

efficiency and emission control rules and regulations in shipping [2] often influence these decisions. Therefore, industrial digitalization eventually supports towards energy efficient, low emission and highly reliable smart shipping fleets.

2 Mathematical Framework

2.1 Ship Performance and Navigation Data

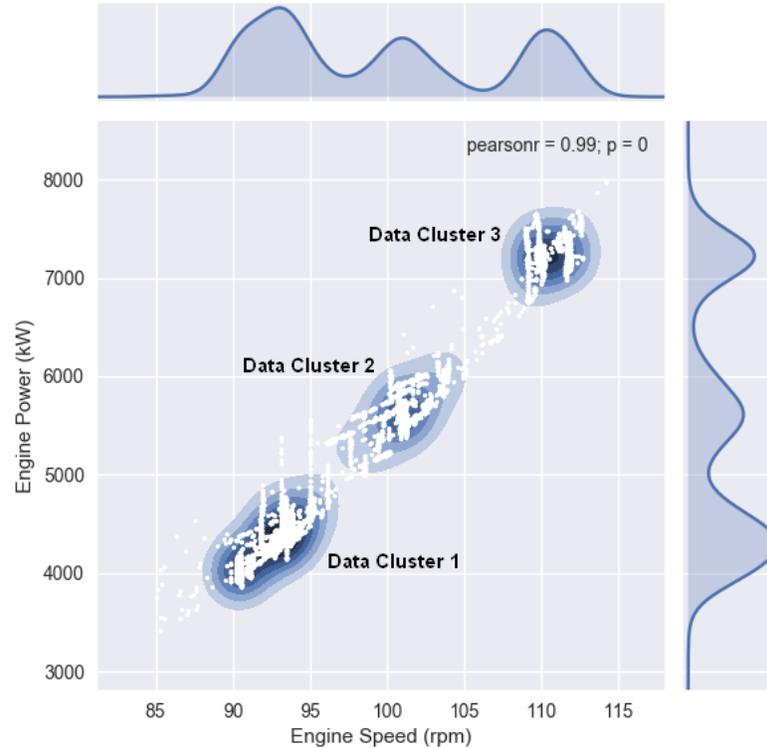


Fig. 2. Ship engine data

A novel mathematical framework to support industrial digitization of shipping is presented in this study. The framework consists of a data flow path, where digital models and advanced data analytics are introduced. Several statistical analyses, machine learning (ML) and artificial intelligence (AI) algorithms are used in this framework [3]. Ship performance and navigation data from a selected vessel are considered to develop this mathematical framework. The vessel is a bulk carrier with ship length: 225 (m) and beam: 32.29 (m) and the respective particulars are presented in [4]. Two parameters, i.e. engine power and engine speeds, of this vessel are presented in Figure 2. A combined kernel density estimation (multivariate KDE) of the same parameters is in the middle plot and the respective univariate KDEs are presented in the top and right plots of the same figure. One should note that the white dots represent the parameter values and the contours represent the density of the data distribution (i.e. multivariate KDE). This plot shows that the engine data are clustered around three Gaussian type distributions and that are denoted as data cluster 1, 2 and 3. These data clusters represent the respective engine modes of this vessel. Therefore, it is concluded that ship performance and navigation data sets are often clustered in a high dimensional space (i.e. with the respective parameters) and those clusters relate to vessel navigation and ship system operational conditions. The same introduce the discreteness (i.e. digital-ness) into the proposed mathematical framework, where the respective digital models are introduced. The distribution of such data clusters made the foundation of digital models and an example such models is presented in Figure 3.

2.2 Digital Models

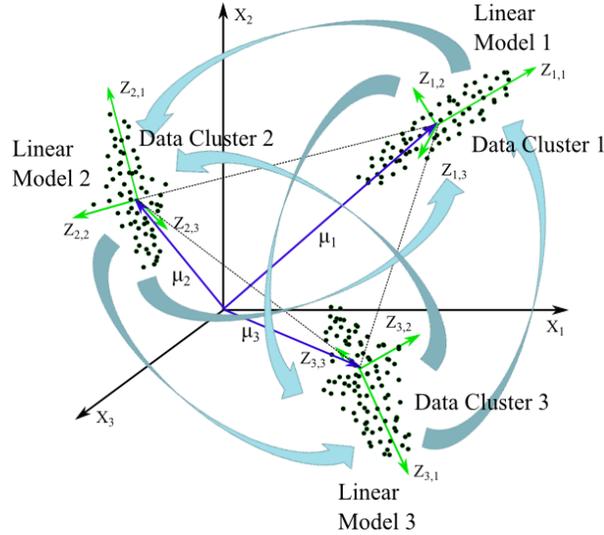


Fig. 3. Digital models

Digital models can be derived purely from the ship performance and navigation data sets. The initial digital models should derive by using a relatively cleaner data set (i.e. less data anomalies) and the same models should be used to improve the quality of the entire data set at a later stage. The quality improved data set should again be used to improve the same digital models. Figure 3 represents a three-dimensional vector space with the right-hand coordinate system of $X_1X_2X_3$. One should note that X_1, X_2 and X_3 represent the parameters of the respective data set. This model consists of three data clusters as denoted in the figure with the respective mean vectors of μ_1, μ_2 and μ_3 . One should note that ship performance and navigation data sets consist of such clusters due to the vessel operational and navigation conditions (i.e. engine modes, trim and draft conditions), as mentioned before. Therefore, the proposed digital models are well suitable to represent such data sets. The first step in developing digital models is to identify the respective data clusters (i.e. how these data clusters are distributed in a high dimensional space). Each data cluster consists of local navigation and operational information of the vessel and ship systems. Such local operational and navigation conditions should be investigated at the second step, where the respective structure of each data cluster (i.e. the relative correlation among ship performance and navigation parameters) should be identified. The structure of each data cluster is denoted by several vectors, therefore the i -th data cluster is represented as $Z_{i,1}, Z_{i,2}$ and $Z_{i,3}$ vectors (see Figure 3). That creates a vector structure (i.e. a data structure) to represent the respective parameter correlations in each data cluster and then into the whole data set. These vectors can also be categorized as singular vectors (i.e. associated with the respective singular values) for each ship performance and navigation data cluster.

Singular values quantify to the importance of each singular vector that holds the information on the respective parameter relationships. Hence, the respective singular vector structure represents to local operational and navigation conditions. That also represent an approximate linear model for each local mode of navigation and operational conditions of the vessel and ship systems. One should note that the singular values and singular vectors can be used to reduce the number of ship performance and navigation parameters by removing the least important singular vectors (i.e. a model reduction technique). Such reduction in ship performance and navigation parameters can play an important role in the data sets collected onboard, since the cost data transfer can be reduced by introducing a less number of parameters. One should note that the model reduction can also be a part of the proposed digital models. Digital models can also be classified as the linearization (i.e. piecewise linearization) of ship performance and navigation conditions around various navigational and operational points. Therefore, ship navigation and operation situations jump from one data cluster, i.e. system state, to another under a high dimensional space (see Figure 3). Digital models can also be categorized as linear models that represent local navigation and operational conditions of the vessel and ship systems. These linear models can be combined with appropriate observes to develop a highly

nonlinear ship performance and navigation models and that is categorized as predictive analytics in this study and that is further discussed in section 3. One should note that the data modeling issues (i.e. data variety) can be addressed by predictive analytics, where various forms of data can be combined into a single model. Furthermore, the structural changes in digital models under various environmental conditions and time-horizons can relate to vessel performance. The understanding of such data structural changes in digital models can further improve the predictive analytics. Since the singular values and vectors are derived from ship performance and navigation data sets, this is a reverse engineering approach that captures vessel and ship system behavior for the data sets.

3 Advanced Data Analytics

3.1 Data Anomalies



Fig. 4. Advanced analytics in shipping

The digital models support the proposed data analytics and presented in Figure 4. Five important data analytics types are proposed under this study to extract the useful information from ship performance and navigation data sets: descriptive analytics, diagnostic analytics, predictive analytics, visual analytics and decision analytics (see Figure 4). One should note that the data quality issues (i.e. data veracity) can be addressed by descriptive and diagnostic analytics, where the respective data anomalies are identified, categorized and recovered under the propose mathematical framework. In general, the descriptive analytics identifies various erroneous conditions and then the diagnostic analytics recovers/removes such conditions from ship performance and navigation data. Therefore, such data analytics creates high quality data sets and that should be further analyzed to extract useful information on vessel navigation and ship system operation behavior. Then, the predictive analytics can be used to forecast and the visual analytics can be used to visualize the information on the same vessel and ship system behavior. Such information creates advanced knowledge in vessel navigation and ship system operation conditions and that will lead to industrial intelligence. Both advanced knowledge and industrial intelligence will support the respective decision analytics. Decision analytics should consist of appropriate key performance indicators (i.e. KPIs) to evaluate the respective navigation and operational actions for various vessels and that can often relate to energy efficiency and emission control rules and regulations.

The digital models interact with these data analytics to improve the data quality and that process is presented in Figure 5. That is denoted as a data anomaly detection and recovery procedure. One should note that this procedure consists of complex interactions among digital models and data analytics, i.e. descriptive and diagnostic analytics and that consists of several steps. Firstly, the raw data send through a data anomaly detection filter 1, where missing data points and preliminary data anomalies are detected and separated. Each parameter will have possible minimum and maximum values and the values beyond that range detect as preliminary data anomalies. The data points with preliminary data anomalies send to a separate group, where the data anomalies against known and unknown sensor and DAQ faults and system abnormalities can be compared [5, 6]. Furthermore, the new information on such data anomalies can also be stored in the respective database. That is a knowledge-base that should develop, further to detect complex sensor and DAQ faults and system abnormalities. One should note these system abnormal events can relate either to vessel navigation or ship system operation conditions, therefore the domain knowledge in shipping can play an important role in this knowledge-base. The data anomalies separated from filter 1 will be directed to data anomaly group 1. The remaining data send through the digital models, where additional data anomalies can be detected (i.e. by data anomaly filter 2). The respective outliers of digital models are detected as data anomalies from this filter. The remaining data send to the cleaned data group and that can be considered as the data that have the highest data quality. However, the data that have secondary data anomalies will send to a separate group, i.e. data anomaly group 2, where

the respective data anomalies against known and unknown sensor and DAQ faults and system abnormalities can be compared, similarly. Then, the data sets from anomaly group 1 and 2 will be transferred through the data recovery filter and then digital models. A considerable amount of data anomalies can be recovered by this step. The digital models, i.e. data structure and accurate ship performance and navigation parameters, are used to estimate unknown or erroneous parameters values in this procedure. Therefore, the recovered data can be send to the cleaned data group and then the whole data set will be used under predictive analytics.

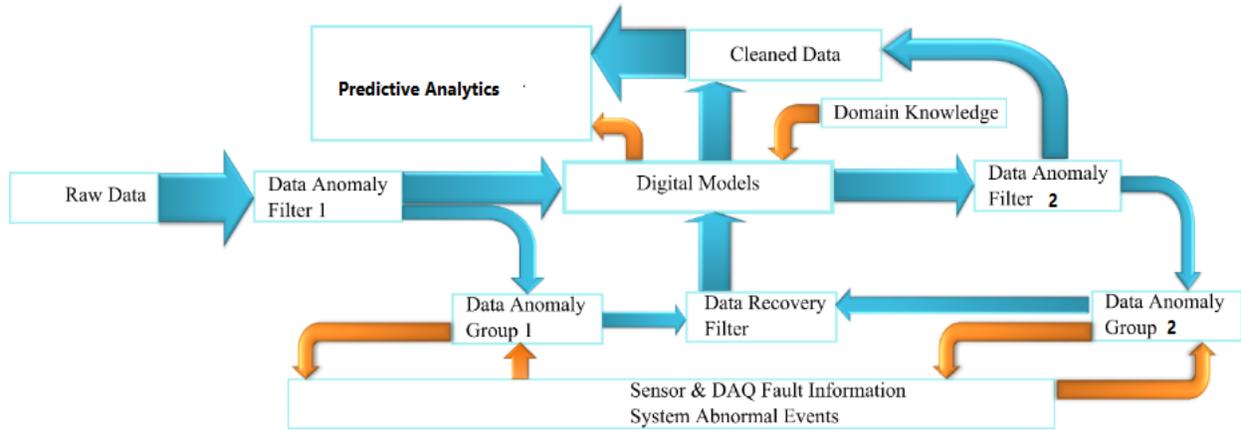


Fig. 5. Data anomaly detection and recovery procedure

3.2 Predictive Analytics

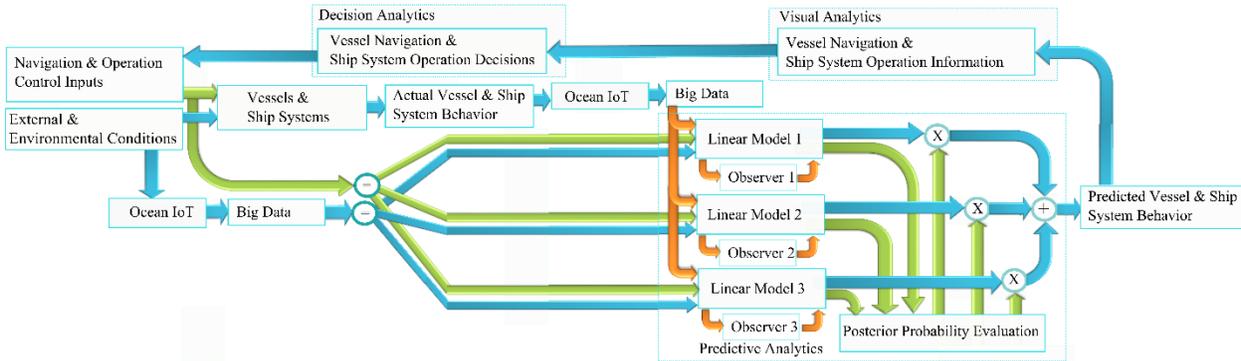


Fig. 6. Development of predictive analytics

The digital models (i.e. linear models) interact with predictive analytics to forecast vessel navigation and ship system operation behavior [7] and that process is presented in Figure 6. These models of the vessel and ship systems connect with several observers and that create the respective predictive analytics. One should note that each digital model represents a local linear model of vessel navigation and ship system operation behavior. Finally, that crates a global nonlinear model to represent vessel and ship system behavior. The inputs to the predictive analytics are the navigation and operation control inputs of the vessel and ship systems and the external and environmental conditions. The external and environmental conditions can be observed as big data sets under ocean IoT. That are also the inputs to the vessel and ship systems. The outputs of the vessel and ship systems are their actual behavior and that can also be observed as big data sets under ocean IoT.

The outputs of the predictive analytics are predicted vessel and ship system behavior. That behavior is converted to vessel navigation and ship system operation information by visual analytics. Visual analytics displays various ship performance and navigation parameter relationships [8], where the information on optimal vessel and ship system performance can be extracted. The same information creates advanced knowledge in vessel navigation and ship system operations and facilitates towards industrial intelligence in shipping (see Figure 1), as mentioned

before. Hence, advanced knowledge in vessel navigation and ship system operations can be used to develop appropriate decision analytics with key performance indicators (KPIs). Finally, the outcome of the decisions analytics influences the navigation and operation control inputs of the vessel and ship systems.

4 Conclusions

A novel mathematical framework to support industrial digitization of shipping is presented with a data flow path, i.e. from Industrial IoT to Predictive Analytics. The proposed data analytics can self-learn (i.e. the data structure can learn itself), self-clean (i.e. data anomalies can be detected, isolated and recovered by considering the outliers of the data structure), self-compress and expand (i.e. the respective parameters in the data sets can be reduced and expanded by considering the same data structure) and self-visualize (i.e. the respective data structures can be used for both vessel and ship system performance observations). Since this framework is developed from ship performance and navigation data sets, this process can also be a reverse engineering approach of vessel and ship systems. Furthermore, that introduces intelligent analytics to the shipping industry and also provides important solutions to the big data challenges in industrial digitalization.

References

1. 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.
2. L. P. Perera and B. Mo, "Emission Control based Energy Efficiency Measures in Ship Operations," *Journal of Applied Ocean Research*, vol. 60, 2016, pp. 29-46.
3. L. P. Perera and B. Mo, "Machine Learning based Data Handling Framework for Ship Energy Efficiency," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 10, 2017, pp. 8659-8666.
4. 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.
5. 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.
6. 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.
7. L. P. Perera and B. Mo, "Ship Speed Power Performance under Relative Wind Profiles," *Maritime Engineering and Technology III*, Guedes Soares & Santos (Eds.), vol. 1, Taylor & Francis Group, London, UK, 2016, pp. 133-141.
8. L. P. Perera and B. Mo, "Visual Analytics in Ship Performance and Navigation Information for Sensor Specific Fault Detection," In *Proceedings of the 36th International Conference on Ocean, Offshore and Arctic Engineering (OMAE 2017)*, Trondheim, June, 2017 (OMAE2017-61118).
9. 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*, vol. 139, no. 2, 2017.
10. L. P. Perera and B. Mo, "Machine Intelligence for Energy Efficient Ships: A Big Data Solution," *Maritime Engineering and Technology III*, Guedes Soares & Santos (Eds.), vol. 1, Taylor & Francis Group, London, UK, 2016, ISBN 978-1-138-03000-8, pp. 143-150.