping industry [21]. These models are also categorized as digital models due to their discreteness in a high dimensional data space and that represents sub-operational conditions of ship performance and navigation conditions of ocean-going vessels. Therefore, such models can support digitalization of the shipping industry.

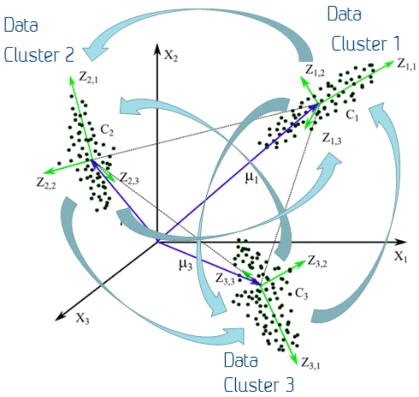

Fig. 2. Digital Models

A simplified view of the proposed digital models in a three-dimensional space is presented in Figure 2. The model development steps consist of identifying various data clusters within ship performance and navigation data set and the structure of each data cluster. That creates a vector structure (i.e. a data structure), i.e. a represents sub-operational conditions of vessel performance and navigation conditions (see Figure 2). The initial digital models can be derived at onshore data centers by considering a relatively cleaner data set, then that can be deployed onboard vessels as a software application. Then, the same model can be used to further improve the data structure, in real-time. Therefore, improved digital models can be a part of this framework to support the respective data handling layers. The clustered data sets (i.e. digital models) can have additional flexibility in real-time because the data sets are much smaller, therefore less computational power is required under the same framework. These data clusters relate to sub-operational conditions (i.e. engine-propeller operating points, trim-draft combination points) of the respective ship operating and navigation situations in a high dimensional space. Since these data clusters represent sub-operational conditions, the respective ship performance and navigation conditions can jump from one data cluster to another under the ship operational and navigation space.

This behavior has noted from the respective ship performance and navigation conditions, where the respective ship performance and navigation parameters are clustered around its data space. Therefore, the same feature introduces the discreteness into these data driven models, i.e. digital models. Furthermore, that can also be denoted as the linearization (i.e. piecewise linearization) of ship performance and navigation conditions under various operational and navigational points. However, this data structure, i.e. digital model, also relate to the number of ship performance and navigation parameters that are collected by onboard IoT. On the other hand, the same data structure relates to ship energy efficiency and system reliability conditions, therefore the respective structural changes under various time-horizons should be observed to quantify the respective efficiency and reliability conditions. Figure 2 represents a digital model in a three-dimensional vector space, i.e. with the right-hand coordinate system of $X_1X_2X_3$. X_1, X_2 and X_3 denote the respective parameters of a selected data set. It is assumed that the data set consists of three data clusters and that are denoted by C_1, C_2 and C_3. The data clusters distributed in a three-dimensional vector space, therefore the respective mean vectors are denoted by μ_1, μ_2 and μ_3. The identification of the respective data clusters (i.e. how these data clusters are distributed in a high dimensional space) is the first step in developing such digital models. The identification of the respective structure of each data cluster (i.e. the parameter relationships/correlations) is the second step in developing such models. In general, the structural vectors, i.e. singular vectors, of the i-th data cluster are denoted as $Z_{i,1}$, $Z_{i,2}$ and $Z_{i,3}$. One should note that these structural vectors, i.e. singular vectors, represent the important covariance directions of the data sets, therefore each singular vector represents the respective parameter correlations and each singular value represents the percentage of the ship performance and navigation information that each vector carried with respect to the set of singular vectors. These parameter correlations, i.e. singular vectors, can be presented under the visual analytics layer in the data handling framework, where the relationships among the respective mean values of each data cluster and the structure of each data cluster, i.e. singular vectors, in relation to ship performance and navigational conditions can be observed.

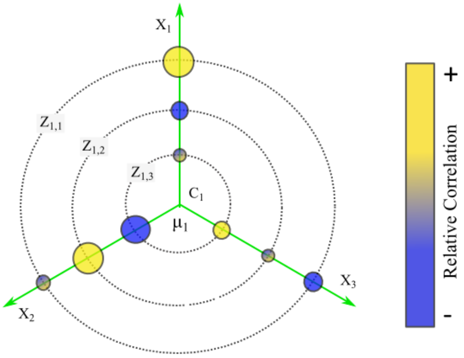

Fig. 3. High dimensional vector space.

There can be various challenges in the development phase of the digital models in shipping. The identification and visualization, i.e. imagination, of such digital models (i.e. data clusters and the respective data structure) can be a challenge, since these models may exist in a high dimensional data space. Each ship performance and navigation parameter can introduce one dimensionality into the data set, therefore the model complexity can go higher with each sensor measurement. On the other hand, the number of absolute data clusters in a data set is unknown and extremely difficult to visualize. Furthermore, these data clusters can have strange shapes, distributions and overlaps. Even though there are various ML and AI algorithms, i.e. based on the distance among data points, have been developed to capture the respective clusters in a data set, the outcome of such algorithms is still dissatisfactory, i.e. failed to identify the proper number of data clusters [22]. Since parameter values are scattered in ship performance and navigation data sets, an appropriate number of data clusters should be identified to succeed in those data driven models. However, some algorithms may find sub-data clusters within a main data cluster and that can introduce erroneous conditions into the proposed digital models. Therefore, the domain knowledge in shipping should extensively be used to verify the respective clusters in ship performance and navigation data sets. The identification of the respective data clusters from ship performance and navigation data sets can be the main difficulty in developing such data driven models. On the other hand, this number of data cluster may also relate to the respective decision supporting application that should be developed onboard the vessel. Furthermore, an appropriate number of ship performance and navigation parameters should also be selected to derive the required number of data clusters, i.e. a proper data structure, where the extensive domain knowledge in shipping can be used. However, the number of ship performance and navigation parameter can have some limits due to onboard IoT and the domain knowledge can also be inadequate in some situations. As the next step of this study, the data structure of each data cluster is identified and visualized to extract the respective parameter relationships on ship performance and navigation conditions.

1.5 Information Visualization

Data visualization is an important tool for decision supporting type onboard applications, i.e. energy efficiency and system reliability, and that is the last layers in the data handling framework. The optimal vessel navigational and ship system operational conditions can be identified by data visualization to support energy efficiency type applications and that reduce fuel consumption and emissions in shipping. Ship performance and navigation conditions should be visualized, appropriately in such situations to identify the optimal operational and navigation conditions of vessels. Those optimal operational and navigation conditions of vessels eventually influence on decision supporting layer of the data handling framework. One should note that the visualization layer is also based on the proposed digital models, i.e. the structural vectors. Therefore, the respective KPIs (i.e. key performance indicator) in the decision supporting layer can also be facilitated by the same structural vectors, i.e. singular vectors, of the data driven models.

Figure 3 represents a simplified version of the structural vectors, i.e. singular vectors, in a data set, where the relative correlations among three parameters in a high dimensional space are presented. One should note that the descending order of structural vectors represent the descending covariance directions, i.e. singular values, that are orthogonal in the data cluster.

Therefore, that covariance information can be transformed into relative correlations among ship performance and navigation parameters as discussed before and that are presented in the same figure. Hence, the respective structural vectors are represented as $Z_{1,1}$, $Z_{1,2}$ and $Z_{1,3}$ (see Figure 3). One should note that this figure represents a three-dimensional vector space and that is different from the conventional vector representation, i.e. three dimensional cartesian coordinate system. The largest and smallest covariance direction of the data cluster are represented by the singular vectors of $Z_{1,1}$ and $Z_{1,3}$, respectively. Hence, the most and least important information of the data cluster is accommodated in $Z_{1,1}$ and $Z_{1,3}$, respectively. Furthermore, various data anomalies may often accommodate in the least important singular vectors of the data set. The color scheme is developed with respect to the relative correlation among three parameters and that can be defined as: $Z_{1,1}$ shows that: when parameter X_1 increases (or decreases) , then parameter X_3 decreases (or increases), and parameter X_2 may not vary much with respect to other parameter variations. Hence, this figure represents the behavior of one parameter with respective to other parameters in the data sets in relation to the respective structural vectors. The same observations, i.e. the relative correlation information, can relate to the optimal ship performance and navigational conditions, where that can be identified by the data structural vectors.

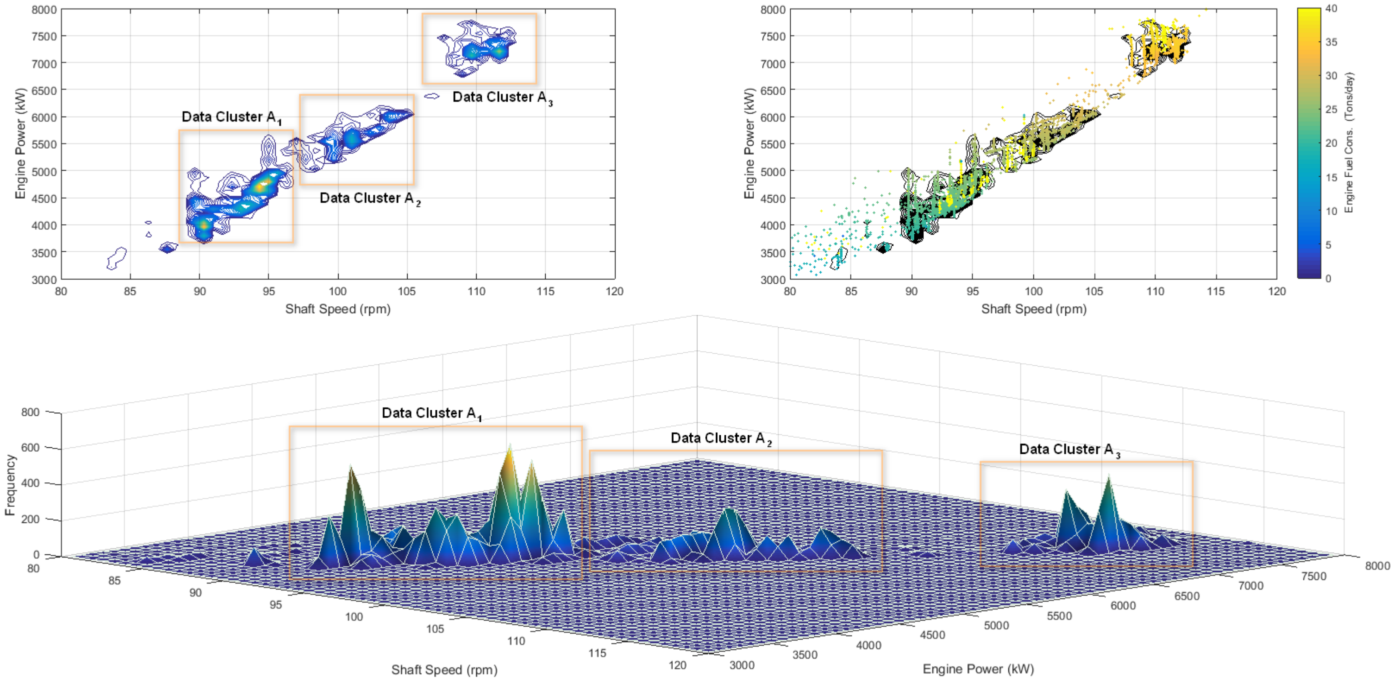

Fig. 4. Data Clustering in Engine Propeller Combinator Diagram.

2 Data Analysis

2.1 Vessel Particulars

As the next step of this study, a data set from a selected vessel is analyzed to present the respective digital model development. The data set consists of ship performance and navigation parameters that is collected from a bulk carrier, to derive the digital models. The vessel consists of following particulars: ship length: 225 (m), beam: 33 (m), gross tonnage: 38.889 (tons), deadweight at max draft: 72.562 (tons). The vessel is facilitated by a 2-stroke main engine (ME) with maximum continuous rating (MCR): 7564 (kW) at the shaft rotational speed of 105 (rpm). Furthermore, the vessel is facilitated with two auxiliary engines with MCR: 850 (kW) at the respective shaft rotational speed of 800 (rpm). The vessel also has a fixed pitch propeller with 6.20 (m) in diameter and 4 blades. One should note that the ship performance and navigation data set consists of 10 parameters (units): STW (speed through water) (Knots), SOG (speed over ground) (Knots), ME (main engine) power (kW), shaft speed (rpm), ME fuel consumption (cons.) (Tons/day), auxiliary (aux.) fuel consumption (cons.) (Tons/day), average (avg.) draft (m), trim (m), and relative (rel.) wind speed (m/s) and direction (deg). This is a time-series data set that has a sampling rate of 15 (min) and collected in approximately 3 (years).

2.2 Digital Model Development

Firstly, an appropriate number of data clusters in the ship performance and navigation data set should be derived. That has been done by identifying the respective operational modes of the main engine under the engine propeller combinatory diagram. It is noted that the engine propeller combinator diagram is a good basis of the proposed digital models. Therefore, the main engine related parameters as statistical distributions (i.e. histograms) to identify the respective operational modes of the main engine are considered. It is previously reported that the histograms for main engine speed, power and fuel consumption parameters are shown as data clusters that relate to the engine operational modes [5]. The same results are presented in another format, where the respective combined histogram (i.e. a combinator diagram) of the engine speed and power parameters is presented in the bottom plot of Figure 4.

The results show that three data clusters that relate to the main engine operational modes do exist in the data sets. Hence, it shows that the engine propeller combinator diagram, i.e. engine power and shaft speed values are combined, can be a good basis for the proposed data driven models (i.e. digital models). The main reason for this selection is that the ship performance and navigation data are clustered around the parameters of engine speed, power and fuel consumption, therefore those three parameters create a three dimensional space with an appropriate number of data clusters, i.e. the mean operating points or engine modes of the marine engine. However, possible clustering relationships among other ship performance and navigation parameters should also be observed in a later stage and that step may increase the accuracy of digital models. However, the number of ship performance and navigation parameters and their relationships can also relate to the respective onboard application, as mentioned before. Hence, the engineering judgments along with the domain knowledge in shipping and statistical parameter distributions should be combined to develop these data clusters and the will further increase the accuracy of data driven models. Such steps can introduce a more meaningful structure into ship performance and navigation data sets and that can improve for the respective onboard application.

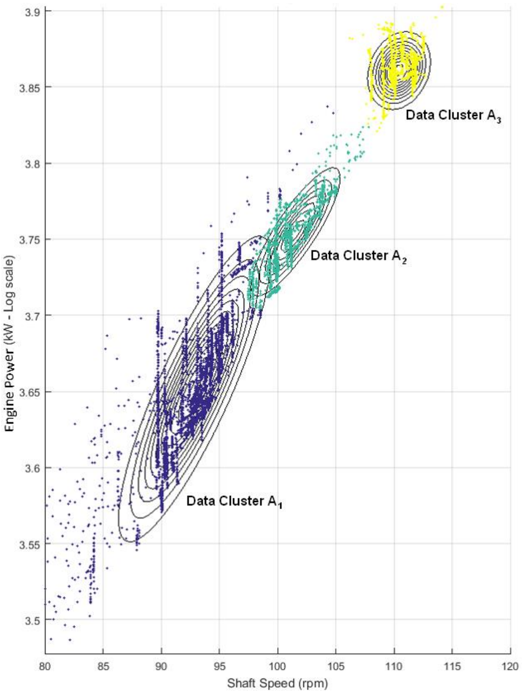

Fig. 5. Engine Propeller Combinator Diagram with GMMs.

2.3 Data Clustering

The results in Figure 4 have concluded that the ship performance and navigation data set is clustered around the three engine modes. The parameters of engine power and shaft speed as a statistical distribution are presented in the bottom plot of the same figure, i.e. an engine propeller combinator diagram. The contour plot of the same engine propeller combinator diagram is presented in the top left plot of the same figure, where the most frequent engine operational regions are highlighted [23]. The same contour plot with the respective engine fuel consumption is presented in the top right plot of the same figure. The plot shows a positive correlation between the parameters of engine power and fuel consumption. Slow ship maneuvering situations, i.e. near zero speed-power values, are removed in this data analysis, therefore a considerable number of sensor and DAQ fault and noise conditions are removed. The respective data clusters that relate to engine operational modes are denoted by clusters

A_1, A_2, and A_3 in the same plot. In the next step, an appropriate algorithm is proposed to identify the respective data clusters in the engine propeller combinator diagram. In addition, another algorithm to identify the structure, i.e. singular values and vectors, of each data cluster identified by the previous step is also proposed. Therefore, that creates the basis for the proposed digital models under the proposed data handling framework.

2.4 Gaussian Mixture Models

Gaussian mixture models (GMMs) with an expectation maximization (EM) algorithm [24] is proposed in this study to identify the respective data clusters in engine propeller combinator diagram. One should note that GMMs represent multivariate Gaussian distributions with the respective mean and covariance values. The outcome of data clustering is presented in Figure 5, where the engine operational modes are denoted by three multivariate Gaussian distributions, i.e. under the GMMs. One should note that these GMMs denoted as contour plots relate to data clusters A_1, A_2, and A_3 of the engine propeller combinator diagram (see Figure 4). The respective data points are also categorized with respective to each data cluster (i.e. GMM) and that are denoted in different colors (i.e. blue, green and yellow). This data classification is done by using the expectation-maximization (EM) algorithm in which calculates the respective parameters of the GMMs under of the engine propeller combinator diagram. Furthermore, it is assumed that other ship performance and navigation parameters have also been clustered around the same engine modes. This step concludes the first step, i.e. the classification, of developing the digital models. The second step, i.e. the structural identification, of developing that digital models is described in the following step.

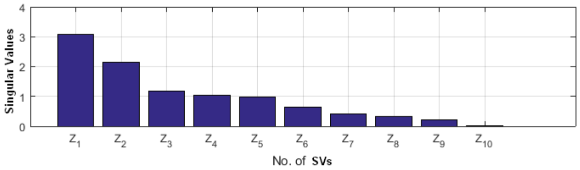

Fig. 6. Singular vectors with values of Data Cluster A_3

2.5 Singular Values and Vectors

Singular value decomposition is used to identify the respective structure of each data cluster under the engine propeller combinator diagram. Singular vectors represent the largest covariance directions, i.e. that are orthogonal, in the data cluster. Hence, the respective structure (i.e. the largest covariance directions that are orthogonal) in the data set can be identified by this approach. When the parameters are projected into each singular vector direction, a set of linearly uncorrelated variables can be derived through possibly a set of correlated variables [25]. The singular value decomposition of data cluster A_3 is presented in this section. The respective singular values and vectors in descending order of the same data cluster are presented in Figures 6 (i.e. scree-plot) and 7. Figure 6 presents an overview of the ship performance and navigation information distribution, i.e. singular values, for each singular vector (SV). The top and bottom singular values are denoted as Z_1 and Z_{10}, respectively. The respective singular vectors are presented in Figure 7 and that represents a high dimensional vector space (i.e. see Figure 3) as discussed previously.

A 10 dimensional vector space is presented in Figure 7 and the respective singular vectors are presented in dotted circles. The outer and inner circles represent the top and bottom singular vectors, respectively. The dotted circles intersect the respective axes that represent the parameters of ship performance and navigation data set. Each intersection consists of a colored circle and that represents the relative correlation among the respective ship performance and navigation parameters. The size of each circle represents the significance, i.e. the strength of the parameter correlation, of that parameter with respect to others and the color of each circle represent the positive or negative sign of the parameter correlation for the same parameter. High positive (HP) correlations are represented by yellow color large circles and high negative (HN) correlations are represented by blue color large circles in the same figure (see the color bar). Therefore, that can be an overview of the relative correlations among the respective ship performance and navigation parameters. One should note that this representation is derived from the respective singular values and vectors of the ship performance and navigation data set.

A brief overview of the singular vectors is discussed in this section. The top singular vector shows that ship resistance increases due to draft increments, where STW and SOG of the vessel decrease. The same conditions decrease shaft speeds, where the engine loading condition is higher. It is also noted that the draft increments are compensated by trim adjustments in this situations by the navigator. Hence, the increment in vessel avg. draft increases trim, decreases STW, and decreases SOG.

The same data cluster, i.e. cluster A_3, as a time series (i.e. with the respective sample number) is presented in Figure 8, where the respective ship performance and navigation parameters are presented. The top singular vector is noted as the most visible feature in this data set and that is marked in a window in the same figure. However, other singular vectors are hard to observe from a time series data set, therefore the singular vector representation in Figure 7 should be used to extract other relevant information from this data cluster. The results show that the most important information in ship performance can be extracted by the singular values and vectors as the proposed digital models. Furthermore, other singular vectors can be discussed in the similar manner [26].

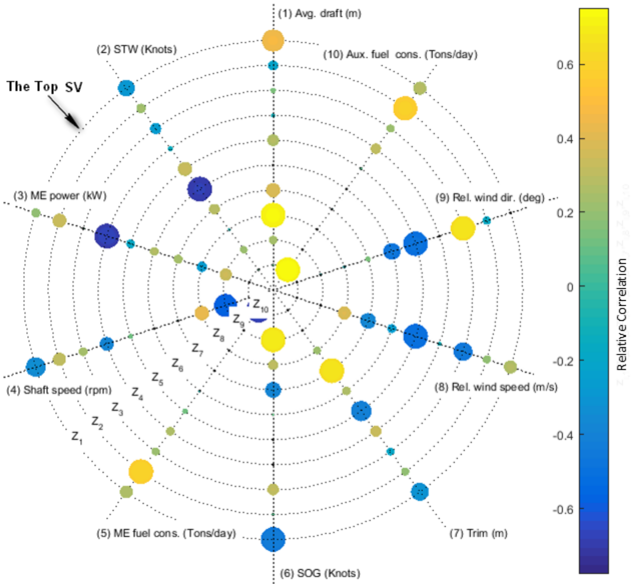

Fig. 7. Singular Vectors of Data Cluster A_3

The 2nd top singular vector shows: a moderate increment in engine speed increases engine power, moderately and fuel consumption in main and auxiliary engines, significantly. These results indicate that the shaft speed increment beyond the mean operating point (i.e. in this engine operating region) may increase fuel consumption, significantly but that may not increase engine power, considerably. The 3rd top singular vector shows: a small decrease in ship speed, decreases rel. wind speed, moderately and increases rel. wind angle, significantly. The 4th top singular vector shows: a moderate increase in shaft speed increases engine power, significantly. One should note that this parameter behavior can be unique to this engine operating mode. The 5th top singular vector shows: a moderate increase in vessel trim increases rel. wind speed, significantly and increases rel. wind direction, significantly. One should note that these features may represent a contradictory relationship with respective to the 3rd top singular vector. That can happen due to several reasons: that may relate to some parameter inconsistency within the respective data cluster. On the other hand, the rel. wind direction range consists from 0 (deg) to ± 180 (deg) and that may have resulted in such strange correlations. One should note that some ship performance and navigation parameters should be transformed into a better format to observe appropriate parameter correlations. On the other hand, this data cluster may consist of additional hidden data clusters and that can also be resulted in such contradictory relationships. The vessel may have operated under special trim and avg. draft conditions and such situations can introduce these hidden data clusters. When such strange parameter correlations due to hidden data clusters can be observed, that may guide towards better data driven models.

The 6th top singular vector shows: a significant decrease in ship speed, decreases rel. wind direction, moderately. The 7th top singular vector shows that: a moderate increase in avg. draft decreases ship speed, moderately, i.e. a situations with increased ship resistance. Furthermore, that reduces rel. wind speed and the draft increments are compensated by trim increments of the vessel, i.e. a slow maneuvering situation. The 8th top singular vector shows: a significant increase in avg. draft increases shaft speed, moderately. This is a situation, where ship resistance increases due to draft increments, therefore the navigator increases engine power to compensate for the speed losses in the vessel. The 9th top singular vector shows: a significant decrease in shaft speed (high) increases SOG, significantly. The 10th top singular vector shows: a significant decrease in ME fuel consumption increases aux. fuel consumption, significantly. It is noted that the bottom singular vectors (i.e. the 9th and 10th top singular vectors) may not represent any realistic parameter correlations due to their low singular values (see Figure 6). The data anomaly regions can often be projected into the bottom PCs, as mentioned before, therefore that may not consist of any useful ship performance and navigation information. The low positive and negative correlations among the same parameters are neglected, however that should also be incorporated into the discussion to see an overall picture of ship performance and navigation behavior. That will also show some complex relationships among ship performance and navigation parameters and the proposed data structure, i.e. digital models, is capable of representing such relationships.

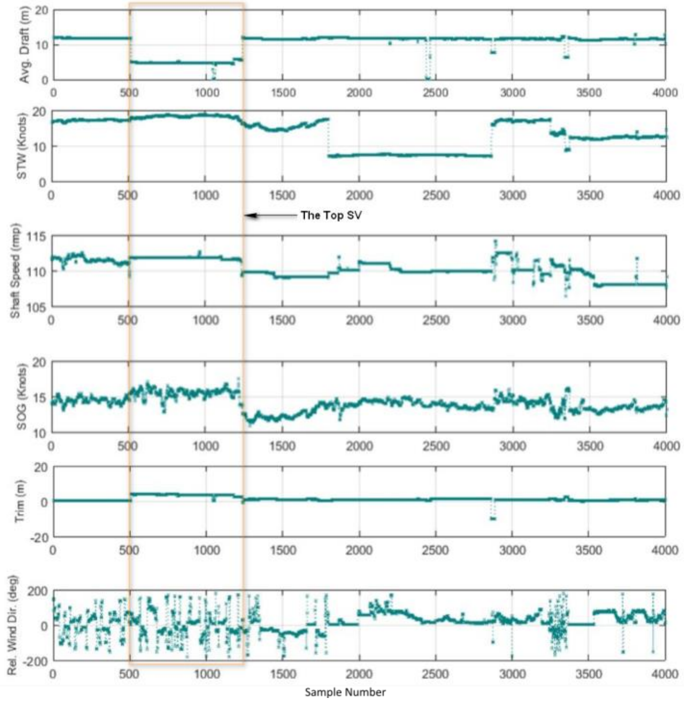

Fig. 8. Data Cluster A_3 with the sample number

2.6 Hidden Data Structure

Some data anomalies and hidden clusters can influence the outcome in digital models and that should further be investigated. Furthermore, such situations should be identified and recovered from the ship performance and navigation data to improve the accuracy of digital models [27]. Hidden data clusters can assign high-dimensional structures due to other parameter behavior into ship performance and navigation data sets and that have been further investigated in this section. The understanding of such hidden data structure can further help to detect various data anomalies that can be outlies of the respective data clusters. Data cluster A_3 is visualized with respect to trim and avg. draft conditions and the results are presented in Figure 9. The top plot represents avg. draft and trim configurations and the respective contour plot with STW is presented in the bottom plot of the same figure. That shows this vessel is operating three types of trim and avg. draft modes with respect to the same engine mode in the engine propeller combinator diagram. Hence, data cluster A_3 can be separated in additional data clusters of A_{31}, A_{32} and A_{31} in a 4 dimensional space by considering these additional parameters. One should note that data clusters A_{31}, and A_{31} may relate to actual ship trim and avg. draft conditions and data cluster A_{31} may relate to a data anomaly situation.

This can be an example to show that these types of data clusters can be hidden with the data sets and appropriate methodologies to discover the same should be investigated. These data cluster relate to engine operation modes and avg. trim-draft configurations, i.e. the domain knowledge, in this vessel and that can assign a meaningful vector structure into the ship performance and navigation data set. Therefore, the accuracy of the respective digital models can be higher. However, some scattered data regions in the engine propeller combinator plot as well as avg. draft-trim configuration plot have also noted and that may relate to various data anomalies. Such data regions can introduce additions challenges in identifying the respective vector structures of ship performance and navigation data sets. However, the same data structure, i.e. digital models, can also be used to detect and recover such data anomalies as discussed under the data handling framework.

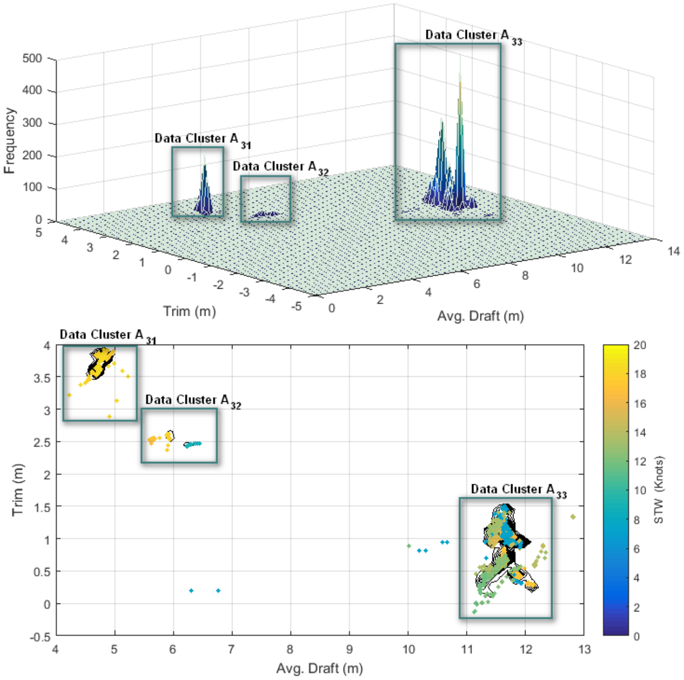

Figure 9. Avg. draft and trim configuration of data cluster A_3

3 Conclusions

This study presents a data handling framework with pre and post processing units to overcome various data handling challenges that are encountered by the shipping industry during the phase of digitalization. The basis of this data handling framework is the proposed digital models and that is also the main contribution of this study. It is also discussed that the proposed digital models can overcome the respective data handling challenges, where conventional mathematical models may fail. It has been shown that the respective digital models can be derived from ship performance and navigation data sets and that is also representation of the data structure. Even though the accuracy of such data driven models can be improved by sampling the ship performance and navigation data at a higher frequency, the shipping industry is not storing such types of data sets. Therefore, the improvements in the digital models under higher frequency data sets are yet to be investigated.

The main core of the data handling framework is the pre and post processing units of ship performance and navigation data. The pre-process improves the quality and reduces the quantity and the post-process further improves the quality and visualizes the information of the data sets. The pre-process reduces the computational burden on the crew onboard and quality improved and reduced data sets delivers to onshore data centers, where the communication costs of the respective data sets can be further reduced. The post-process reduces the complexity in handling big data sets, where high skilled crew to analyze such data sets may not require, i.e. but the digital models. Both processes consist of several data handling layers that are vital to ship onboard and shore based data handling processes and supported by the proposed data driven models, i.e. digital models, to overcome the respective challenges in real-time. These models exist in a high dimensional parameter space consisting of several data clusters and each data cluster may have of a unique data structure that relate to sub-operational conditions of vessel performance and navigation conditions. It has shown that the respective information on such sub-optimal conditions can be extracted as singular values and vectors, in which can be visualized. These vectors represent the most important ship performance and navigation relationships. It is also noted that the selection of ship performance and navigation parameters and their ranges should be done properly to make the singular vectors more meaningful.

One should note that a majority of the parameters in the selected data set relate to ship performance and navigation conditions. However, several parameters that relate to weather and environmental conditions, i.e. rel. wind speed and direction, have also been considered in this study. In general, high winds can create rough weather conditions [28] and that can degrade ship performance and navigation conditions. However, a good parameter correlation among rel. wind conditions and other ship performance and navigation parameter have not bee observed in this study. The rel. wind direction is introduced into this data set in a way that it may not correlate, properly with other parameters, i.e. since wind angle change from 0 (deg) to 360 (deg). Therefore, the external influences, i.e. weather and environmental conditions, should be introduced into the data sets in a particular manner that should have a good correlation with other parameters. This same concept can reiterate as the measurements of ship performance and navigation data should be transformed in a particular manner that can demonstrate the optimal parameter relationships and correlations. However, this issue, i.e. the optimal parameter transformation, is still investigating and the respective outcomes will be presented in the future research studies. Therefore, the same outcome along with the proposed digital models can be used for weather routing type applications in shipping. These digital models are well-suitable for weather routing type applications due their special features of self-learning (i.e. data clusters and the structure of each data cluster), self-cleaning (i.e. sensor and DAQ fault removal and compression, data recovery, data regression & integrity verification), self-compression & expansion (i.e. parameter reduction and expansion). These models can also be a part of the proposed data handling framework [29] and that may open a novel path towards digitalizing the shipping industry [30] under high dimensional data spaces.

Acknowledgement

An initial version of this paper is presented at the 36th International Conference on Ocean, Offshore and Arctic Engineering (OMEA 2017), June, 2017, Trondheim, Norway. This work has been conducted under the project of "SFI Smart Maritime (237917/O30) - Norwegian Centre for improved energy-efficiency and reduced emissions from the maritime sector" that is partly funded by the Research Council of Norway.

Reference

[1] E. Uhlemann, "Connected-Vehicles Applications Are Emerging," in IEEE Vehicular Technology Magazine, vol. 11, no. 1, pp. 25-96, March 2016.

[2] J. Zhang, F. Y. Wang, K. Wang, W. H. Lin, X. Xu and C. Chen, "Data-Driven Intelligent Transportation Systems: A Survey," in IEEE Transactions on Intelligent Transportation Systems, vol. 12, no. 4, pp. 1624-1639, Dec. 2011.

[3] O.J. Rodseth, L.P. Perera, and B. Mo, ”Big data in shipping - Challenges and opportunities,” In Proceedings of the 15th International Conference on Computer Applications and Information Technology in the Maritime Industries (COMPIT 2016), Lecce, Italy, May, 2016, pp. 361-373.

[4] L.P. Perera and B. Mo, ”Data Compression of Ship Performance and Navigation Information under Deep Learning,” In Proceedings of the 35th International Conference on Ocean, Offshore and Arctic Engineering (OMAE 2016), Busan, Korea, June, 2016, (OMAE2016-54093).

[5] L.P. Perera and B. Mo, ”Data Analytics for Capturing Marine Engine Operating Regions for Ship Performance Monitoring,” In Proceedings of the 35th International Conference on Ocean, Offshore and Arctic Engineering (OMAE 2016), Busan, Korea, June, 2016, (OMAE2016-54168).

[6] A. R. J. Ruiz and F. S. Granja, “A short-range ship navigation system based on ladar imaging and target tracking for improved safety and efficiency, ”IEEE Trans. Intell. Transp. Syst., vol. 10, no. 1, pp. 186–197, Mar. 2009.

[7] L. P. Perera and B. Mo, "Machine Intelligence Based Data Handling Framework for Ship Energy Efficiency," in IEEE Transactions on Vehicular Technology, vol. 66, no. 10, pp. 8659-8666, Oct. 2017.

[8] IMO, 2007. Development of an e-Navigation Strategy, IALA and e-navigation, Subcommittee on Safety of Navigation, Report of the Correspondence Group on e-Navigation, NAV/53/13/3.

[9] L.P. Perera, ”Statistical Filter based Sensor and DAQ Fault Detection for Onboard Ship Performance and Navigation Monitoring Systems,” In Proceedings of the 8th IFAC Conference on Control Applications in Marine Systems (CAMS 2016), Trondheim, Norway, September 2016, pp. 323-328.

[10] L.P. Perera, "Marine Engine Centered Localized Models for Sensor Fault Detection under Ship Performance Monitoring," In Proceedings of the 3rd IFAC Workshop on Advanced Maintenance Engineering, Service and Technology, (AMEST'16), Biarritz, France, vol. 49, no. 28, October, 2016, pp. 91-96.

[11] Z. Li and Q. He, "Prediction of Railcar Remaining Useful Life by Multiple Data Source Fusion," in IEEE Transactions on Intelligent Transportation Systems, vol. 16, no. 4, pp. 2226-2235, Aug. 2015.

[12] L.P. Perera and C. Guedes Soares, ”Weather Routing and Safe Ship Handling in the Future of Shipping,” Journal of Ocean Engineering, vol. 130, 2017, pp. 684-695.

[13] L.P. Perera, M.M. Machado, A. Valland, D.A.P. Manguinho, "Failure Intensity of O shore Power Plants under Varying Maintenance Policies," Engineering Failure Analysis, 2019, (DOI: 10.1016/j.engfailanal.2019.01.011).

[14] C. Zhiwei, S. Yufeng, Z. Tingdi and S. Fangfang, "Reliability analysis of phased-mission complex systems for warship," 2016 Prognostics and System Health Management Conference (PHM-Chengdu), Chengdu, 2016, pp. 1-6.

[15] Soner, O., Akyuz, E. & Celik, M. J Mar Sci Technol (2018). https://doi.org/10.1007/s00773-018-0574-y

[16] Petersen JP, Jacobsen DJ, Winther O (2012) Statistical modelling for ship propulsion efficiency. J Mar Sci Technol 17(1):30–39

[17] C. B. Dickinson, "A Method of Propulsion Plant Performance Evaluation for Marine Applications," in Industry Applications, IEEE Transactions on , vol. IA-10, no. 2, pp. 316-324, March 1974.

[18] T. Nakatani, T. Miwa, N. Yamatani, K. Sasaya, D. Okada, T. Kaneda, E. Kanayama, and E. Ura, "Dynamics analysis and optimal control of a marine diesel engine," in Control, Automation and Systems (ICCAS), 2013 13th International Conference on, pp. 1261-1265, Oct. 2013.

[19] P. D. Osborne, D. B. Hericks, and J. M. Cote, "Full-Scale Measurements of High Speed Passenger Ferry Performance and Wake Signature," in OCEANS 2007, pp. 1-10, 2007.

[20] D. G. Trodden, A. J. Murphy, K. Pazouki, J. Sargeant, "Fuel usage data analysis for efficient shipping operations," Ocean Engineering, vol. 110, part B, December 2015, Pages 75-84.

[21] E. B. Besikci, O. Arslan, O. Turan, and A. I. Oler, "An artificial neural network based decision support system for energy efficient ship operations," Computers & Operations Research, vol. 66, pp. 393–401, 2016.

[22] L. Mak, M. Sullivan, A. Kuczora, and J. Millan, "Ship performance monitoring and analysis to improve fuel efficiency," in Oceans - St. John's, Sept. 2014, pp.1-10.

[23] A. K. Jain, Data clustering: 50 years beyond K-means, Pattern Recognition Letters, vol. 31, no. 8, June 2010, pp. 651-666.

[24] L.P. Perera and B. Mo "Marine Engine Centered Data Analytics for Ship Performance Monitoring," Journal of Offshore Mechanics and Arctic Engineering-Transactions of The ASME, 2016 (DOI: 10.1115/1.4034923).

[25] L.P. Perera and B. Mo, "Marine Engine Operating Regions under Principal Component Analysis to evaluate Ship Performance and Navigation Behavior.," In Proceedings of the 8th IFAC Conference on Control Applications in Marine Systems (CAMS 2016), Trondheim, Norway, September 2016, pp. 512-517.

[26] L.P. Perera and B. Mo, "Digitalization of Sea going Vessels under High Dimensional Data Driven Models," In Proceedings of the 36th International Conference on Ocean, Offshore and Arctic Engineering (OMAE 2017), Trondheim, June, 2017 (OMAE2017-61011).

[27] J. E. Jackson, "Principal components and factor analysis: part i-principal components." Journal of Quality Technology, vol. 12, no. 4, pp. 201–213, 1980.

[28] L.P. Perera, B. Mo, and M. P. Nowak "Visualization of Relative Wind Profiles in relation to Actual Weather Conditions of Ship Routes," In Proceedings of the 36th International Conference on Ocean, Offshore and Arctic Engineering (OMAE 2017), Trondheim, Norway, June, 2017 (OMAE2017-61120).

[29] L.P. Perera and B. Mo, "Data Analysis on Marine Engine Operating Regions in relation to Ship Navigation," Journal of Ocean Engineering, vol. 128, 2016, pp. 163-172.

[30] Y. Lv, Y. Duan, W. Kang, Z. Li and F. Y. Wang, "Traffic Flow Prediction with Big Data: A Deep Learning Approach," in IEEE Transactions on Intelligent Transportation Systems, vol. 16, no. 2, pp. 865-873, April 2015.