Marine Engine Operating Regions under Principal Component Analysis to evaluate Ship Performance and Navigation Behavior

Lokukaluge P. Perera and Brage Mo

Norwegian Marine Technology Research Institute (MARINTEK), Energy Systems and Technical Operations, Trondheim, Norway (e-mail: prasad.perera@marintek.sintef.no, brage.mo@marintek.sintef.no).

Abstract: Marine engine operating regions under principal component analysis (PCA) to evaluate ship performance and navigation behavior are presented in this study. A data set with ship performance and navigation information (i.e. a selected vessel) is considered to identify its hidden structure with respect to a selected operating region of the marine engine. Firstly, the data set is classified with respect to the engine operating points (i.e. operating modes), identifying three operating regions for the main engine. Secondly, one engine operating region (i.e. a data cluster) is analyzed to calculate the respective principal components (PCs). These PCs represent various relationships among ship performance and navigation parameters of the vessel and those relationships with respect to the marine engine operating region are used to evaluate ship performance and navigation behavior. Furthermore, such knowledge (i.e. PCs and parameter behavior) can also be used for sensor fault identification and data compression/expansion types of applications as a big data solution in shipping.

Keywords: Principal Component Analysis, Big Data, Marine Engine Operations, Ship Performance Monitoring, Structured Data.

1. INTRODUCTION

Ship performance monitoring is with sensors and data acquisition systems facilitated to collect large data sets. These large data sets, often categorized as "big data", are analyzed in real-time to extract the information of ship performance and navigation behavior. Furthermore, the respective data handling process should consist of several improvement steps that include sensor fault identification, data compression and expansion to improve the quality of the process. These large scale data sets are often as unstructured formats considered due to various nonlinearities among ship performance and navigation parameters. Therefore, an appropriate structure should be investigated to implement the improvement steps in a real-time data handling process of these data sets. That data structure can also be used to improve the quality of the data sets under the proposed improvement steps (i.e. sensor fault identification, data compression and expansion). Hence, the results of ship performance and navigation data can further be improved.

Such unstructured data sets are associated with various conventions ship performance and navigation models to estimate their respective parameters. However, the respective parameter estimation processes may degrade due to the complexities in such empirical models and these conventions ship performance and navigation models may have difficulties in handling large data sets. Furthermore, the system-model uncertainties under ship performance and navigation parameters can further degrade the outcome of the respective estimation process. Therefore, a methodology to identify the respective structure in a selected data set of ship performance and navigation information is proposed and that may initiate the respective steps towards future ship performance and navigation models.

This study proposes to learn the respective models from large data sets of ship performance and navigation information by considering Principal component analysis (PCA) (Shlens, 2014). PCA is a non-parametric method that extracts relevant information from chaotic type data sets and reduces the initial size of the data set to improve the content visibility. In general, the respective variance values are among the parameters in a selected data set identified by PCA. These variance directions that are orthogonal represent the respective principal components (PCs) of the data set. The top and bottom PCs held the most and least important information of the data set. Therefore, the most important information of the data set accommodates to the top PCs and that is by projecting the same data set into the selected top PCs done. Therefore, the new data set is a representation and that consists of the most important information of the old data set. Therefore, the bottom PCs are often ignored during this process because that may not consist of any important information of the data set.

The descending order of the PCs represents the order of significance (i.e. the order of variance) in the data set. When the same data set is into the least PCs projected, sensor faults and other erroneous data regions are often into these least PCs separated (Perera, 2016). Therefore, the new data set that is projected into the top principal components may have less sensor faults and other erroneous data regions. The same approach can also improve the data quality, where the least PCs can be used to identify sensor faults and other erroneous data regions in the same data set. Furthermore, that information can be used to improve the quality of the data set and identify the respective faulty sensors.

The PC structure is as the basis for the respective models used and those models can be both for sensor fault identification used and data compression/expansion types of applications in shipping as a big data solution. However, a proper structure of the data set in ship performance and navigation information should be used in both situations (i.e. sensor fault identification and data compression/expansion). An inadequate structure of the data set can further degrade the outcome of the respective models and sensor fault identification and data compression/expansion steps. One should note that PCA has some limitations on finding accurate parameter relationships in some situations. This parameter inaccuracy may relate to the data point distribution and that may result in unusual parameter relationships within the data set. To overcome such challenges, a data clustering approach around main engine operating points is proposed in this study.

There are several steps to implement before PCA. In general, the respective data set should have an approximate Gaussian type distribution (i.e. appropriate mean and variance) to get appropriate results from PCA. If the data set consists of various data clusters, then that can introduce erroneous conditions in PCA. Therefore, the respective data clusters should be identified, so that PCA can be implemented for each data cluster, separately to improve results. Since marine engines of vessels are operating around various operating points, operating points are appropriate mean values for such data clusters in ship performance and navigation information. The proposed data clustering approach consists of Gaussian mixture models (GMMs) with an expectation maximization (EM) algorithm and uses to identify such marine engine operating regions (Perera and Mo, 2016a). Then, the data set should equally be centered and scaled (i.e. standardized), where each parameter is subtracted and divided by the sample mean and standard deviation values. The parameter variance related erroneous conditions can be avoided by this step of PCA. If the same parameters represent unusual relationships, then the data set should be further investigated to capture additional clustering dimensions. Therefore, a methodology to identify the respective structure in a data set of ship performance and navigation information is presented in the following sections and that can handle large data sets and implement in real-time data handling process (Perera and Mo, 2016b, c & d).

2. PRINCIPAL COMPONENT ANALYSIS

An overview of PCA is in this section presented. A ship performance and navigation data set (i.e. m number of parameters), denoted as:

$$X(t) = [x_1(t) \quad x_2(t) \quad \dots \quad x_m(t)] \tag{1}$$

where $x_1(t), x_2(t), \dots, x_m(t)$ with $x_i(t) \in R^n$ are the respective parameters of ship performance and navigation information.

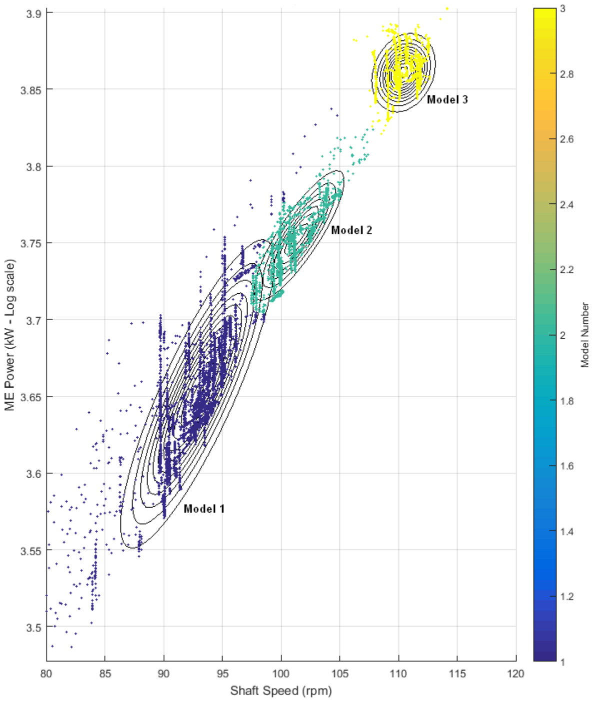

Fig. 1. Marine engine operating region classification.

The sample mean, $\bar{x}$, and variance S_x of the same data set are denoted as:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$
$$S_x = \frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})(x_i - \bar{x})^T \tag{2}$$

This data set is into a new data set transformed by considering the following transformation steps under PCA, denoted as:

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = u^T \bar{x}$$
$$S_y = \frac{1}{n}\sum_{i=1}^{n} (y_i - \bar{y})(y_i - \bar{y})^T = u^T S_x u \tag{3}$$

where $\bar{y}$ is the mean of the new data set, and S_y is the respective variance of the transformed data set and u is a unit variance vector that uses to project the old data set into the new data set. Hence, that also satisfies:

$$u^T u = I \tag{4}$$

PCA maximizes the variance of each variance direction (i.e. principal component direction) of the new data set. Hence, the trace of S_y should be maximized and denoted as:

$$\text{Max.} \;\; trace(S_y) = \text{Max.} \;\; trace(u^T S_x u) \tag{5}$$

The Lagrange multiplier that satisfy (5) can be written as:

$$L = trace(u^T S_x u) = \sum_{i=1}^{n} \left[u_i^T S_x u_i + \lambda(1 - u_i^T u_i)\right] \tag{6}$$

The derivatives of the Lagrange multiplier in (6) can be resulted in:

$$S_x u_i = \lambda_i u_i \\ u_i^T u_i = 1 \quad (7)$$

where λ_i is the eigenvalues and u_i is the respective eigenvectors of S_x. One should note that the largest and smallest eigenvalues represent the top and bottom PCs, respectively.

3. DATA ANALYSIS

A data set of ship performance and navigation information is of a selected vessel considered. The respective parameter can be summarized as: average (avg.) draft (m), speed through water (STW) (Knots), main engine (ME) power (kW), shaft speed (rpm), main engine (ME) fuel consumption (cons.) (Tons/day), speed over ground (SOG) (Knots), trim (m), relative (rel.) wind speed (m/s) and direction (deg) and auxiliary (aux.) fuel consumption (cons.) (Tons/day). The respective data set is from a bulk carrier collected with following particulars: ship length: 225 (m), beam: 32.29 (m), gross tonnage: 38.889 (tons), deadweight at max draft: 72.562 (tons). The vessel is powered by 2 stroke ME with maximum continuous rating (MCR) of 7564 (kW) at the shaft rotational speed of 105 (rpm). Furthermore, the vessel has a fixed pitch propeller diameter 6.20 (m) with 4 blades.

As the first step of this process, are the respective clusters of the data set of ship performance and navigation information investigated. An approach based on Gaussian mixture models and an expectation-maximization algorithm is in this step implemented to cluster the respective data set in ship performance and navigation information. Three specific operating points of the marine engine is identified in this data set and an overview of the approach is presented in Perera and Mo (2016a). The results (i.e. the respective data clusters) are in Figure 1 presented as an engine power (kW-log scale) and shaft speed (rpm) diagram. One should note that the respective data clusters are by Model 1, 2 and 3 denoted. These data clusters relate to three operating points related of the marine engine (i.e. vessel operating points). It is expected that other parameters in ship performance and navigation information should also relate to the same engine operating points.

The third data clusters (i.e. model 3) is considered for PCA. The calculated PCs are in Figure 2 and the i-th PC is denoted as:

$$Z_i = \begin{bmatrix} z_{i,1} & z_{i,2} & z_{i,3} & z_{i,4} & z_{i,5} & z_{i,6} & z_{i,7} & z_{i,8} & z_{i,9} & z_{i,10} \end{bmatrix} \quad (4)$$

where $z_{i,1}, z_{i,2}, z_{i,3}, .z_{i,10}$ represent the respective vector components of the i-th PC. One should note that the top and bottom PCs are Z_1 and Z_{10}, respectively. Hence, the respective vector components of each PC are further in the next step investigated by appropriate data visualization and the results are presented in Figure 2. One should note that this figure represents a 10 dimensional vector space, where the respective PCs (i.e. eigenvectors) are in a polar coordinate

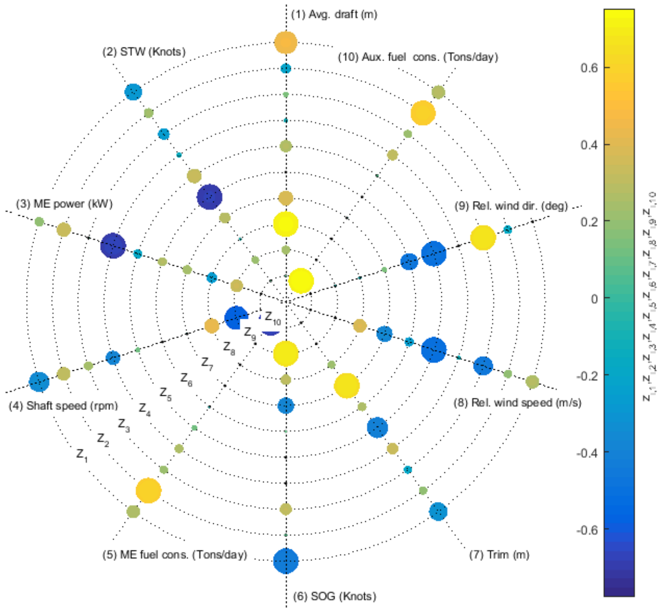

Fig. 2. PCs for Model 3.

system demonstrated. Each PC is by a dotted circle presented, where the top PC has the highest radius. Each axis that is intersecting these circles represents a parameter from the data set of ship performance and navigation information. The respective vector components of each PC are by colored circles presented and the same circle radius represents the significance of that component with respect to other components within the same PC. This figure also represents an overview of the correlations among the respective parameters of ship performance and navigation information. A summary of these correlations among the respective parameters in the selected region of the marine engine (i.e. Model 3) is presented in Table 1.

	$z_{i,1}$	$z_{i,2}$	$z_{i,3}$	$z_{i,4}$	$z_{i,5}$	$z_{i,6}$	$z_{i,7}$	$z_{i,8}$	$z_{i,9}$	$z_{i,10}$
Z_1	**HP**	**MN**	LP	**MN**	LP	**HN**	**MN**	LP	L	LP
Z_2	LN	LP	**MP**	**MP**	**HP**	L	LP	LP	LN	**HP**
Z_3	L	LN	L	LP	LP	LP	LN	**MN**	**HP**	LP
Z_4	L	L	**HN**	**MN**	LP	L	LP	L	L	LP
Z_5	LP	LP	L	L	L	L	**MP**	**HP**	**HP**	L
Z_6	L	**HN**	LP	L	L	L	LP	LN	**MN**	L
Z_7	**MP**	LP	LP	L	L	**MN**	**HP**	**MN**	L	L
Z_8	**HP**	L	LN	**MP**	L	LP	L	LP	L	L
Z_9	L	L	LP	**HN**	L	**HP**	L	L	L	L
Z_{10}	L	L	L	L	**HN**	L	L	L	L	**HP**

Table 1: PCs with parameter relationships

This table consists of the respective PCs and their vector components. Furthermore, the following notations are to capture the correlations among the respective parameters within each PC introduced: HP: high positive, MP: medium positive, LP: low positive, HN: high negative, MN: medium negative, LN: low negative and L: low. One should note that HP represented by yellow color large circles and HN represented by blue color large circles in Figure 2. Therefore, the same notation system is used to evaluate ship performance and navigation behavior with respective the

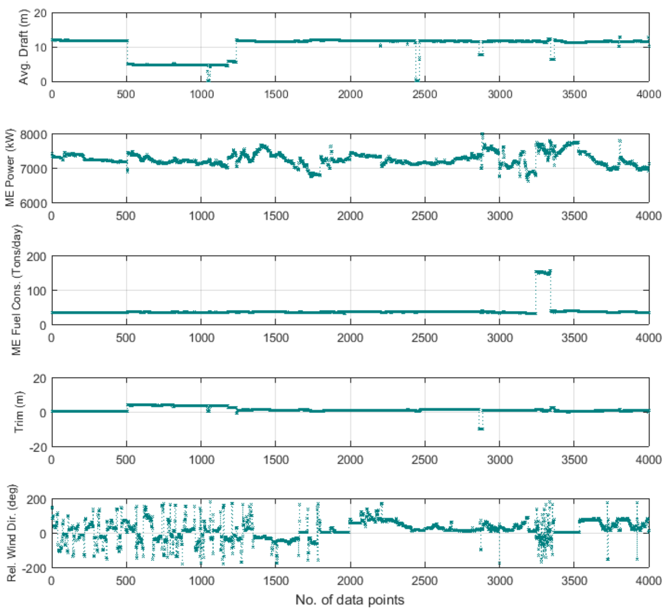

Fig. 3. Ship performance and navigation data.

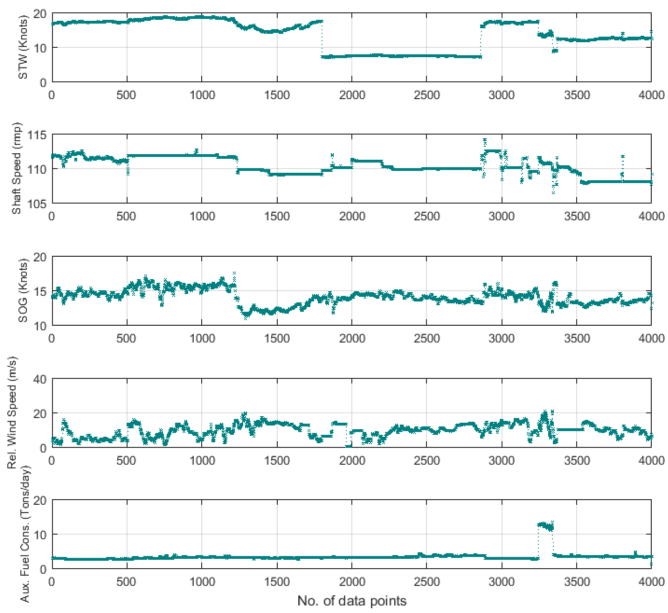

Fig. 4. Ship performance and navigation data.

derived PCs. Furthermore, these results are also compared with respective parameters as a time series data set and the results are presented in Figures 3 and 4. One should note that the x-axis of these figures presents a number (no.) of the data points and the time duration between two consecutive data points is 15 (min).

The respective PCs with their vector components are in this section further investigated (see Figure 2 and Table 1). The 1st PC represents: when avg. draft (high) increases (HP), STW (medium) decreases, shaft speed (medium) decreases, SOG (high) decreases, and trim (medium) decreases. The 1st PC shows that ship resistance has increased due to the draft increments, where STW and SOG of the vessel are also decreased. The same conditions have decreased shaft speeds due to high engine loading conditions. Furthermore, the draft increments are compensated by the trim adjustments of the vessel and that is by this PC also noted. These results can by the ship performance and navigation data be verified that are presented in Figure 3 and 4. In general, when vessel avg. draft increases, then trim increases, STW decreases, and SOG decreases.

The 2nd PC represents: when engine power and shaft speed (medium) increases, ME fuel consumption (high) increases, and aux. fuel consumption (high) increases. The 2nd PC shows a moderate increment in engine shaft speed increases engine power levels, moderately and fuel consumption in both main and auxiliary engines, significantly. This results shows that shaft speed increments beyond the mean operating point in this engine operating region may not increase engine power, considerably but that may increase the respective fuel consumption, significantly. Similarly, these results can be by the results that presented in Figure 3 and 4 verified. In general, when engine power and shaft speed increases, ME fuel consumption increases, and aux. fuel consumption increases. However, unusual data regions (i.e. high data values) in both ME and AE fuel consumption plots. These data regions are also influenced on the respective PCs. It is believed that these regions are accumulated due to sensor erroneous conditions, therefore that (i.e. the erroneous data regions) should be removed from the data set to improve the validity of the respective PCs.

The 3rd PC represents: when rel. wind speed (medium) decreases, then rel. wind angle (high) increases. Therefore, the 3rd PC shows that when the vessel increases its speed, then rel. wind speed increases and rel. wind angle decreases (i.e. the vessel encounters high head wind conditions with the speed increments). Similarly, these results can be by the data presented in Figure 3 and 4 verified. The fourth PC represents: when ME power (high) increases, then shaft speed (medium) increases. Therefore, the forth PC shows, the shaft speed increments increase engine power and the results can be verified by the data that are presented in Figure 3 and 4. The 5th PC represents: when vessel trim (medium) increases, then relative wind speed (high) and direction (high) increase. The 5th PC shows that the trim values are used under calm water conditions, where relative wind speed is slower and angle is higher. A larger wind angle represents the vessel is moving in moderate or slow speeds, therefore the vessel does not encounter any high head winds.

The 6th PC represents: when STW (high) decreases, then relative wind angle (medium) decreases. The 6th PC shows, a positive correlation between STW and relative wind direction of the vessel and that relationship is similar to the previous PC. The same results can be verified by the data that are presented in Figure 3 and 4. The 7th PC represents: when avg. draft (medium) increases, then SOG (medium) decrease, trim (high) increases, and rel. wind speed (medium) decreases. Similarly, the 7th PC shows that ship resistance has increased due to the draft increments, therefore SOG is also decreased. The same conditions have reduced rel. wind speed, as discussed previously. Furthermore, the draft

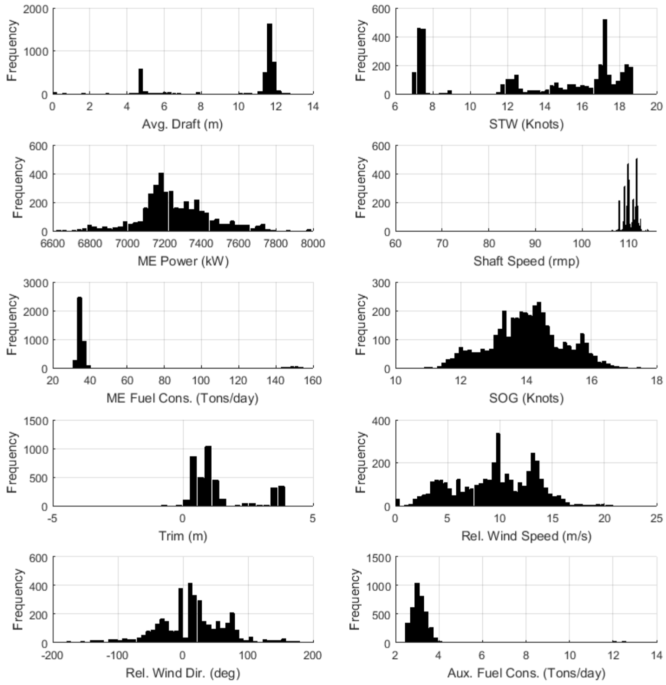

Fig. 5. Statistical distribution of the respective parameters.

increments are compensated by the trim variations under slow maneuvering conditions of the vessel.

The 8th PC represents: when avg. draft (high) increases, then shaft speed (medium) increases. The 8th PC shows that ship resistance has increased due to the draft increments, therefore shaft speed is increased to compensate the speed losses in the vessel. Similarly, these results can be by the data presented in Figure 3 and 4, verified. The 9th PC represents: when shaft speed (high) decreases, then SOG (high) decreases. The 10th PC represents: when ME fuel consumption (high) decreases, then aux. fuel consumption (high) increases. The bottom PCs may not represent any useful information about the respective parameter relationships as mentioned before. Therefore, a proper interpretation for the bottom PCs should not be expected. Furthermore, that can accumulate data erroneous conditions of ship performance and navigation information, therefore such parameter relationships should be ignored.

The low positive and negative correlations among the respective parameters are from the above discussion ignored. However, those effects (i.e. low positive and negative correlations) should also be incorporated in the respective parameter relationship to see an overall picture of ship performance and navigation information. However, that can complicate the outcome of the respective PCs. A majority of these PCs can be with the data plots in Figures 3 and 4 verified. However, some erroneous data regions have also influenced in these results and should, as mentioned before be removed. Therefore, the respective tools to identify and remove such data erroneous situations to be developed, to that improve the accuracy of the respective PCs.

As the next step of this process, the time series data presented in Figures 3 and 4 are as statistical distributions grouped, and the results are in Figure 5 presented. A majority

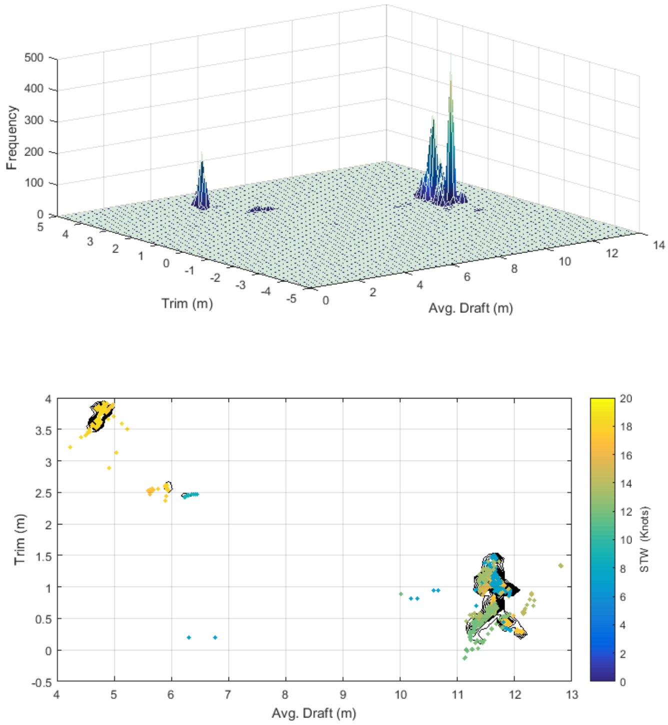

Fig. 6. Avg. draft and trim configuration of the vessel.

of the parameters in ship performance and navigation information shows single statistical distributions. However, the average draft and trim values of the vessel consist of two separate distributions. These trim and draft configurations are further studied and the combined statistical distribution for the same parameters is in the top plot of Figure 6 presented. The figure shows that these two distributions relate to each other, when the draft values are approximately 11-12 (m) and 4-5 (m), then the trim values are approximately 0-1.5 (m) and 3.5-4 (m), respectively. Therefore, the vessel is operating around specific avg. draft-trim combinations and that is as several data clusters observed. The respective contour plot (i.e. a top view) of the same plot with vessel STW is in the bottom of Figure 6 presented.

One should note that the vessel is having high STW under low draft conditions and low STW under high draft conditions due to ship resistance effects. However, additional less frequent avg. draft-trim regions of the vessel can be in the same plot noted. One should note that these regions may consider as transition regions, where the vessel operation conditions shift from one operating point (i.e. the avg. draft or trim value) to another. It is believed that these transition regions may not influence significantly on overall ship performance and navigation behavior. It is concluded that the same data set should be further clustered with respect to vessel avg. draft-trim conditions to further improve the respective PCs. Hence, the data clustering approach that is under marine engine-operating regions presented should also be with vessel avg. draft-trim configurations incorporated. One should note that this combination creates 6 data clusters that relate to vessel operation conditions. Such data clusters may create different set of PCs and that can further improve the knowledge of ship performance and navigation behavior. That (i.e. PCs and parameter behavior) further improves onboard sensor fault identification and data compression/expansion types of applications in vessels and that is as the future work of this study considered.

5. CONCLUSION

Marine engine operating regions under principal component analysis (PCA) are by considering a data set of a selected vessel to evaluate ship performance and navigation behavior studied. The data set with ship performance and navigation information is in this analysis to identify its hidden structure with respect to a selected operating region of the marine engine used. That consists of several steps. Firstly, the data set is with respect to the engine operating points classified, that identifies three engine operating regions. The proposed data clustering approach consists of Gaussian mixture models (GMMs) with an expectation maximization (EM) algorithm. Then, one engine operating region is selected (i.e. a data cluster) to calculate the respective principal components (PCs).

These PCs represent various parameter relationships among ship performance and navigation information, therefore ship performance and navigation behavior within a selected marine engine operating region is in this study discussed. However, it is also noted that the selected region of marine engine operations can be further divided by considering vessel avg. draft and trim conditions. Therefore, the clustered data set can be further clustered by considering the respective avg. draft and trim conditions. The vessel is operating around specific avg. draft and trim values within this engine-operating region and that is the main reason of this hidden data structure. Therefore, identifying such hidden structures within ship performance and navigation data sets is also play an important role in the proposed approach.

The main advantage in this approach is that ship performance and navigation behavior can be observed from with respect to engine/propeller operating regions and then avg. draft/trim regions. Therefore, this approach can also be considered as the required development steps towards data driven ship performance and navigation models by considering the respective vessel operating points (i.e. engine operating points and trim-draft operating points). Hence, the respective models are linearized around each of these vessel operating regions (i.e. engine/propeller operating regions and the avg. draft/trim regions) by this approach. One should note the parameter relationships in ship performance and navigation data can vary due to these vessel operating regions and that are represented by various vectors (i.e. PCs).

However, that can facilitate a piecewise linear function capture the nonlinear behavior of ship speed power performance and navigation capabilities. PCA with ship performance and navigation data can be an important part of such nonlinear functions and that is proposed as the future work of this study. This extended knowledge (i.e. PCs and parameter behavior) of the respective data set of ship performance and navigation information can be for sensor fault identification and data compression/expansion types of applications as a big data solution (Rodseth et al., 2016) in the shipping industry used.

6. ACKNOWLEDGEMENT

This work has been 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 conducted.

REFERENCES

Perera, L.P., (2016) "Marine Engine Centered Localized Models of Ship Performance Monitoring for Sensor Fault Detection and Isolation," In Proceedings of the 3rd IFAC Workshop on Advanced Maintenance Engineering, Service and Technology, (AMEST'16), Biarritz, France, October, 2016.

Perera, L.P. and Mo, B., (2016a), "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 (OMAE2016-54168).

Perera, L.P. and Mo, B. (2016b) "Ship Speed Power Performance under Relative Wind Profiles," In Proceedings of the 3rd International Conference on Maritime Technology and Engineering (MARTECH 2016), Lisbon, Portugal.

Perera, L.P. and Mo, B., (2016c) " Machine Intelligence for Energy Efficient Ships: A Big Data Solution," In Proceedings of the 3rd International Conference on Maritime Technology and Engineering (MARTECH 2016), Lisbon, Portugal.

Perera, L.P. and Mo, B., (2016d), "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, (OMAE2016-54093).

Shlens, J. (2014) "A tutorial on principal component analysis," arXiv preprint arXiv:1404.1100.

Rodseth, O.J., Perera, L.P., and Mo, B., (2016),"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, pp. 361-373.