Journal of Maritime Research, Vol. VII. No. 3, pp. 3-14, 2010
Copyright © 2010. SEECMAR

Printed in Santander (Spain). All rights reserved
ISSN: 1697-4840

VESSEL CONDITION MONITORIZATION THROUGH SATELLITE USING PRINCIPAL COMPONENT ANALYSIS

J.L. Larrabe[1,2], M.A. Gómez[1,3], I. Sotés[1,2], F.J. Alvarez[1,4], M.C. Rey[1,4], V.E. Mielgo[1,4], I. Sotes[1,2], X. Basogain[1,2]

Received 27 November 2009; in revised form 4 December 2009; accepted 15 August 2010

ABSTRACT

Abstract - Offshore telecommunications are possible in virtually all parts of the world via satellite. Satellite communications are increasing the bandwidth and lowering the cost, but are still far from levels that are in earth. Principal components analysis (PCA) is a statistical technique that has found application in fields such as biometrics or compression of images, being a common tool for the search of pattern of multidimensional data sets. The hypothesis of this study was the possibility of using the PCA theory to compress, with sufficient accuracy, the large amount of data that are collected on board, and then send them all via satellite, in a cheaper and faster way than traditionally. The materials used were 44 samples of 182 different signals, obtained from 19 different equipment on board the LNG vessel "Castillo de Villalba". With these data the PCA algorithm was applied, using a computer program developed by the authors, generating new data packets to send by satellite. Different strategies were used in order to ensure that the correlation coefficient was equal or greater than 0.95 in the comparison between original data and reconstructed data onshore. Results obtained by grouping the 182 signals showed 46.9% reduction in data sent via satellite, with a mean r = 0.95 ± 0.08. This strategy is appropriate for decision making about tele-diagnosis and maintenance of the ship using ground equipment. Consequently, significant savings are achieved in telecommunication costs and telecommunication times.

Keywords: Vessel, telecommunications, satellite, compression, lossy.

[1] Universidad del País Vasco, Tel. 946014847, Fax. 946017700, María Díaz de Haro 68 48920 Portugalete, Spain. [2]Professor, Email: jl.larrabe@ehu.es, [3]Professor, Email: cnpgosom@lg.ehu.es, [4]Researcher, Email: jl.larrabe@ehu.es.

INTRODUCTION

Communications at sea are in the process of evolution. Satellite coverage is practically worldwide for the main marine technologies used today, such as Iridium, Inmarsat, Very Small Aperture Terminal (VSAT), etc. Moreover, much progress is being made in increasing the bandwidth as well as in reducing the required hardware prices and the communications fees. This paper presents a Lossy Compression Technique or data non-exact compression method by applying principal component analysis (PCA). It is proposed for monitoring the condition of equipment on board of surface vessel or underwater vehicles, which allows a lower cost of communication, reducing the used satellite time or bandwidth (Calvo et al. 2009; Gomariz et al. 2009; Horak 2007; Organización, I 2008; Organización, I & Organización de Aviación Civil Internacional (Montreal 2008).

The first part of the introduction shows the main communication systems used in the shipbuilding industry, such as Iridium, Inmarsat and VSAT and bandwidths and prices available for communication are analyzed. Subsequently, we examine the procedures for lossless and lossy data compression, indicating the characteristics of each. Finally, an introduction to the procedure of principal components analysis is done.

The International Convention for the Safety of Life at Sea (SOLAS), adopted under the support of the International Maritime Organization (IMO), an agency of the UN, the implementation on board of the Global Maritime Distress and Safety (GMDSS). It is a set of safety procedures, equipment and communication protocols designed to enhance safety and simplicity of navigation and rescue boats at risk. It is in force at merchant and passenger ships since 1999. The system seeks to carry out the following: to alert, including position determination of the unit in distress, search and rescue coordination, locating the provision of maritime information, general communications and communications bridge to bridge. The network of satellites operated by Inmarsat, under the supervision of the IMO, is a key element of the GMDSS system in order to provide telephony and data services to users worldwide (coverage spans the globe except the polar regions above 70 degrees of latitude), via special terminals which communicate to ground stations through twelve geosynchronous telecommunications satellites. Inmarsat launches into space every few years a new generation of satellites, with increasingly more powerful features. Hardware prices vary between 3,000 and 20,000 Euros, having a variety of types of terminals (Fleet F77, Fleet F55, Inmarsat M, Inmarsat C, Inmarsat B, etc). The newest product line is the Inmarsat BGAN (Broadband Global Area Network) terminals to enable data transmission of up to 432 kbps in terminals ranging from the size of a small notebook and weighing 0.9 kg. The estimated costs of a voice call via the Inmarsat systems are about 1.2 € per minute and the broadband data transmission have an average of 5.4 € per Megabyte. These costs are approximate and vary depending on telecommunication provider (Horak 2007; Organización, I 1995;

Organización, I 2008; Organización, I & Organización de Aviación Civil Internacional (Montreal 2008).

Thanks to the Inmarsat system, vessels and platforms can be communicated around the world to a similar level than land users, although at a higher cost. This is enabling new schemes in the operation of ships, improving security, enabling onboard personnel reduction and optimizing its operation.

Recently, new alternatives have been developed for offshore data transmission, such as the Iridium system and VSAT. Iridium is the name of a communication network which has 66 satellites in total arranged, in 11 longitudinal orbits of 6 satellites each. This network was designed by Satellital Movil Phone Services from Motorola, with global coverage (including the poles) and, it was put on service on 1998. Initially, it was a very expensive service because of the high price of the hardware. For example, an antenna reached 2,700€, and the cost of communication about 5.42 € per minute. Now, Iridium has the OpenPort service directed to maritime market which will be able to transmit up to 128 Kbps worldwide; the cost of the terminal reaches 4,600 € and the transmission –reception data costs will be 3.90 € per Megabyte (Elbert 1999; Elbert 2008; Horak 2007; Organización, I 2008).

VSAT is a data transmission technology between an earth station and a satellite, with directional dish antenna, that is smaller than 3 meters. Data rates typically range from narrowband 54 kbps to more than 18 Mbps. VSATs are used for maritime communications with a special design to withstand tough marine conditions by adjusting antenna position and maintaining lock on the satellite allowing for the ship's motion and turning. VSAT network consists of a constellation of satellites and a terrestrial network. Depending on the chosen frequency coverage is global (C band 3-7 GHz) or Regional (Ku band 12-14 GHz), with all shore coverage and a few miles from the coast. The cost of the terminal is between 45,000 to 200,000 € and the transmission–reception costs are determined by a flat rate in terms of 3,000-15,000 € per month, in agreement to contract for services with the telecommunication operator (Horak 2007; Organización, I 2008). The figure 1 shows an example of comparison of communication costs per month [€] between Iridium Openport, Imarsat FeetBroadBand 500 and VSAT technologies. For large amount of data traffic, VSAT solution is an economical solution.

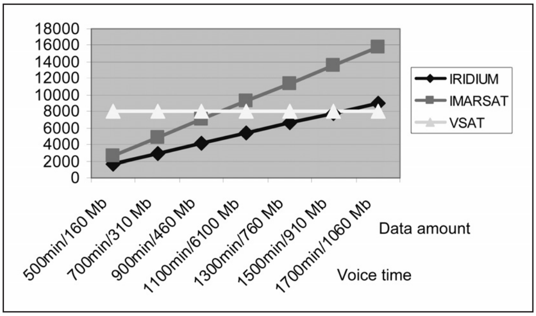

Figure 1. Data transmission from vessel: example of comparison of communication costs per month [€] between Iridium Openport, Imarsat FBB500 and VSAT technologies.

In order to reduce fees of satellite communication at sea, there are two ways to compress information: lossless and lossy way. The lossless compression algorithms usually use statistical redundancy to represent the data without error. The compression methods are possible because most real-world data has statistical redundancy. For example, in English text, the letter 'e' is much more common than the letter 'z', and the probability that the letter 'q' is followed by the letter 'z' is very small.

Another compression technique, called lossy data compression, is possible if some loss of fidelity is acceptable. In general, lossy data compression will be guided by research on how people perceive the data in question. For example, the human eye is more sensitive to subtle variations in brightness than variations in colour. JPEG image compression works, in part, by "rounding off" method to lessen information. Lossy data compression provides a way to obtain better compression rates of a given amount of data. In some cases, data reliability is required and lossy compression will be questioned; in other cases, fidelity can be sacrificed to reduce the amount of data as much as possible (Grailu, Lotfizad, & Sadoghi-Yazdi 2009; Steinberg 2009).

Lossless data compression is reversible, so that the original data can be rebuilt, while lossy data compression accepts some loss of data in order to achieve greater compression. Repeatedly lossy compressing and decompressing the file will cause it to progressively lose quality. Lossy methods are most often used for compressing sound, images or videos. This is because these types of data are intended for human interpretation where the mind can easily "fill in the blanks" or see past very minor errors or inconsistencies. In practice, lossless data compression also gets to a point where compressing again does not work, although lossy extreme algorithms, such as removing the last byte of a file will always compress the file up to the point where it becomes empty.

Principal Component analysis (PCA) is a useful statistical technique that has found application in fields such as face recognition and image compression, and is a common technique for finding patterns in large data (Belogolovy et al. 2009; Johnson, Rose-Pehrsson, & Morris 2004; Prasad 2007; Xiong, Liang, & Qian 2007).

It is a way of identifying patterns in data and expresses it in such a way that highlights their similarities and differences. Because patterns can be difficult to find in data of high dimension, where the luxury of graphical representation is not available, PCA is a powerful tool for data analysis (Lee, Qin, & Lee 2007; Manly 2005; Mason & Young 2002).

The other main advantage of PCA is that once you have found these patterns in the data, you can compress it, reducing the number of dimensions, without much loss of information (Johnson, Rose-Pehrsson, & Morris 2004; Prasad 2007; Xiong, Liang, & Qian 2007).

For this work, the hypothesis was based on the possibility to use the theory of PCA and perform lossy compression of large amounts of collected data by the control system onboard ships and send it via satellite saving time or reducing costs.

OBJECTIVES

1. To develop a computer program for compression and decompression of the algorithms data collected on board using PCA.
2. To find the best strategy to compress data using PCA.

MATERIALS AND METHODS

The data used were collected on board a ship which transported liquefied natural gas LNG (Castillo de Villalba) through its integrated automation system (IAS, Norcontrol, Norway). This device generates a spreadsheet file every 12 hours, which represents the condition of 182 different signals of the 19 major subsystems of the vessel: Main Turbine, Boiler Common, Boiler No. 1 Boiler No 2, Turbo Generator No 1, Turbo Generator No 2, Diesel Generator, Boiler Water Readings, Feed Cond. System, Evaporators, Water Tanks, Fuel Oil, Marine Diesel Oil, Gas Oil, Sludge and Bilge, Others, LD Compressors and Fridges-Air Conditioning.

It was considered a pilot-pilot travel Saint-Nazaire (France) - Bonny (Nigeria) with a duration of 22 days, representing 44 different values for each signal, representing 44 different values for each signal and obtaining data from 8008 double precision floating point (64 bits). The data set keeps the correlations between the signals for the normal operation of the ship: loading, unloading, ballast voyage and travel in cargo. The original data were stored in a matrix called $[Data]_{mxn}$, where the samples or observations were rows (m) and columns (n) were dimensions, previously cited as signals (Torokhti & Friedland 2009).

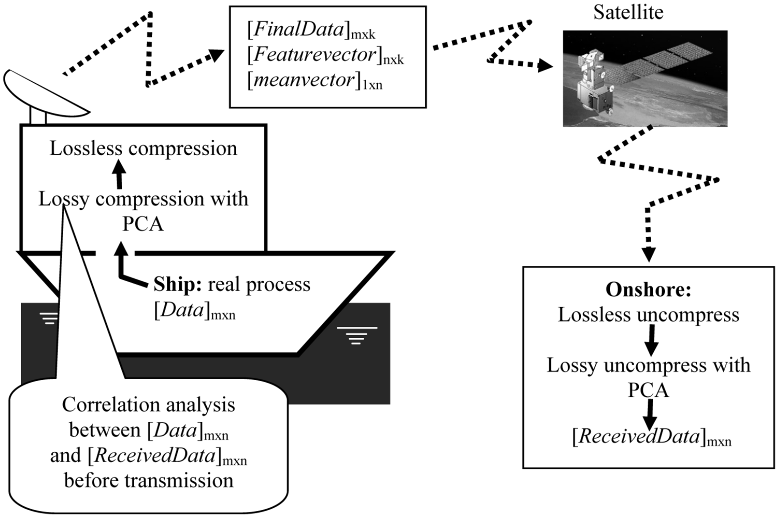

Figure. 2. Block diagram of the transmission method using lossy compression with PCA.

Two computer programs were developed by the authors (Labview 8.2, National Instrument, Austin TX). One of them performs PCA and generated the packets to send by satellite offshore. The second program uncompressed data onshore. All results were displayed graphically and saved in file format compatible with spreadsheet programs. Figure 2 shows the block diagram of the transmission method.

Several steps were given to determine PCA (Lee, Qin, & Lee 2007; Mason & Young 2002; Prasad 2007; Xiong, Liang, & Qian 2007). First, from each dimension, was subtracted the mean. This produced a data set ($[DataAdjust]_{mxn}$) whose mean was zero:

$$DataAdjust_{ij} = Data_{ij} - \overline{\{Data_{xj} \mid x = 1,\ldots,m\}} \tag{1}$$

where i was observation, j was dimension and m was total number of observations.

Second the $[covariance\ matrix]_{nxn}$ of $[DataAdjust]_{mxn}$ was evaluated. The eigenvalues λ and eigenvectors $\boldsymbol{x}$ for $[covariance\ matrix]_{nxn}$ were calculated as third step: eigenvalues λ_j were used for calculation of [% of total variance] (V_j) for each component j:

$$V_j = 100 \frac{\lambda_j}{\sum_{x=1}^{n} \lambda_x} \qquad \sum_{x=1}^{n} \lambda_x = n \tag{2}$$

After components were chosen to form feature vector: Eigenvalues λ and eigenvectors $\mathbf{x}$ are sorted in descending order. Component with highest λ was considered principal component and $[Featurevector]_{nxk}=(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k)$, where $\mathbf{x}_i$ was a column oriented eigenvector, contains chosen components (k).

To derive new dataset was performed the transpose of [*Featurevector*] and [*DataAdjust*] in order to get original data in terms of chosen components:

$$[FinalData^T]_{kxm} = [FeatureVector^T]_{kxn} \ x \ [DataAdjust^T]_{nxm} \tag{3}$$

FinalData has eigenvectors as coordinate axes.

After choosing proper components, was prepared the package of double precision numbers (64 bit) to send once for satellite:

$$\begin{aligned} &[FinalData]_{mxk} \\ &[Featurevector]_{nxk} \\ &[meanvector]_{1xn} \end{aligned} \tag{4}$$

These data were received on shore and the procedure of reconstruction of the data settles down in destination:

$$[ReceivedData]_{mxn} = ([Featurevector]_{nxk} \ x \ [FinalData^T]_{kxm})^T + [meanvector]_{mxn} \quad (5)$$

having $[meanvector]_{mxn}$ matrix all m rows equal to $[meanvector]_{1xn}$. This procedure yields original data using the chosen components (figure 2).

Simplifying for not going into details of programming, the designed computer application chosen in sequence the eigenvectors from highest to lowest eigenvalue and calculated the mean r correlation coefficient of all the variables from the matrix with the real data $[Data]_{mxn}$ and the received matrix $[ReceivedData]_{mxn}$. When r was greater or equal than a given threshold, the data package to send was prepared. In the case of this work, the threshold chosen was $r \geq 0.95$. The figure 3 shows the front panel of that program (Khoo 2000; Peña Sánchez de Rivera 1995).

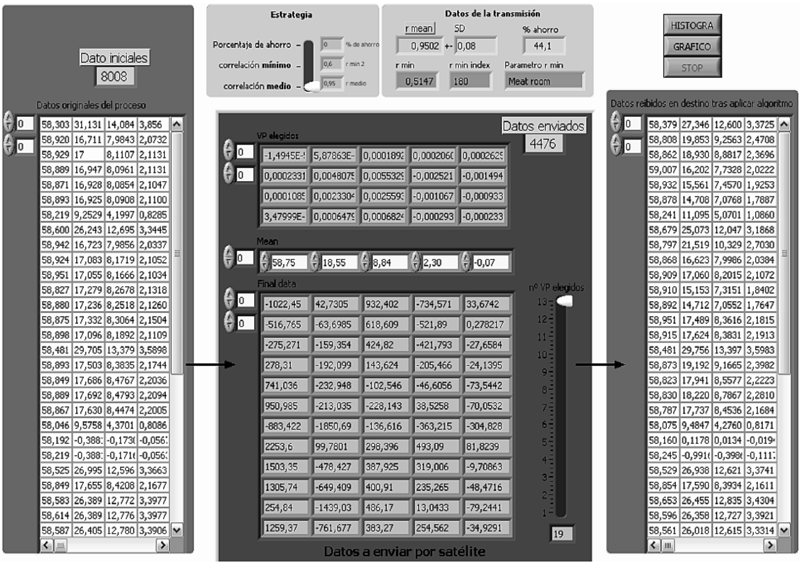

Figure 3. Front panel of the designed computer application. In the left side are the real data collected onboard. In the centre are the three packets sent by satellite. In the right side it is shown received data.

Two methods were analyzed: on the one hand all 182 signals were compressed together (table 1) and only one packet to be sent was created. On the other hand each vessel subsystem was treated independently, i.e. 34 signal of main turbine, 12 signals of Boiler nº 1, etc. (table 2). Now 19 different packets were created to be send. In all cases the following items were evaluated: number of principal components used, coefficient of correlation, number of original data, number of sent data and file space saved taking into account that each value needed 64 bits for its representation. The Histogram of coefficient of correlation r obtained sending all collected onboard variables (182) and separately are shoed in the figure 4 and 5.

RESULTS

	Number of Signals	Number of principal components used	Coefficient correlation	Number of original data	Number of sent data	File space saved
All equipments together	182	18	0.95±0.08	8008	4250	46.9%

Table 1. Obtained transmision results sending all collected onboard variables together.

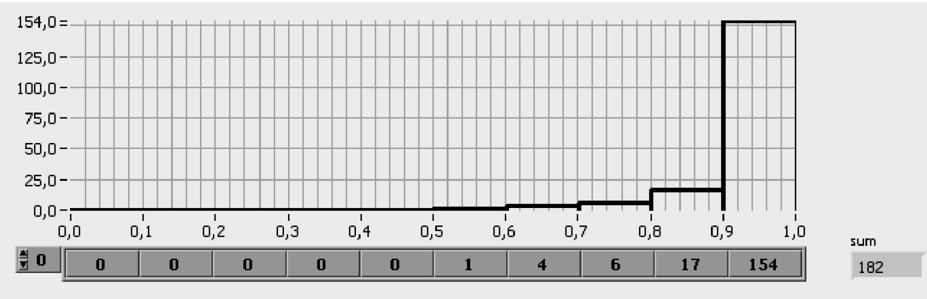

Figure 4. Histogram of coefficient of correlation *r* obtained sending all collected onboard variables (182) together. 154 (84.6 %) of them presented *r* between 0.9 and 1. All signal presented *r* > 0.5.

		Number of Signals	Number of principal components used	Coefficient correlation	Number of original data	Number of sent data	File space saved
1	Main Turbine	34	6	0.95±0.11	1496	502	66.4 %
2	Boiler Common A	4	4	1±0	176	196	-11.4 %
3	Boiler Common B	3	2	1±0	132	97	26.5 %
4	Boiler nº1	12	6	0.96±0.09	528	348	34.1 %
5	Boiler nº2	12	7	0.95±0.11	528	404	23.5 %
6	Turbo Generator No 1	12	2	0.97±0.04	528	124	76.5 %
7	Turbo Generator No 2	12	3	0.99±0.02	528	180	65.9 %
8	Diesel Generator	17	2	0.97±0.08	748	139	81.4 %
9	Boiler Water Readings	4	4	1±0	176	196	-11.4%
10	Feed Cond. System	12	7	0.95±0.12	528	404	23.5 %
11	Evaporators	2	2	1±0	88	94	-6.8 %
12	Water Tanks	7	5	0.98±0.03	308	262	14.9 %
13	Fuel Oil	18	9	0.95±0.08	792	576	27.3 %
14	Marine Diesel Oil	3	3	1±0	132	144	-9.1 %
15	Gas Oil	4	3	0.96±0.08	176	148	15.9 %
16	Bilge and Sludges	2	2	1±0	88	94	-6.8 %
17	Others	3	2	1±0.01	132	97	26.5 %
18	LD Compressors	17	5	0.96±0.09	748	322	57.0 %
19	Fridges - Air Conditioning	4	4	1±0	176	196	-11.6 %
	Total	**182**			**8008**	**4523**	**43.5%**

Table 2. Obtained transmision results sending each vessel subsystem independently.

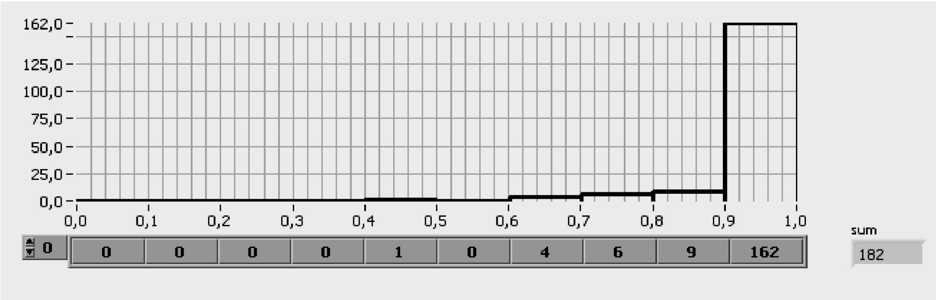

Figure 5. Histogram of coefficient of correlation *r* obtained sending each vessel subsystem independently. 162 (89 %) of them presented *r* between 0.9 and 1. All signal presented *r* > 0.4.

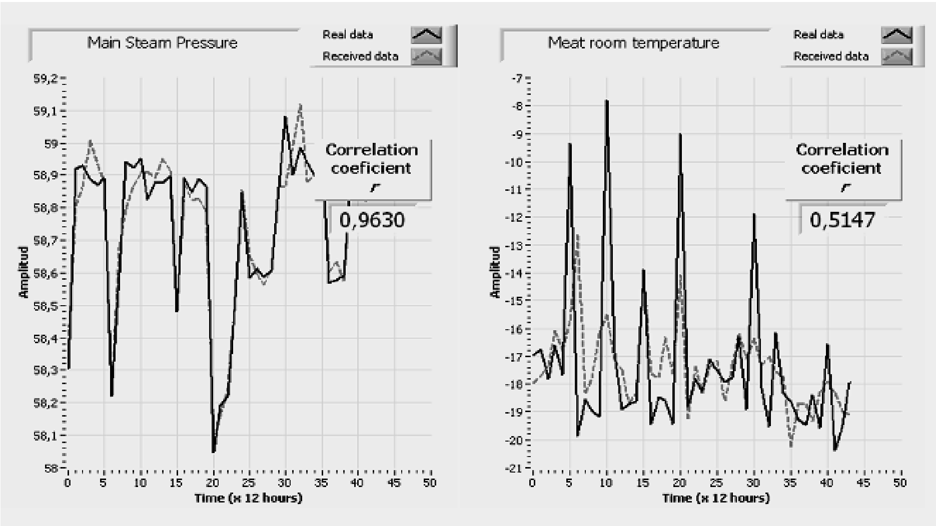

Figure 6. Two examples of onboard real signals (black line) and onshore received signal (red-dashed line) after PCA compression was applied. It is shown the case of sending all the 182 signals. The graph on the right represents the signal with the worst coefficient of correlation and in the left a signal with medium coefficient of correlation.

DISCUSSION

The best savings rates on the volume of transmitted data are obtained in power generation systems. These systems find strong correlations between different variables within it. The number of monitored signals of each subsystem is an important parameter, since those with less than 4 signals sampled data have the worst savings in the amount of data sent.

Between the two transmission strategies studied in this work, to send the full set of signals requires less data transmission rates (table 1: 46.9 % of saved money) than in separate subsystems (table 2: 43.5 % of saved money). However, this last strategy ensures that the dispersion of the correlation coefficients of the signals is less, improving fidelity in reception (figures 4 and 5).

It would be interesting to extend this study to other types of vessel, with other types of subsystems and operation forms.

CONCLUSION

The software developed for transmission using PCA significantly reduces the amount of data sent via satellite, reducing time and cost of communication in case of transmission of all signals together. Alternatively, PCA technique may increase the number of samples sent for a defined time and cost.

For some subsystems of the ship, the separate transmission of their signals is advantageous, bringing savings of 81.4% in the amount of data sent obtaining very high mean correlation coefficient ($r = 0.97 \pm 0.08$). For other subsystems, due to a low correlation between variables, the PCA is not advantageous with regard to sending the raw data. The software should detect these situations and use the most economical way.

PCA compression strategy is appropriate for making onshore maintenance decisions about onboard equipment with a strong correlation between sampled signals as propulsion subsystem, generation plant, etc, reducing communication cost.

ACKNOWLEDGMENTS

Our sincere thanks to the Castillo de Villalba ship crew and especially to D. Pablo Vegas, for his invaluable help.

REFERENCES

Belogolovy, A.; Belyaev, E.; Sergeev, A. & Turlikov, A. 2009: “Video Compression for Wireless Transmission: Reducing the Power Consumption of the WPAN Hi-Speed Systems”, *Smart Spaces and Next Generation Wired/Wireless Networking, Proceedings*, vol. 5764, pp. 313-322.

Calvo, O.; Sousa, A.; Bibiloni, J.; Curti, G.; Acosta, G. & Rozenfeld, A. 2009: “Low-Cost Autonomous Underwater Vehicle for Underwater Acoustic Inspections”, *Journal of Maritime Research*, vol. 6, no. 2, pp. 37-52.

Elbert, B.R. 1999: *The satellite communication applications handbook* Artech house, Norwood, Massachusetts.

Elbert, B. R. 2008: *Introduction to satellite communication*, 3rd ed edn, Artech House, Norwood, Massachusetts.

Gomariz, S.; Prat, J.; Gaya, P. & Sole, J. 2009: “An Autonomous Vehicle Development for Submarine Observation”, *Journal of Maritime Research*, vol. 6, no. 2, pp. 23-36.

Grailu, H.; Lotfizad, M. & Sadoghi-Yazdi, H. 2009: “A lossy/lossless compression method for printed typeset bi-level text images based on improved pattern matching”, *International Journal on Document Analysis and Recognition*, vol. 11, no. 4, pp. 159-182.

Horak, R. 2007: *Telecommunications and data communications handbook* Wiley-Interscience, Hoboken, New Jersey.

Johnson, K. J.; Rose-Pehrsson, S. L. & Morris, R. E. 2004: “Monitoring diesel fuel degradation by gas chromatography-mass Spectroscopy and chemometric analysis”, *Energy & Fuels*, vol. 18, no. 3, pp. 844-850.

Khoo 2000: *Physiological control systems analysis, simulation, and estimation* IEEE Press etc., New York etc.

Lee, J. M.; Qin, S. J. & Lee, I. B. 2007: “Fault detection of non-linear processes using kernel independent component analysis”, *Canadian Journal of Chemical Engineering*, vol. 85, no. 4, pp. 526-536.

Manly, B. F. J. 2005, *Multivariate statistical methods: a primer* Chapman and Hall, London etc.

Mason, R. L. & Young, J. C. 2002, *Multivariate Statistical Process Control with Industrial Applications* Society for Industrial and Applied Mathematics, Philadelphia.

Organización, M., I 1995: *G.M.D.S.S handbook handbook on the Global Maritime Distress and Safety System*, 2nd ed edn, International Maritime Organization, London.

Organización, M., I 2008: *Codigo NGV 2000 Código internacional de seguridad para naves de gran velocidad : resolucion MSC.97(73) adoptada el 5 diciembre de 2000, enmendada mediante las resoluciones MSC.175(79) y MSC.222(82)*, Edición de 2008, 2a ed edn, Organizacion Marítima Internacional, Londres.

Organización, M., I & Organización de Aviación Civil Internacional (Montreal, C. 2008): "Manual IAMSAR manual internacional de los servicios aeronáuticos y marítimos de búsqueda y salvamento", vol. Edición de 2008, 7a ed.

Peña Sánchez de Rivera, D. 1995: *Estadística modelos y métodos*, 2a ed. rev. edn, Alianza, Madrid.

Prasad, G. 2007: "On fusion of PCA and a physical model-based predictive control strategy for efficient load-cycling operation of a thermal power plant", *Optimal Control Applications & Methods*, vol. 28, no. 4, pp. 231-258.

Steinberg, Y. 2009: "Coding and Common Reconstruction", *Ieee Transactions on Information Theory*, vol. 55, no. 11, pp. 4995-5010.

Torokhti, A. & Friedland, S. 2009: "Towards theory of generic Principal Component Analysis", *Journal of Multivariate Analysis*, vol. 100, no. 4, pp. 661-669.

Xiong, L.; Liang, J. & Qian, J. X. 2007: "Multivariate statistical process monitoring of an industrial polypropylene catalyzer reactor with component analysis and kernel density estimatione", *Chinese Journal of Chemical Engineering*, vol. 15, no. 4, pp. 524-532.