Statistical Filter based Sensor and DAQ Fault Detection for Onboard Ship Performance and Navigation Monitoring Systems

Lokukaluge P. Perera

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

Abstract: Statistical filter based sensor and data acquisition (DAQ) fault detection is presented in this study. The parameters of a large-scale data set of ship performance and navigation information are considered as statistical distributions and principal component analysis (PCA) is used to identify the hidden structure of the same data set. This data set relates to a specific operating region of the main engine, where ship performance and navigation conditions can be linearized. The structure derived under PCA is further investigated to identify the respective sensor and DAQ fault situations as the main contribution. That is done by projecting the same data set into the respective principal components, where a new set of ship performance and navigation parameters is derived. Then, the respective parameter variance values of the new data set are calculated and the thresholds that relate to the same variance values for detecting sensor and DAQ fault situations are derived. Finally, the data set of ship performance and navigation information is analyzed through these fault thresholds and the successful results on identifying complex fault situations are presented in this study. Hence, this approach can be used to develop advanced sensor and DAQ fault detection and isolation methodologies of ship performance and navigation monitoring systems.

Keywords: Sensor Fault Detection, Principal Component Analysis, Big Data, Ship Performance and Navigation Monitoring.

1. INTRODUCTION

Ship performance and navigation monitoring is facilitated with sensors and data acquisition (DAQ) systems to collect large scale data sets. However, these sensors and DAQ systems often associate with various sensor and DAQ faults and introduce erroneous regions into the respective data sets. These erroneous data regions may degrade the outcome of ship performance and navigation monitoring systems. Hence, this study proposes to identify various sensor and DAQ fault situations in such real-time monitoring platforms and isolate from the respective data sets. That improves the quality of the data sets and the final results of ship performance and navigation monitoring systems.

Several studies on sensor fault detection and isolation approaches in onboard ship performance and navigation monitoring systems are presented in the recent literature (Lajic and Nielsen (2009) and Lajic et al. (2009)). However, such approaches depend on several empirical mathematical models of ship performance and navigation monitoring to identify sensor and DAQ faults. It is also believed that these models may have difficulties in handling large scale data sets and the respective system-model uncertainties can further degrade the fault identification process. Therefore, a model learning approach, where the respective model is derived within the data set, is proposed in this study to overcome such situations. That can facilitate a better solution to the challenges that are encountered under empirical model based sensor fault detection methods.

The parameters of a large-scale data set with ship performance and navigation information are considered as statistical distributions in this study. Then, principal component analysis (PCA) is used to identify the hidden structure within the same data set and that are considered as the respective models of sensor and DAQ fault identification. PCA is considered as a non-parametric method that extracts relevant information from chaotic type data sets, where the initial size of the data set can reduce to improve the content visibility. The data variance directions among ship performance and navigation parameters, orthogonal to each other, are presented by the respective principal components (PCs). The descending order of the PCs represents the order of significance (i.e. the order of variance) in the respective data set (i.e. the top to bottom PCs held from the most to least important information). Therefore, the most important information of the data set accommodates in the top PCs by projecting the same data set into them (i.e. a selected set of top PCs).

The bottom PCs are often ignored during this process because that may not consist of any important information of the data set in some situations. However, it is noted that when the data set is projected into the bottom PCs, sensor and DAQ system faults and other erroneous data regions often separate into these PCs (Perera, 2016). Therefore, the projected data set (i.e. into the top principal components) may have less erroneous data conditions. That improves the data quality, where the projected data set into

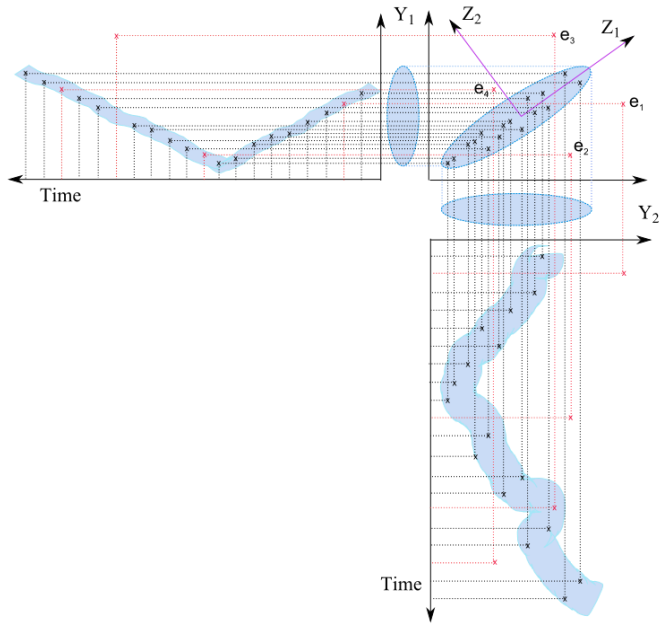

Fig. 1. Two parameter data set under PCA

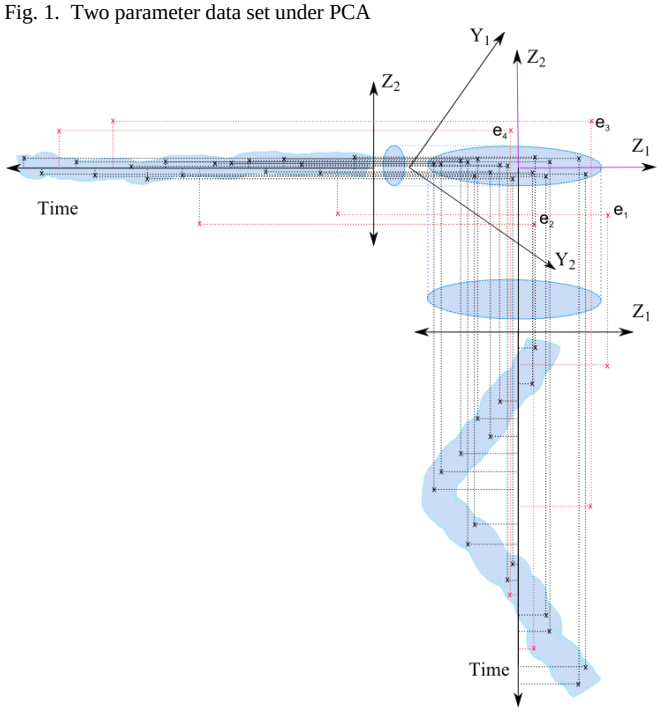

Fig. 2. Projected data set into PC axes

the bottom PCs can be used to identify sensor and DAQ faults. Furthermore, that information can be used to isolate such erroneous data intervals from the respective data set. Hence, the PC structure identified within the data set is used as the basis for the respective models and that are used to identify sensor and DAQ fault situations in this study. These models can handle large scale data sets and implement in real-time monitoring platforms (Perera and Mo, 2016 a & b).

2. SENSOR FAULT DETECTION

An overview of PCA with respect to a large scale data set of ship performance and navigation information is presented in Perera and Mo (2016c and 2016d). A simplified version of PCA is considered in this section and presented in Figure 1 and 2 as a two parameter monitoring situation. It is assumed that two sensors measure the respective parameters in real-time and denote as Y_1 and Y_2. The blue shaded areas in the Y_1 and Y_2 vs. time plots represent the actual values of each parameter with sensor noise. The respective sensors read the parameter values (i.e. denoted as 'x') from the same shaded regions as presented in the figure. The blue ovals (i.e. next to Y_1 and Y_2 axes in Figure 1) represent the selected thresholds that relate to the respective variance/standard deviation of each parameter. One should note that these thresholds are used to identify sensor and DAQ faults. e_1, e_2, e_3 and e_4 denote four sensor and DAQ faults situations that are introduced during this process. The faults, e_1 and e_3, are located far beyond the respective thresholds, therefore those situations can be identified as sensor and DAQ fault situations. That (i.e. the threshold) should be a realistic data range for each parameter to identify such fault situations, where the measurements beyond this range for the respective parameter is categorized as a fault situation. Sensor noise can also be incorporated in such situations, therefore appropriate threshold values with respect to variance/standard deviation values of the parameters should be selected.

The faults, e_2 and e_4, cannot be identified by this method because those faults are still within the thresholds. Therefore, an additional sensor and DAQ fault detection method based in statistical filters is considered to overcome such situations. That method is described in the following section, where such sensor and DAQ faults are identified. It is assumed that these two parameters, Y_1 and Y_2, have a positive correlation (see Figure 1) and that statistical relationship is considered as an appropriate mathematical model to capture other sensor and DAQ fault situations. The respective correlation (the blue inclined oval) between these two parameters is presented in Figure 1 in the Y_1 and Y_2 plot. One should note also that this data region represents the same data set without its time stamp. However, the whole data set can have a time stamp that may relate to the time and duration that it was collected. The respective PCs of the same data set are denoted by Z_1 and Z_2 in the same figure. Z_1 and Z_2 represent the top and bottom (i.e. least) PCs of the same data set, respectively. Then, the data set is transformed into these two PCs and the results are presented in Figure 2.

One should note that the same data set is represented by two new parameters of Z_1 and Z_2 in this figure. Those are the new basis of the original data set and a linear relationship between the parameters in both situations under PCA is assumed. Therefore, both data sets have the same dimensions. The blue ovals (i.e. next to Z_1 and Z_2 axes in Figure 2) represent the selected thresholds that relate to the respective variance/standard deviations of the new parameters. The bottom principal component in this situation is Z_2, therefore the sensor and DAQ fault conditions, e_1, e_2, e_3 and e_4, are observed beyond its variance/standard deviation related threshold value. However, one sensor and DAQ fault, e_1, is also noted beyond the variance/standard deviation related threshold value of the top PC (i.e. Z_1). The

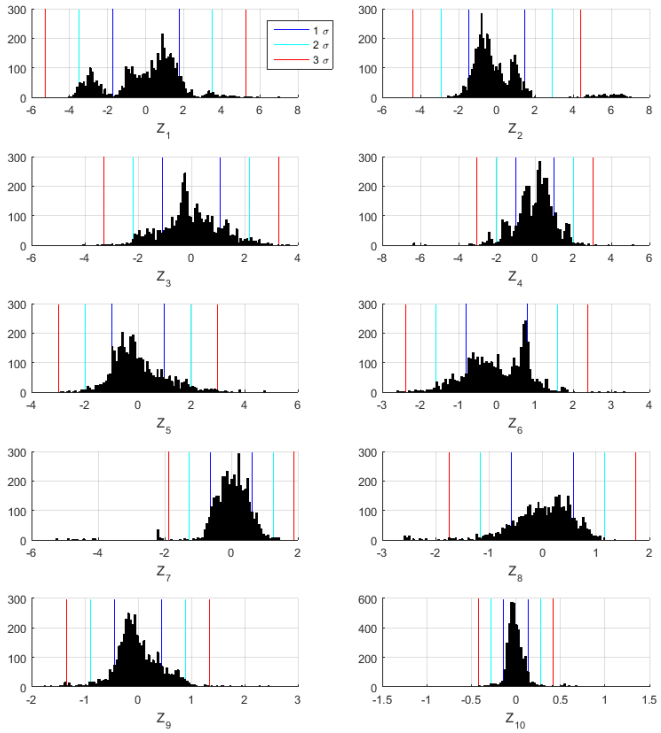

Fig. 3. The data set projected into the PCs.

respective fault numbers, sensor numbers and parameter variance/standard deviation related threshold values, where each sensor fault is detected, are summarized in Table 1. The results show that all sensor and DAQ fault situations are separated into the bottom PC. This is approach is illustrated to identify complex sensor and DAQ fault situations and these erroneous data points should be filtered to improve the quality of the respective data sets (Perera and Mo, 2016b and 2016c). However, the PCA capabilities not only detecting sensor and DAQ faults but also identifying the respective sensors are further investigated in this study. However, the results presented in Table I may not provide a clear relationship to identify such sensor specific faults.

Fault No.	Sensor No.	Y_1	Y_2	Z_1	Z_2
e_1	2		x	x	x
e_2	2				x
e_3	1	x			x
e_4	2				x

Table 1: Sensor faults and observations

3. DATA ANALYSIS

A data set of ship performance and navigation information is collected from a bulk carrier 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 main engine (ME) with maximum continuous rating (MCR) of 7564 (kW) at the shaft rotational speed of 105 (rpm). The vessel has a fixed pitch propeller diameter 6.20 (m) with 4 blades. A clustered data set of ship performance and navigation is derived from the previous data set by considering a specific operating region of the main engine of the vessel (Perera et. al, 2016c). This data set relates to the specific main engine operating region: shaft speed

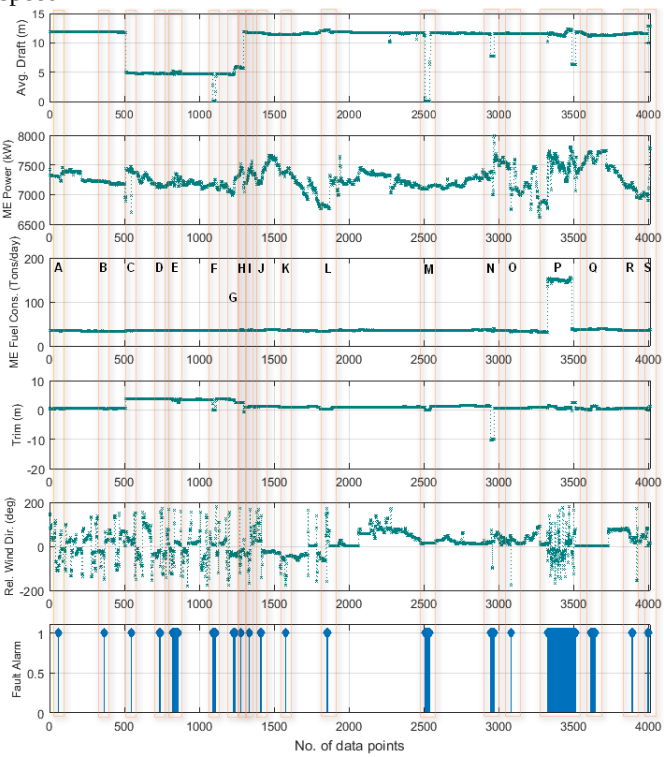

Fig. 4. The respective parameters as point distributions with the fault alarm

from 105 (rpm) to 115 (rpm) and engine power from 6500 (kW) to 8000 (kW).

This data set is equally centered and scaled (i.e. standardized), therefore each parameter is subtracted and divided by its sample mean and standard deviation values. the data set consists of the respective parameters: average (avg.) draft (m), main engine (ME) power (kW), main engine (ME) fuel consumption (cons.) (Tons/day), trim (m), relative (rel.) wind direction (deg), speed through water (STW) (Knots), shaft speed (rpm), speed over ground (SOG) (Knots), relative (rel.) wind speed (m/s) and auxiliary (aux.) fuel consumption (cons.) (Tons/day). That is studied under PCA and projected into the respective PCs. The results (i.e. the respective histogram) as statistical distributions are presented in Figure 3. One should note that each plot represents a new parameter that is derived from the old data set by considering the respective PC structure. The respective PCs from the top to bottom (i.e. least) are presented from $Z_1, Z_2, ..., Z_{10}$. The respective standard deviation values of σ, 2σ and 3σ for each new parameter are also presented in the same figure. Therefore, $Z_1, Z_2, ..., Z_{10}$ introduce the respective statistical filters for sensor and DAQ fault situations. In general, each parameter consists of a Gaussian type distribution, except in Z_1. The new parameter, Z_1, is a combination of two Gaussian type distributions, approximately. One should note that PCA performs its best with Gaussian type distributions. However, this issue can be addressed at the initial step, where a proper data clustering methodology to avoid multi-Gaussian type distributions should be implemented. Therefore, ship performance and navigation data should be clustered appropriately to overcome such situations in such situations.

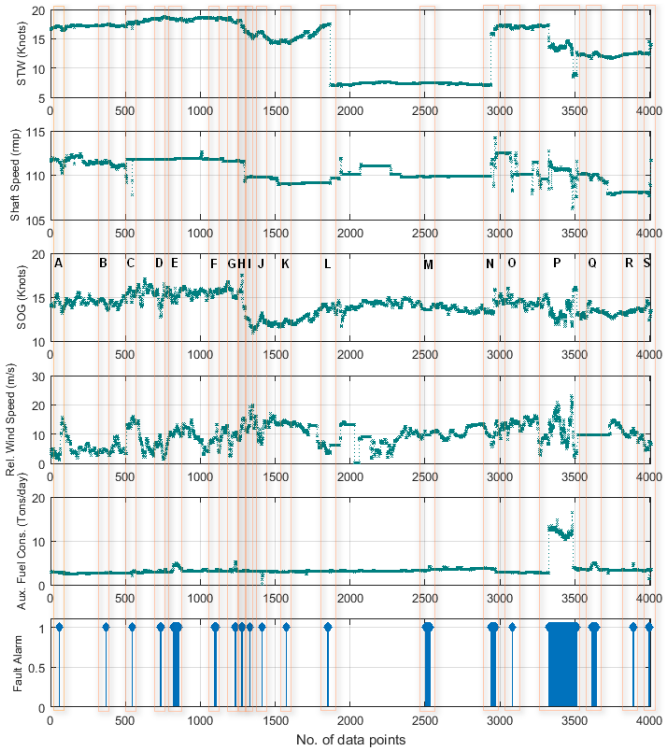

Fig. 5. The respective parameters as point distributions with the fault alarm

The value of 3σ is considered as the threshold limit for sensors and DAQ faults for each new parameter. Therefore, the data points projected beyond this range are considered as the fault situations. Even though sensor and DAQ fault situations can be identified by this method, the respective sensors that are responsible for each situation are to be investigated. The results are presented in Figure 4 and 5 with the respective parameters. The respective sensors that measure these parameters are described in Perera *et al.*, (2015). The bottom plots of both figures represent the combined sensor and DAQ fault situations that are identified as fault alarms. These data points (i.e. the faults) are projected beyond the 3σ values of each new parameter and all faults are combined in these plots. These same sensor and DAQ faults with respect to each statistical filter are presented in Figure 6.

The respective sensor and DAQ faults are denoted from A to S windows to investigate these situations further. A summary of the observations that relates to such fault situations is presented in this section. In general, a considerable parameter variation under fault window A (with 3 fault alarms) is not observed and that situation is noted by filter Z_5 (see Figure 6). Similarly, a considerable parameter variation under fault window B (with 1 fault alarm) is not observed in this situation and that is noted by filter Z_5. The respective parameter variations (i.e. a sudden drop in ME power, fuel consumption, STW, shaft speed and aux. fuel consumption) under fault window C (with 3 fault alarms) are observed by both filters Z_4 and Z_{10}. A considerable parameter variation under fault window D (with 7 fault alarms) is not observed in this situation that is noted by filter Z_9. The respective parameter variations (i.e. a considerable increase in aux. fuel consumption) under fault window E (with 39 fault alarms) are observed by filter Z_{10}. One should note that these fault alarms relate to erroneous data points.

The respective parameter variations (i.e. a sudden drop in avg. draft and trim) under fault window F (with 15 fault alarms) are observed by filter Z_7. A sector of the same fault window is also identified by filter Z_8 (with 4 fault alarms). The respective parameter variations (i.e. a considerable increase in aux. fuel consumption) under fault window G (with 11 fault alarms) are observed by filter Z_{10}. The respective parameter variations (i.e. a considerable increase in SOG and reduction in rel. wind direction) under fault window H (with 4 fault alarms) are observed by filter Z_9. The respective parameter variations (i.e. a considerable increase in rel. wind speed and reduction in rel. wind direction) under fault window I (with 3 fault alarms) are observed by filter Z_9.

The respective parameter variations (i.e. a considerable increase in rel. wind direction and aux, fuel consumption) under fault window J (with 4 fault alarms) are observed by filter Z_{10}. The respective parameter variations (i.e. a considerable reduction in rel. wind direction with high wind speed) under fault window K (with 3 fault alarms) are observed by filter Z_3. The respective parameter variations (i.e. an unusual increase and reduction situation in rel. wind direction) under fault window L (with 7 fault alarms) are observed in this situation. The first part of this fault window is identified by filters Z_5 (with 3 fault alarms) and the second part of this fault window is identified by filters Z_6 (with 4 fault alarms). The respective parameter variations (i.e. a considerable reduction in avg. draft, trim and constant repeated values in STW, SOG, rel. wind speed and direction) under fault window M (with 32 fault alarms) are observed in this situation. This fault window is identified by filters Z_7 and Z_8, and the same is partially identified by filter Z_9.

Fault window N is divided into several sections such as from N(a) to N(f), because several sensor and DAQ faults are overlapped in this situation and identified by various filters, separately. The respective parameter variations (i.e. a considerable variation in rel. wind direction) under the first part (with 11 fault alarms) and (i.e. a considerable increment in ME power and fuel consumption, shaft speed and SOG) the second part (with 8 fault alarms) of fault window N (a) are observed in this situation by filter Z_4. The respective parameter variations (i.e. a considerable reduction in rel. wind direction with respect to high STW and shaft speed) under fault window N (b) (with 20 fault alarms) are observed by filter Z_5.

The respective parameter variations (i.e. a considerable reduction in rel. wind direction with respect to high STW and shaft speed) under the first part (with 1 fault alarm) and (i.e. a considerable reduction in ME power, fuel

consumption, and shaft speed) the second part (with 1 fault alarm) of fault window N (c) are observed in this situation by filter z_6. The respective parameter variations (i.e. a considerable reduction in avg. draft and trim) under fault window N (d) (with 28 fault alarms) are observed by filter z_7. The respective parameter variations (i.e. a considerable reduction in rel. wind speed and increment in shaft speed) under fault window N (e) (with 20 fault alarms) is observed by filter z_8. The respective parameter variations (i.e. a considerable reduction in SOG under high shaft speed, STW and low rel. wind speed) under fault window N (f) (with 8 fault alarms) are observed by filter z_9. The respective parameter variations (i.e. a considerable reduction in rel. wind direction, STW, shaft speed and aux. fuel consumption,) under fault window O (with 1 fault alarms) are observed in this situation. This fault window is identified by both filters z_3 and z_9.

Fault window P is also divided into several sections from P(a) to P(i), because several sensor and DAQ faults are overlapped in this situation and identified by various filters, separately. A considerable parameter variation under fault window P (a) (with 7 fault alarms) is not observed in this situation and that is noted by filter z_1. The respective parameter variations (i.e. a considerable increment in ME and aux. fuel consumption) under fault window P (b) (with 150 fault alarms) are observed by filter z_2. The respective parameter variations (i.e. some variations in rel. wind direction) under fault window P (c) (with 3 fault alarms) are observed by filter z_3. The respective parameter variations (i.e. a considerable reduction in ME power and shaft speed) under fault window P (d) (with 2 fault alarms) are observed by filter z_4.

The respective parameter variations (i.e. some variations in rel. wind direction) under fault window P (e) (with 1 fault alarms) are observed by filter z_5. The parameter variations (i.e. some variations in rel. wind direction) under fault window P (f) (with 11 fault alarms) are observed by filter z_6. The parameter variations (i.e. a considerable reduction in shaft speed) under fault window P (g) (with 4 fault alarms) are observed by filter z_8. The parameter variations (i.e. a considerable reduction in STW under high shaft speed and SOG) under fault window P (h) (with 24 fault alarms) are observed by filter z_9. The parameter variations (i.e. some variations in ME and aux fuel consumption) under fault window P (i) (with 24 fault alarms) are observed by filter z_{10}.

The respective parameter variations (i.e. a considerable increase in aux fuel consumption) under fault window Q (with 31 fault alarms) is observed by filter z_{10}. The respective parameter variations (i.e. a considerable increase in aux fuel consumption) under fault window R (with 4 fault alarms) are observed by filter z_{10}. The respective parameter variations (i.e. a considerable increase in avg. draft, and STW, and reduction in trim, SOG, and aux. fuel consumption) under fault window R (with 3 fault alarms) are observed by filter z_{10}. A summary of filter numbers vs the fault windows are presented in Table 2. Furthermore, each filter with its identified sensor and DAQ fault situations are presented in Figure 6. Therefore, the above conclusions can be verified with the table and figure.

	1	2	3	4	5	6	7	8	9	10
A					X					
B					X					
C				X						X
D									X	
E										X
F							X	X		
G										X
H									X	
I			X							
J										X
K			X							
L					X	X				
M							X	X	X	
N				X	X	X	X	X	X	
a				X						
b					X					
c						X				
d							X			
e								X		
f									X	
O			X						X	
P	X	X	X	X	X	X		X	X	X
a	X									
b		X								
c			X							
d				X						
e					X					
f						X				
g								X		
h									X	
i										X
Q										X
R										X
S										X

Table 2: The filter numbers vs. fault windows.

5. CONCLUSION

A statistical filter based sensor and DAQ fault detection methodology is presented for ship performance and navigation monitoring systems. Even though various sensor and DAQ faults situations are extensively discussed in this study, this methodology should further be developed to identify the receptive faulty sensors. However, the proposed approach has shown promising results with respect to the large data sets of ship performance and navigation information and that are analyzed to evaluate sensor and DAQ fault situations, further. Some inconclusive fault situations are also observed in this study due to complex parameter interactions. Therefore, the proposed threshold values for the respective statistical filters (i.e. 3σ) should be adjusted appropriately overcome such situations. The main advantage in this approach is that the sensor and DAQ fault detection models are derived within the respective data set, where the operating regions of the main engine are used to cluster the data set and the structure of each data cluster is further investigated under PCA. Then, the respective models (i.e. for fault identification) are derived by such information. One should note that vessel performance and navigation conditions are linearized around various vessel/engine operating regions in such models. Therefore, the behavior of

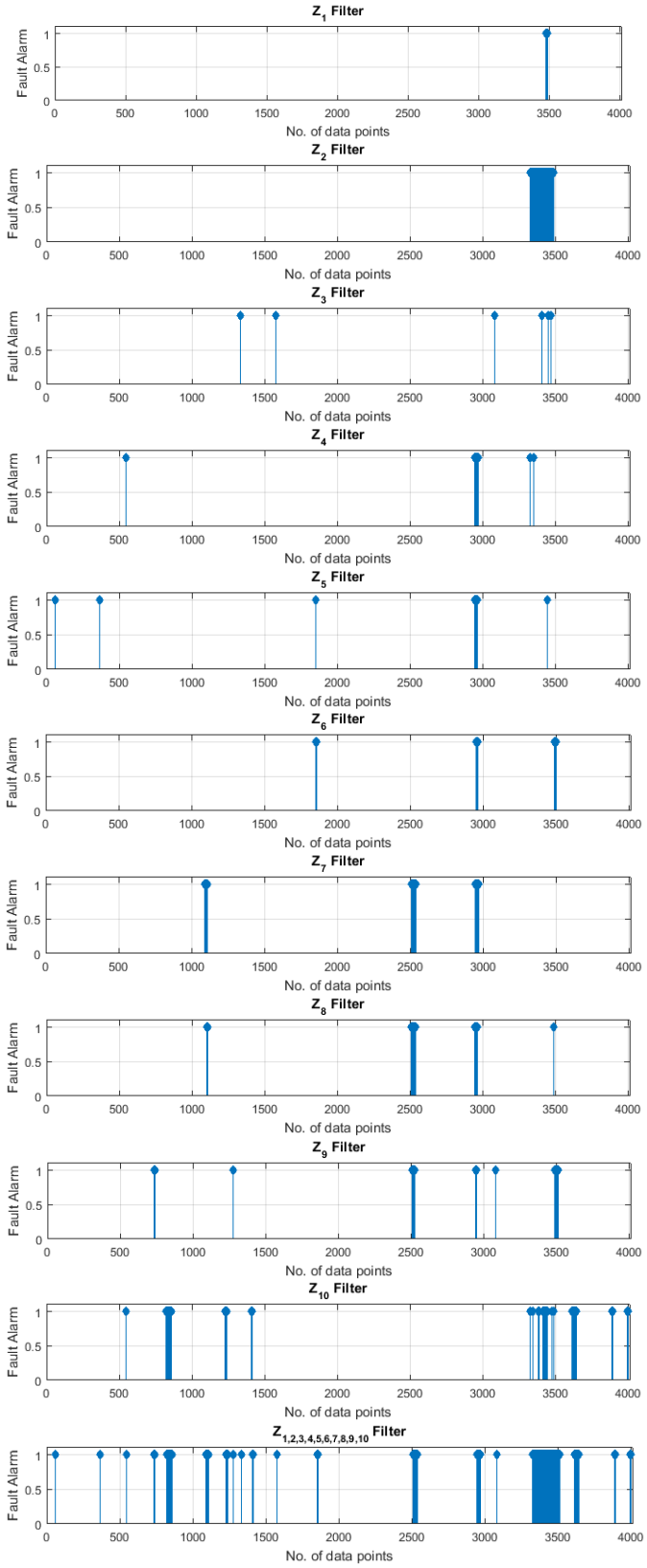

Fig. 6. The data projected into the PCs

each parameter in ship performance and navigation data can vary due to these vessel operating regions. Similarly, that can facilitate a piecewise linear function to capture nonlinear behavior of ship speed power performances. Time-dependence relationships among the parameters in each data cluster are ignored during this approach, where the respective data clusters are considered as statistical distributions. However, these data clusters can still denoted with common time stamps (i.e. daily, weekly or monthly time stamps) as required. The data set is tested as a time series to vary its capabilities in detecting sensor and DAQ fault situations. It is also noted that these sensor and DAQ faults may overlap and that may complicate the fault identification process. Therefore, this study should be continued to identify the solutions to such complex sensor and DAQ fault situations.

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., Kristjansson, L.A., Jonvik, P.C. and Svardal, J.O., (2015), "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., (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) " 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) "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.