# PRINCIPAL COMPONENT ANALYSIS TO COMPRESS ACQUIRED DATA OFFSHORE

**Juan Luis Larrabe[1], Miguel Angel Gómez[1], Francisco José Alvarez[2], Elena Gastiasoro[2], M. Carmen Rey[2], Victoria E. Mielgo[2], Enrique Hilario[2]**

**(1) Dept. Navigation Sciences, Engineers and Shipbuilders, H.T.S. Nautical and Naval Machines, Maria Diaz de Haro, 68.Portugalete 48920 Bizkaia, SPAIN**

**(2) Research Unit; Hospital of Cruces, Barakaldo, Bizkaia SPAIN**

**Tfn: 034 94 6014847 email: jl.larrabe@ehu.es**

*Abstract - Telecommunications offshore have connectivity in virtually all parts of the globe via satellite, with increasing bandwidth and lower cost, but still far from levels that are onshore. The principal component analysis (PCA) is a statistical technique that has found application in fields such as biometrics or compression of images, being a common tool for finding patterns in multidimensional data sets. The hypothesis for this work was that it was possible to use the theory of PCA to compress, with sufficient accuracy, the large amount of data that are collected on board to a vessel and then sent by satellite in a more economical or rapid way than the traditional one. The material used were 44 samples of 182 different signals, collected from 19 different equipment on board to "Castillo de Villalba" Liquid Natural Gas carrier vessel. 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 coefficient of correlation r between original and reconstructed data onshore were equal or greater than 0.95. The results showed that it was possible to save 46.9% in the number of data sent via satellite, in the case of grouping all the 182 signs, with a mean r = 0.95 ± 0.08. This strategy is appropriate for onshore vessel equipment telediagnostic and maintenance decision making, with telecommunication cost or time savings.*

*Keywords - 1. Ship, 2. Telecommunication, 3. Satellite, 4. Compression, 5. Non-exact.*

## I. INTRODUCCIÓN

Communications at sea are in the process of evolution. The coverage by the satellite is virtually the entire globe for major technologies used today, either Iridium, Inmarsat or very small aperture terminals (VSAT). Moreover, where more progress is being made is in the bandwidth, increasing it, as well as the dwindling cost of communications and hardware equipment necessary. This paper presents a data non-exact compression method by applying principal components analysis (PCA) for monitoring the condition of equipment on board which allows lower cost of communication or reduce the occupation time of the bandwidth for a given amount of data sampled.

In order to reduce offshore satellite communication fees, there are two ways to compact information: exact and non-exact computer data compression. Exact compression algorithms usually exploit statistical redundancy to represent the sender's data more concisely without error.

Another compression technique, called non-exact data compression or perceptual coding, is possible if some loss of fidelity is acceptable. Generally, a non-exact data compression will be guided by research on how people perceive the data in question. Non-exact data compression provides a way to obtain the best fidelity for a given amount of compression. In some cases, transparent (unnoticeable) compression is desired; in other cases, fidelity is sacrificed to reduce the amount of data as much as possible.

Principal Component analysis (PCA) is a statistical technique with application in fields such as face recognition and image compression, and is a common technique for finding patterns in data of high dimension.

The other main advantage of PCA is that once you have found these patterns in the data, is possible compress it, by reducing the number of dimensions, without much loss of information.

For this work, the departure hypothesis was based in the possibility to use PCA theory to manage great quantity of data collected onboard by vessel control system, to compress and send it through satellite.

## II. OBJECTIVES

1.To develop a computer program to perform PCA compress and uncompress algorithm with collected date onboard.

2.To find the best strategy to compress data using PCA.

## III. MATERIALS AND METHODS

The data used were collected on board a ship for transporting 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.

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 1 shows the block diagram of the transmission method.

The designed computer application chosen in sequence the eigenvectors from highest to lowest eigenvalue and calculated the mean correlation coefficient r 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 ≥ 0.95. The figure 2 shows the front panel of the that program.

## IV. RESULTS

TABLE 1
OBTAINED TRANSMISION RESULTS SENDING ALL COLLECTED ONBOARD VARIABLES TOGETHER

| | Number of Signalsv | 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% |

## V. 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, it is advantageous the transmission of their signals separately, bringing savings of 81.4% in the amount of data to send with very high mean correlation coefficient (r = 0.97 ± 0.08). For other subsystems, due 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 less cost 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.

## VI. ACKNOWLEDGMENTS

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

***REFERENCES***

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

*[2] M. ORGANIZACIÓN, I, 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 ed. Londres: Organizacion*

*Marítima Internacional, 2008.*

*[3] M. ORGANIZACIÓN, I AND C. Organización de Aviación Civil Internacional (Montreal, "Manual IAMSAR. Manual internacional de los servicios aeronáuticos y marítimos de búsqueda y salvamento," vol. Edición de 2008, 7a ed 2008.*

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

*[5] M. ORGANIZACIÓN, I, G.M.D.S.S handbook. Handbook on the Global Maritime Distress and Safety System, 2nd ed ed. London: International Maritime Organization, 1995.*

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

*[7] K. J. Johnson, S. L. Rose-Pehrsson and R. E. Morris, "Monitoring diesel fuel degradation by gas chromatography-mass Spectroscopy and chemometric analysis," Energy & Fuels, vol. 18, no. 3, pp. 844-850, May2004.*

*[8] G. Prasad, "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, July2007.*

*[9] B. F. J. Manly, Multivariate statistical methods*
*a primer. London etc.: Chapman and Hall, /19.*

*[10] L. Xiong, J. Liang, and J. X. Qian, "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, Aug.2007.*

*[11] J. M. Lee, S. J. Qin, and I. B. Lee, "Fault detection of non-linear processes using kernel independent component analysis," Canadian Journal of Chemical Engineering, vol. 85, no. 4, pp. 526-536, Aug.2007.*

*[12] R. L. Mason and J. C. Young, Multivariate Statistical Process Control with Industrial Applications. Philadelphia: Society for Industrial and Applied Mathematics.*

*[13] A. Torokhti and S. Friedland, "Towards theory of generic Principal Component Analysis," Journal of Multivariate Analysis, vol. 100, no. 4, pp. 661-669, Apr.2009.*

*[14] J. M. Lee, S. J. Qin, and I. B. Lee, "Fault detection of non-linear processes using kernel independent component analysis," Canadian Journal of Chemical Engineering, vol. 85, no. 4, pp. 526-536, Aug.2007.*

*[15] B. F. J. Manly, Multivariate statistical methods a primer, 3nd. ed ed. Boca Raton etc.: Chapman and Hall/CRC, 2005.*

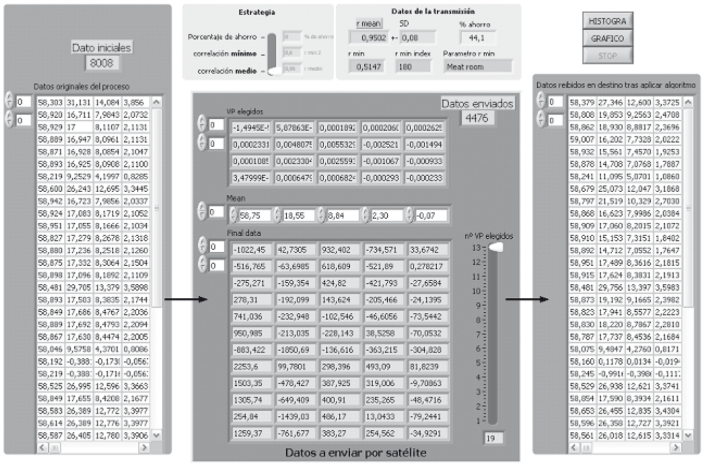


***Figure. 1. Front panel of the designed computer application. In the left side are the real data collected onboard. In the center are the three packets sent by satellite. In the right side it is shown received data.***

*[16] R. L. Mason and J. C. Young, Multivariate statistical process control with industrial applications. Philadelphia etc.: SIAM etc., 2002.*

*[17] D. Peña Sánchez de Rivera, Estadística. Modelos y métodos Vol. 2, 2 ed., 6 reimp. ed. Madrid: Alianza, 1999.*

*[18] G. Prasad, "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, July2007.*

*[19] L. Xiong, J. Liang, and J. X. Qian, "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, Aug.2007.*

TABLE 2
OBTAINED TRANSMISION RESULTS SENDING EACH VESSEL SUBSYSTEM INDEPENDENTLY

| | | **Number of Signals** | **Num 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%** |