Marine Engine Centered Localized Models for Sensor Fault Detection under Ship Performance Monitoring

Lokukaluge P. Perera

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

Abstract: Sensor fault detection under marine engine centered localized models of an engine propeller combinator diagram is presented in this study. The proposed approach consists of two detection levels to identify of sensor fault situations in an onboard data acquisition system of a vessel. Each parameter in ship performance and navigation data can have a realistic data range (i.e. a threshold relates to the variance), where the parameter can vary. If the sensor reads a value beyond this parameter range, then that data point is categorized as a sensor fault situation by the first fault detection level. However, some sensor faults are located within this data range and that cannot identify by this detection level. Such complex sensor fault situations are detected by the second fault detection level by considering the proposed localized models. These localized models are derived with respect to the operating regions of an engine-propeller combinator diagram, where the respective data points are clustered by Gaussian mixture models with an expectation maximization algorithm. Each data cluster is examined through principal component analysis and projected into the bottom principal component to identify such complex sensor fault situations. A data set of ship performance and navigation information of a selected vessel is used through these sensor fault detection levels and the successful results on identifying such sensor fault situations are also presented in this study.

Keywords: Sensor Fault Detection, Principal Component Analysis, Big Data, Statistical Data Analysis, Gaussian Mixture Models, Engine-Propeller combinator Diagram.

1. INTRODUCTION

Modern integrated bridge systems (IBSs) are equipped with various sensors and data acquisition (DAQ) systems to monitor vessel performance and navigation information. Those systems collect large quantities of ship performance and navigation data and analyze to observe optimal vessel operation and navigation conditions. However, these large scale data sets consist of various sensor related erroneous intervals and that may degrade the results of the respective data analyses. If such erroneous data intervals are detected in an early stage of the data handling process, that can be removed to improve the quality of the data set. Furthermore, the respective faulty sensors can also be detected in such situations and the required maintenance actions can be taken. That step eventually improves the quality of the collected data sets and the final results of the data analyses. This study proposes a sensor fault detection structure in an onboard DAQ systems of a vessel consisting of several fault detection levels.

There are several sensor fault detection approaches with respect to ship performance and navigation information under various DAQ systems are presented in the recent literature (Lajic and Nielsen (2009) and Lajic et al. (2009)). These studies often depend on various mathematical models that relate to ship kinematics and dynamics. However, the accuracy of such models can be challenged under various navigation situations and the respective model performance can also degrade under large scale data sets. A model learning methodology to detect sensor faults situations is considered in this study and that consists of identifying the respective mathematical models from the respective data set of ship performance and navigation information. Hence, this approach has the capability to handle large scale data sets and that is the main contribution of this study.

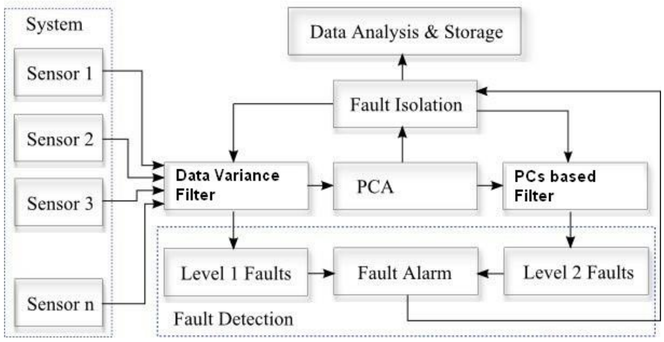

Fig. 1. Sensor fault detection structure

An overview of the proposed sensor fault detection structure is presented in Figure 1. This structure is developed by considering the respective studies of ship performance and navigation monitoring (Perera and Mo, 2016a, 2016b, 2016c, 2016d) and that consist of various data analysis and visualization tools and techniques. The figure consists of two layers to identify of the respective sensor fault situations in an onboard DAQ system. The real-time data collected by the respective sensors transfer through these two layers. The first layer consists of a data variance filter (i.e. sensor faults level 1) that is attached to the respective fault alarm. The second layer consists of a principal components (PCs) based filter. Principal component analysis (PCA) (Sperduti, 2013) is used with the respective data set to derive this filter.

PCA is a non-parametric method for extracting relevant information from a chaotic type data set, where the respective structure of the data set can be identified. The same structure can be used to reduce the size of the data set and that also improves the content visibility. This structure is considered as the new basis of the same data set and that consists of a linear combination of the original basis (i.e. the initial parameters of the data set). This new basis has the same dimensions as the original data set and represents by the respective principal components (PCs). Various parameter relationships (i.e. correlation, covariance, dependence etc.) can also be observed under PCA. The largest to smallest variance directions in a data set are represent from the top to bottom PCs. The top PCs consist of the most important information of the data set. Hence, the top PCs in a data set select to represents the entire data set in some situations, where the least important (i.e. the bottom) PCs are neglected (i.e. data compression).

It is also noted that erroneous data intervals can often be observed under the bottom PCs because such regions are projected far beyond the respective parameter variance values of the bottom PCs. Hence, a majority of sensor fault situations can often be detected by the bottom PCs as further described in the second level sensor faults (i.e. level 2). Then, the fault alarm executes the fault isolation procedure, where these erroneous data intervals should filter to improve the quality of the respective data set (Perera and Mo, 2016c). Finally, the cleaned data set will transfer for data analysis and storage facilities for further processing. These sensor fault levels (i.e. level 1 and 2) are further discussed in the following sections.

2. SENSOR FAULT DETECTION

Two sensor fault levels are introduced in this study: level 1 & 2. Two types of sensor fault situations are detected under level 1: i) the repeated data points and ii) the data points beyond the selected thresholds that relate to the variance values of the respective parameters. It is noted that sensor and DAQ systems may crate data intervals with repeated values (i.e. frozen data intervals) in real-time data handling processes (Perera and Mo, 2016a). These frozen data intervals are detected by observing the repeated data values and should remove from the respective data set. e.g. it is noted that wind sensors in vessels may repeat some data values due to high vibration conditions under rough weather navigation situations. In general, sensors should not repeat such values due to measurement noise and that often approximates to white Gaussian distributions. However, this type of faults (i.e. repeated data points) can occur either due to sensor or DAQ system faulty situations.

Each parameter in ship performance and navigation information has a realistic data range where the parameter can vary. This respective range can be derived either from maximum and minimum values or selected threshold values that relate to the variances of each parameter. If the sensor reads values beyond this parameter range, then those data points are categorized as sensor fault situations. This simple concept is used in this level to identify another sensor fault situation and that is categorized as the data variance filter. The respective parameters of ship performance and navigation information of a selected vessel are presented in Table 1. The table consists of minimum and maximum values of each parameter and these parameter ranges are used to identify sensor fault situations under the same fault level. However, appropriate threshold values can also be introduced to identify the decision boundaries that relate to the variance values (i.e. minimum and maximum values) of fault level 1.

	Parameter	Min.	Max.
1.	Avg. draft (m)	0	15
2.	STW (Knots)	3	20
3.	ME power (kW)	1000	8000
4.	Shaft speed (rpm)	20	120
5.	ME fuel cons. (Tons/day)	1	40
6.	SOG (Knots)	0	20
7.	Trim (m)	-2	6
8.	Rel. wind speed (m/s)	0	25
9.	Rel. wind direction (deg)	2	360
10.	Aux. fuel cons. (Tons/day)	0	8

Table 1: Ship performance and navigation parameters

More complex sensor fault situations undetected by level 1 are captured by level 2. This step is named as a PCs based filter designed under PCA as mentioned before. An overview of such filter design is presented in this section. A two sensor situation, where two parameters are measured by sensors in a DAQ system, is presented in Figure 2 to explain such complex fault situations. Two parameters that are measured by two sensors with a selected sampling period are denoted as Y_1 and Y_2. The actual parameter values with sensor noise are presented by the respective shaded regions in the Y_1-time and Y_2-time plots. The measured values (i.e. sensor measurements) are denoted by "x" in this figure. The respective variance for each parameter is presented in blue oval shapes (i.e. next to Y_1 and Y_2 axes) in each plot. Four sensor fault situations, beyond actual measurements, are introduced and denoted as e_1, e_2, e_3 and e_4. One should note that sensor faults e_1 and e_3 can be detected by fault level 1 because those two values are beyond the thresholds (i.e. relate to the variance) of each parameter. However, sensor faults e_2 and e_4 cannot be detected by fault level 1, therefore fault level 2 is introduced to capture such events. The respective data set without timestamp is presented in the Y_1-Y_2 plot and that is used for PCA.

It is assumed that both parameters have a positive correlation as presented in Figure 2 and this relationship is visible in the Y_1-Y_2 plot. Such parameter relationships among

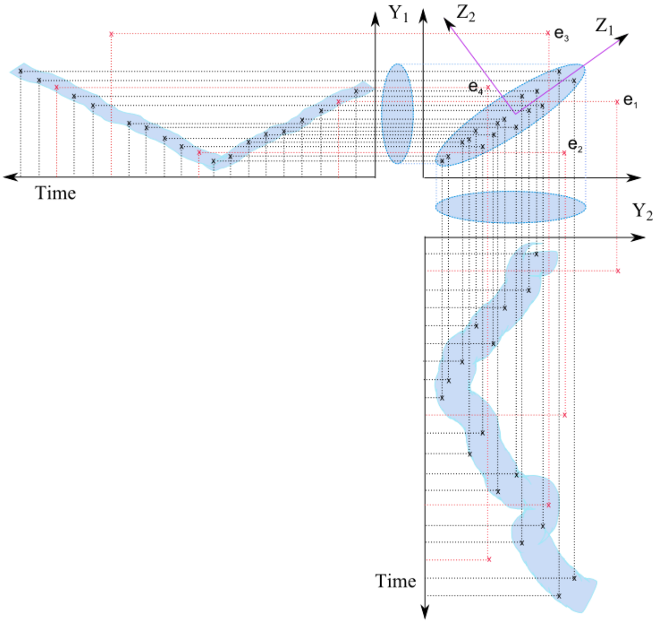

Fig. 2. Two parameter data set under PCA

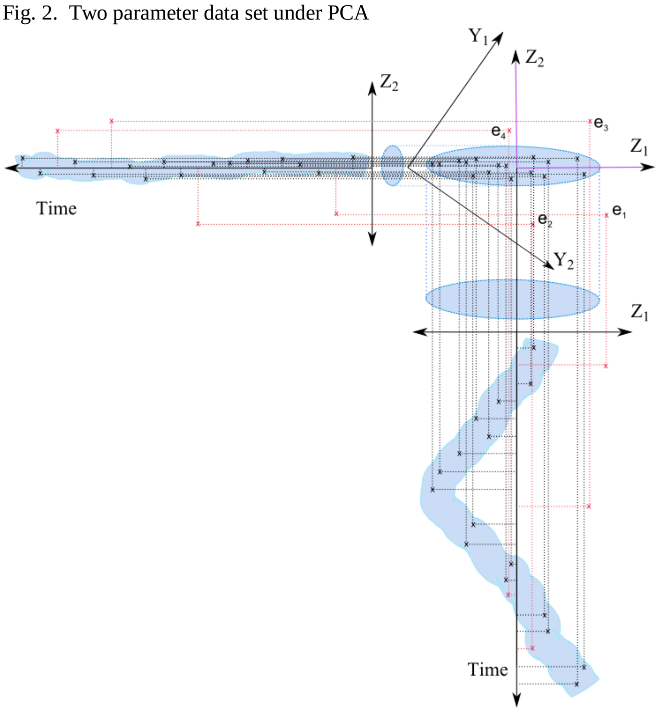

Fig. 3. Two parameters projected into PC axes

ship performance and navigation information should be derived to identify second level sensor faults. However, these parameter relationships should be localized to improve the accuracy of the sensor fault detection structure. In the next sept, the PCs of the two parameter data set should be calculated to identify the respective data structure. The results are presented as Z_1 and Z_2 in the same figure. One should note that Z_1 (i.e. the top PC) represents the largest covariance direction and Z_2 (i.e. the bottom PC) represents the second largest covariance direction that is normal to Z_1.

Since Z_2 is the bottom PC, it is expected that the respective sensor fault situations can often be observed under this axis. This can be done by rotating the same data set into

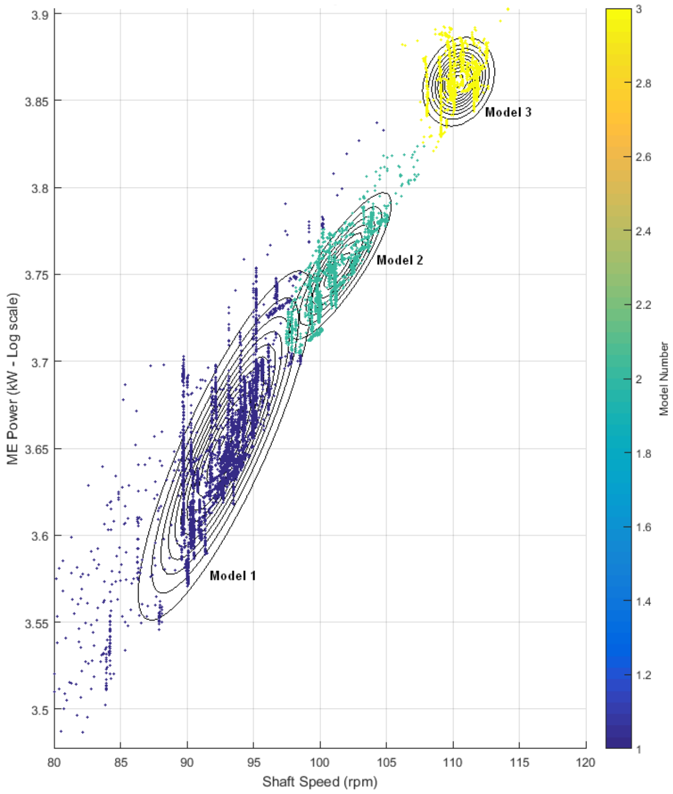

Fig. 4. Marine engine combinator diagram with localized models

the respective PC directions and the results are presented in Figure 3.

Each principal component has its variance value that uses to derive the respective threshold value (i.e. an appropriate decision boundary) for sensor fault detection. The variance related threshold values are presented as blue oval shapes (i.e. next to Z_1 and Z_2 axes). It is expected that the respective sensor faults should be projected beyond these thresholds and the results are also presented in Figure 3. As presented in the figure, all sensor fault situations are projected beyond the threshold value of Z_2 (i.e. the bottom PCs). Sensor fault situations undetected by level 1 are detected by this level. Hence, this study proposes to use the bottom PC to identify an appropriate set of sensor fault situations from a data set of ship performance and navigation information by fault level 2. One should note that this example presents a two parameter situation, however a multi-parameter situation with various sensor faults may need additional tools to identify such situations. However, the success of this approach relates to the mathematical models that represent the relationships among the respective ship performance and navigation parameters.

These parameter relationships are categorized as mathematical models and the accuracy of such models in ship performance and navigation information influences on the proposed fault detection process (i.e. fault level 2). It is believed that a single mathematical model (i.e. ship performance and navigation information) is inadequate to capture realistic ship navigation under various seakeeping conditions. Hence, a multiple model approach with ship performance and navigation data is considered and the respective models are derived with respect to an engine propeller combinator diagram of a selected vessel. Various data clusters in the engine propeller combinator diagram represent such models and called as "localized models."

3. LOCALIZED MODELS

The model development steps under a ship performance and navigation data set are presented in this section. Marine engine centered localized models are derived, where an engine-propeller combinator diagram of a selected vessel is considered as the basis for such model development. The respective combinator diagram with engine propeller operating data is presented in Figure 4. Both main engine (ME) power (kW – log scale) and propeller shaft speed (rpm) values are presented in this diagram. It is noted that three operating regions are frequently used by this vessel, therefore Gaussian mixture models (GMMs) with an expectation maximization (EM) algorithm is used to identify such regions (Perera and Mo, 2016d). These operating regions are classified as localized models and that use to create the respective data clusters. An overview of GMMs with an EM algorithm is presented in the following section. The data point assignment for each GMM that represents a frequent operating region of the engine-propeller combinator diagram is done by an EM algorithm. The EM algorithm is used to calculate the respective model parameters of GMMs and consists of two iterative levels: expectation and maximization. In the expectation step (i.e. E-step), the probability that each data point belongs to the respective data cluster is evaluated. In the maximization step (i.e. M-step), that data point is accommodated in the respective data cluster that has the highest probability by updating its mean and covariance values. This method assigns each data point exactly to one operational region (i.e. a GMM) of the engine propeller combinator diagram. Therefore, the boundaries of each operating region of the engine-propeller combinator diagram are determined and these regions classify localized models. The E-step is initiated by considering a multivariate GMM and denoted as (Perera and Mo, 2016d):

$$p_j(x;\mu_j,\Sigma_j)=\frac{1}{(2\pi)^{\frac{n}{2}}\left|\Sigma_j\right|^{\frac{1}{2}}}e^{-\frac{1}{2}(x-\mu_j)^T\Sigma_j^{-1}(x-\mu_j)} \tag{1}$$

where x is the input data set and $p_j(x;\mu_j,\Sigma_j)$ is the PDF of a multivariate Gaussian distribution with, μ_j and Σ_j, the mean and covariance values and n is the number of data points of the j-th data cluster, respectively. The probability of i-th data point belongs to j-th cluster can be written as:

$$w_j^{(i)}=p\left(z^{(i)}=j\mid x^{(i)};\phi,\mu,\Sigma\right) \tag{2}$$

One should note that (2) calculates the "soft guess value" for the parameter, $z^{(i)}$. Considering the Bayes rule and (1), the posterior probability of the parameter, $z^{(i)}$ given the parameter, $x^{(i)}$, can be written as:

$$p\left(z^{(i)}=j\mid x^{(i)};\phi,\mu,\Sigma\right)=\frac{p\left(x^{(i)}\mid z^{(i)}=j;\mu,\Sigma\right)p\left(z^{(i)}=j;\phi\right)}{\sum_{l=1}^{k}p\left(x^{(i)}\mid z^{(i)}=l;\mu,\Sigma\right)p\left(z^{(i)}=l;\phi\right)} \tag{3}$$

where $p\left(z^{(i)}=j;\phi\right)$ is the prior probability of the j-th data cluster and k is the number of data clusters. The equal prior probability of each data cluster is assumed, initially. One should note that (3) represents a multivariate Gaussian distribution with μ_j and Σ_j are the mean and covariance values, respectively. The M-step can be written as:

$$\begin{aligned}\phi_j&=\frac{1}{n}\sum_{i=1}^{n}w_j^{(i)}\\ \mu_j&=\sum_{i=1}^{n}w_j^{(i)}x^{(i)}\Big/\sum_{i=1}^{n}w_j^{(i)}\\ \Sigma_j&=\sum_{i=1}^{n}w_j^{(i)}\left(x^{(i)}-\mu_j\right)\left(x^{(i)}-\mu_j\right)^T\Big/\sum_{i=1}^{n}w_j^{(i)}\end{aligned} \tag{4}$$

This step updates the respective data cluster (i.e. a GMM) by calculating the new mean and covariance values with respect to each data point. This iterative process should stop either at the end of the training data set or the approximately stable prior and posterior mean and covariance values.

The results (i.e. the GMMs) that relate to each operating region are presented as contours in Figure 4 and the respective data clusters that belong to each model are also presented in the same. Three models (i.e. Model 1, 2 and 3) are presented in three different colors and categorized as localized models in the engine propeller combinator diagram. Model 3 is selected for PCA. One should note that PCA is used in this step to identify the respective structure of the clustered data set (i.e. model 3). Even though model 3 is derived by considering two parameters, the actual data set of ship performance and navigation information consists of 10 parameters and that are presented in Table 1. The rel. wind direction is adjusted during this model development, therefore relative wind from starboard and port sides represent by 0 to 180 (deg) and 0 to -180 (deg), respectively.

4. PRINCIPAL COMPONENT ANALYSIS

In PCA, the most important feature in a data set is considered as the variance values among the respective parameters. These data variance directions (i.e. coordinate system) represent the respective PCs. The same data set often projects into the new coordinate system (i.e. change the basis of the data set) to observe the most important feature (i.e. the variance) in some situations. One should note that the change of basis will not change the data quality, but the representation of the data set. Therefore, the same data set can represent under a difference coordinate system (i.e. PCs) in this approach. However, this study proposes to use this new representation of the data set to identify the respective sensor fault situations. Furthermore, sensor noise and parameter redundancy situations in the old data set can be observed by this coordinate transformation.

The descending order of the PCs represents the order of significance (i.e. the order of variance) in the data set. Hence, the top PCs are used in various industrial and research applications, since that consist of the most important information of the respective data set. In general, the bottom PCs are ignored by such applications, since that consist of the

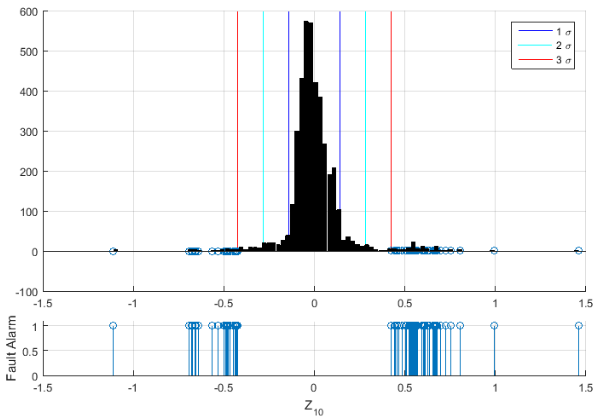

Fig. 5. The data set projected into the bottom PC.

least important information of the respective data set. The importance in the bottom PCs is illustrated in this study and that consists of identifying various sensor fault situations.

Model 3 (i.e. the clustered data set) in the engine propeller combinator diagram (see Figure 3) is considered for PCA as the next step. The clustered data set is equally centered and scaled (i.e. standardized), where each parameter is subtracted and divided by the respective sample mean and standard deviation values. Hence, each parameter can influence equally because of this step and the standardized data set is used to derive the respective PCs. The bottom PC is considered in the next step to identify the respective sensor fault situations. The respective histogram of the data set that is projected into the bottom PC (i.e. Z_{10}) is presented in the top plot Figure 5. One should note that this data distribution has the lowest variance value with compared to other PCs. Furthermore, the respective standard deviation values of σ, 2σ and 3σ are also marked in the same figure. The threshold value of 3σ is considered as the respective limit for sensors fault situations.

Therefore, the data points that are projected beyond this threshold value are considered as sensor fault situations and presented in the same figure. The zoomed view of the same data points is presented in the bottom plot and sensor fault situations are denoted as discrete pulses. Finally, these sensor faults are compared with the respective parameters of ship performance and navigation information as a time series to evaluate the success of the proposed approach. The results are presented in Figure 6, where the respective parameters (i.e. 10 parameters) are presented in the top 10 plots and the detected faults situations (i.e. fault alarm) are presented in the bottom plot. One should note that these plots are presented with respect to the number (No.) of data points (i.e. another representation of the time series) and the time interval between two consecutive data points is 15 (min). Two sensor faults situations are detected in this data set and that are framed by two windows.

In the first sensor fault situation, several parameters (ME power, ME fuel consumption, STW, shaft speed, and auxiliary engine fuel consumption) represent some unusual behavior (i.e. a sudden drop in the parameter value) and that situation is detected by the PCs based filter (i.e. fault level 2). One should also note that these types of multiple sensor

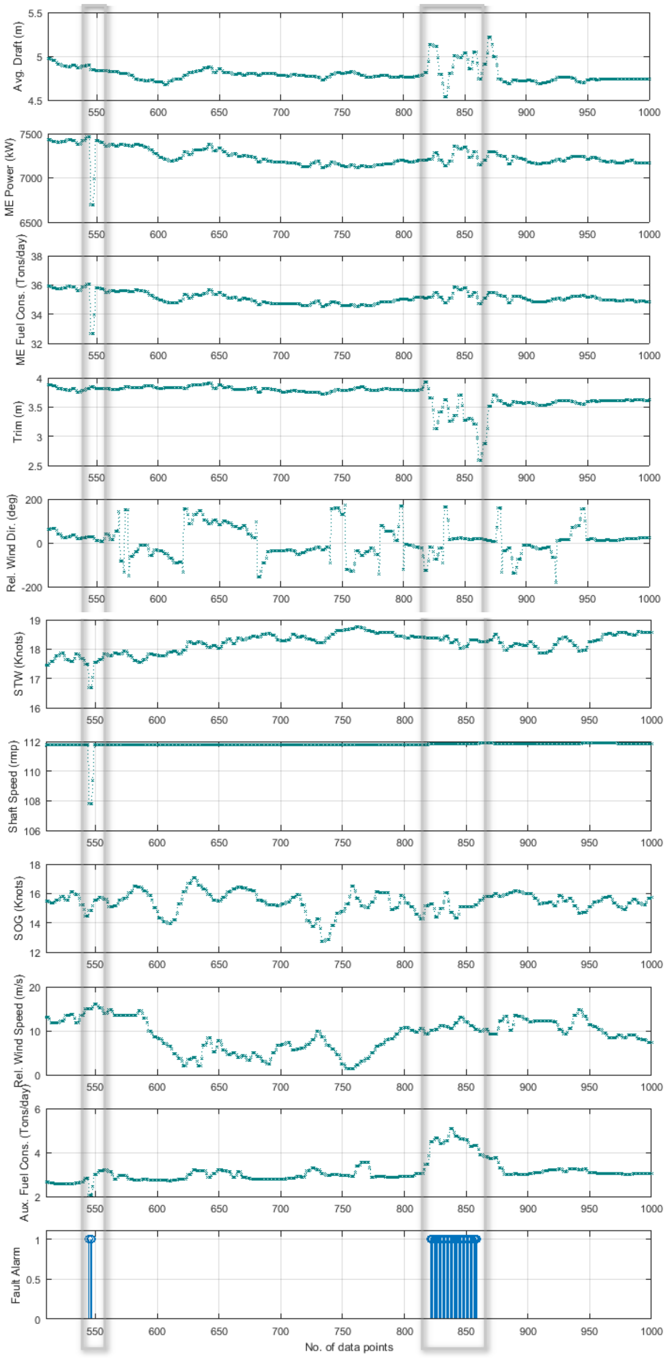

Fig. 6. Ship performance and navigation parameters with sensor faults.

failure situations may relate to DAQ system faults rather than sensor faults. In a future version of this study, the classification of such sensor and DAQ system faults will be investigated. In the second sensor fault situation (i.e. a data interval), several parameters (i.e. average draft, trim) represent some unusual behavior and the auxiliary engine fuel consumption represents considerably higher values, therefore that situation is detected as a sensor fault situation. One should note that the trim and draft values are derived by the same draft measurement sensors, therefore a considerable relationship between these two parameters are expected. Therefore, this specific situation also relates to multiple sensor fault situations. However, this method should be further developed to identify the respect multiple sensor faults situations, where the respective faulty sensors should be recognized by considering other PCs. One should also note that the relative (rel.) wind direction (dir.) of the vessel consists of large frequent variations. However, those variations are not detected as sensor fault situations because that represents common behavior for the respective parameter. Hence, this results show that the PCs based filter is smart enough to distinguish the respective parameter behavior from the sensor faults and that is also a considerable contribution in this method.

5. CONCLUSION

Sensor fault detection under localized models in the combinator diagram of a selected vessel is presented in this study. The proposed approach consists of two detection levels to identify various sensor fault situations in an onboard data acquisition system. Each parameter of ship performance and navigation data has a realistic data range (i.e. variance), where the parameter can vary. This range also relates to the respective parameter variance and that is used as the threshold value for sensor fault detection. If the sensor reads a value beyond this parameter range, then that data point is categorized as a sensor fault situation by the first detection level. However, more complicated sensor fault situations are detected by the second detection level by considering the localized models. The localized models (i.e. data clustering) are developed under the engine propeller combinator diagram by considering GMMs with an EM algorithm. The internal structure of the clustered data points is examined through PCA.

The engine and propeller can have unique operating regions under the combinator diagram. This method can be seen as a piecewise linearization approach, where linearized models along the engine propeller combinator diagram are derived. Finally, the data set is projected into the bottom PC to identify the sensor fault situations in the data set of ship performance and navigation information. A data set of ship performance and navigation information in a selected vessel is analyzed through proposed sensor fault detection levels and the successful results on identifying several fault situations (i.e. fault level 2) are also reported in this study. Hence, the proposed approach has shown successful results with respect to a large data set of ship performance and navigation information collected by an onboard DAQ system (Perera et. al., 2015a, b).

The proposed localized models satisfy the required sensor fault levels as presented in this study. However, a considerable model development steps should be taken in the future work to improve the proposed approach under the engine propeller combinator diagram. One should note that these localized models in the engine propeller combinator diagram are developed under a two dimensional space. Higher dimensional localized models should be derived to identify complex sensor fault situations as required. Even though various sensor faults situations are detected by the proposed detection levels, this methodology should further be developed to identify the receptive faulty sensors. It is believed that higher dimensional localized models can be used to identify not only sensor faults but also related sensors. Hence, that approach can also be used to overcome multiple sensor failure situations, where several faulty sensors can be identified. Such situations can also relate to DAQ system failures, where higher dimensional localized models should be able to capture such variations. Even though these situations are detected as sensor fault situations, such situations can also relate to system or component failures. Appropriate mathematical models that relate to various system and component failures should be derived to overcome such situations, where sensor faults from system and component failures can be separated.

6. ACKNOWLEDGEMENT

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.

REFERENCES

Lajic, Z., Blanke, M. and Nielsen, U. D., (2009), "Fault Detection for Shipboard Monitoring – Volterra Kernel and Hammerstein Model Approaches," 7th IFAC Symposium on Fault Detection, Supervision and Safety of Technical Processes, Spain, pp. 24-29.

Lajic, Z., and Nielsen, U. D., (2009) "Fault detection for shipboard monitoring and decisions support systems," in Proceedings of the ASME 28th International Conference on Ocean, Offshore and Arctic Engineering, Honolulu, Hawaii, June 2009, pp. (OMAE2009–79 367).

Perera, L.P., Mo, B. and Kristjansson, L. A. (2015a) "Identification of Optimal Trim Configurations to improve Energy Efficiency in Ships," In Proceedings of the 10th IFAC Conference on Manoeuvring and Control of Marine Craft (MCMC 2015), Copenhagen, Denmark, pp. 267-272.

Perera, L.P., Mo, B., Kristjansson, L.A., Jonvik, P.C., and Svardal, J.O., (2015b) "Evaluations on Ship Performance under Varying Operational Conditions," In Proceedings of the 34th International Conference on Ocean, Offshore and Arctic Engineering (OMAE 2015), Newfoundland, Canada, (OMAE2015-41793).

Perera, L.P. and Mo, B. (2016a) "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., (2016b) " 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., (2016c), "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).

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

Sperduti, A., (2013) "Linear autoencoder networks for structured data," in the 9th International Workshop on Neural-Symbolic Learning and Reasoning.