7th ECCOMAS Thematic Conference on Smart Structures and Materials
SMART 2015
A.L. Araúo, C.A. Mota Soares, et al. (Editors)
© IDMEC 2015

COMPUTATIONAL STATISTICAL MONITORING OF HYDROCARBON TRANSPORTATION LINES

Magda Ruiz*, **Luis E. Mujica***, **Juan M. Mejía**†

*Escola Universitària d'Enginyeria Tècnica Industrial de Barcelona. Consorci Escola Industrial de Barcelona. Universitat Politècnica de Catalunya. BARCELONATECH. Departament of Applied Mathematics III
Comte d'urgell 187. Barcelona, Spain
magda.ruiz@upc.edu
luis.eduardo.mujica@upc.edu

†Universidad Nacional de Colombia, Facultad de Minas.
Villa Flora, Medellín, Antioquia, Colombia
jmmejiaca@unal.edu.co

Keywords: Transportation lines, Real-Time Transient Model, Principal Component Analysis.

Summary: *Safety and reliability of hydrocarbon transportation lines around the world represents a critical aspect for industry, operators and population. Lines failures caused by external agents, corrosion, inadequate designs, among others, generate impacts on population, environment, infrastructure and economy, besides it may be catastrophically. Therefore, it is essential to constantly monitor operating conditions and hydraulic lines to faults and thus to take measures to mitigate the failure.*
"Real-Time Transient Model" (RTTM) is recognized as one of the most comprehensive, accurate and sophisticated methods to detect leaks in pipelines. This method is based on the numerical solution of the system of equations that describe the phenomenological mass transport, momentum and energy in pipelines, coupled with the thermodynamic behavior of fluids flowing inside. An RTTM makes it possible to calculate mass flow, pressure, density and temperature at every point along the pipeline in real-time with the help of mathematical algorithms. A leak changes the hydraulics of the pipeline, and therefore changes the pressure or flow readings after some time. Local monitoring of pressure or flow at only one point can therefore provide simple leak detection. However, for an advanced monitoring: classification, location or even identification of different kind of leaks, an uni-variate monitoring is not sufficient.
To solve the mentioned drawback, the goal of this work is to develo1p a tool that engages multivariate (monitoring in different points) statistical analysis based on Principal Components Analysis (PCA) and phenomenological simulation based on RTTM to infer the hydraulic behavior of flow lines, fault detection and estimation of fluid integrity discharges to the environment.

1. INTRODUCTION

Hydrocarbons started to play a prominent role in the global economy. The growing demand for hydrocarbons needs to find news hydrocarbon reservoirs. However the extraction, transportation, storage and refining tasks are highly complex. The successful of the oil companies depends of the strategies developed for each task. In this context, hydrocarbons transportation is a factor of strategic relevance.

There are two technical options for transportation: tanks and pipelines. These last are static as opposed to the tanks, which are movable. The pipelines connect source and target units. These are pipes sequentially connected, buried on the terrain or over the surface. Normally, the pipelines operate completely full of product at all times and keeping the motion. In these structures, it can be found a variety of faults due to the accumulation of sediments as wax or paraffin, leakage, rust and inappropriate designs due to the changing elevations along the line, among others.

Hydrocarbons are volatile and flammable and any possible failure in these structures could be catastrophic for the population and environment, leading to severe businesses and structural losses. Therefore, it is essential to rely on fast and accuracy tools for detection of damages in the structure in order to proceed to control and mitigate the problem.

To establish a monitoring system feasible to transport the hydrocarbons through pipelines is not a simple task for any company. The problem is increasingly complex in order to the optimize pumping and minimize costs associated with the operation while maximizing reliability. This type of study is relatively new (last decade), since the technology developed fifty years ago is still implemented. Nowadays, the market dynamics and environmental standards have demanded the application of more advanced techniques. These new techniques have proposed simulations of events with different work environments and common faults. Consequently, situations of high environmental and economic risk have begun to be quickly evaluated. However, currently there are no representation approaches that allow to locate leakages efficiently and quickly where the pipeline is losing hydrocarbon [1].

In this way, the methodology developed in this work uses Real-Time Transient Model (RTTM) and Multivariate Statistical Analysis (MSA) based on Principal Component Analysis (PCA). RTTM describes the phenomenological mass transport, momentum and energy in pipelines. Then the flow, pressure, density and temperature along the pipeline can be obtained. MSA is a widely compress tool for feature extraction which maximize the variance and minimize the correlation among the variables. The goal of this work is to detect and localize leakages by means of simulations of the hydrocarbon flux in undamaged and damage conditions. To be more realistic, the mathematical model includes the dynamic behaviour of the pressure in different locations for dead oil in horizontal topography.

The paper is organized as follows: Description of hydrocarbon transportation, RTTM and PCA are described in Materials (Section 2). Next, the methodology is presented in Section 3 where the procedure to leakages location is explained. In Section 4, results are presented and analyzed. Finally, some Conclusions are summarized.

2. MATERIALS

2.1 Hydrocarbon transportation

Once a hydrocarbon reservoir is located, the exploitation starts. When the hydrocarbon is extracted, immediately after this should be transferred to refining centers. The transportation of hydrocarbons is carried out in two phases: Crude hydrocarbon from the reservoir to the refinery and, final product from the refinery to centers of consumption. In both phases, the need of transportation over long distances can be solved by using two possible solutions: By tanks or by pipelines. In the first case, the conveyance overland are on roads (trucks) or railways and, by sea in tankers. The choice between the two transport options depends on the investment, operating cost and mainly, the certain of the continuity of the hydrocarbon flux. Additionally, two extra analysis have been studied: technical and strategic. The technical term makes emphasis on the reliability of the structure., while the strategic terms makes relation to political issues.

To select whatever long-distance transport solution, it must take into account the fact that new resources are nearly always found far away from the centers of consumption. In this way, the hydrocarbon transport becomes a crucial factor in the exploration strategy of the oil companies, in terms of its technical feasibility and the economic competitiveness of the available transport solution. Thereby, this work is focused in the transportation of hydrocarbons by pipelines. Pipelines are certainly an inflexible option in comparison with the tanks since they need a very high initial investment. Due to the permanent nature of the pipeline structures they are vulnerable, needing an active and effective protection across hostile territory. However, this is the best option when the reservoir is located far away from the destination and/or the surrounded topography prevents the overland transportation.

Normally, pipelines are neither buried or under sea and do not interfere or at least should not interfere with human activities. But, it comes to light only when an accident occurs which causes large economic losses, environmental disasters and possible loss of life. In consequence, pipelines are required to be monitored as soon as possible to avoid the accumulation of sediments (wax or paraffin), leakage or rust. In this context, the monitoring in this kind of structures has to be effective due to the hydrocarbons are highly flammable. There are several varieties of hydrocarbons with specific physical properties such as density, viscosity and composition. For instance: Light crude oil: has a low density and flows freely; Heavy crude oil: does not flow easily; Natural gas; Bitumen; among others.

2.2 Real-Time Transient Model - (RTTM) for leakage detection

To detect leaks, comparison between inlet and outlet flow measurements is not enough. These measurements differ significantly during start-up and shutdown of the pump (pressure in the pipeline changes) and even during a period of 'stationary' operation due to omnipresent transients in the pipeline. To avoid false alarms, the minimum detectable leak rate should be higher than the difference between inlet and outlet flow during normal operation. Knowing that transient can be significant, the minimum detectable leak rate has to be relatively high.

To overcome these limitations Real-Time Transient Model (RTTM) based systems calculate the flow in the pipeline from the pressure and temperature at the inlet and outlet. Transients are present along the pipeline, but they are also in the system. The calculated flow is then compared to the measured flow. This difference should be around zero, otherwise a leak is present. Such a difference is much easier to identify and is more reliable as an indicator. It permits a lower minimum detectable leak rate and therefore a low number of false alarms per year.

Localization of leakage is more than comparison between simulated and measured flows, from the dynamic of these flows it can be inferred the localization of the leakage, and even its magnitude. One option is to develop an inverse RTTM able to calculate parameters of the pipeline by using the measured flow. However, if the calculation of flows by means of RTTM is computational expensive, the inverse calculation is even more.

The simulated cases correspond to a pipeline connects from a hydrocarbon reservoir to a single destination (one way) that transport only one hydrocarbon (Heavy crude oil) in horizontal topography. These phenomenological simulations reproduce as closely the response (flow and pressure) of the pipeline. The simulation contains information to optimize the pumping rate, the momentum and energy of the pipeline. Additionally, these kind of simulations have a high number of inputs and constraints to consider that growing exponentially with the level of detail to get in the pipeline. Each phenomenological modelling simulated data set has the same length, same outputs variables and environmental pressure. However, the main disadvantage of RTTM method is its high computational cost. The other option is to simulate several scenarios by using RTTM and train some kind of classifier or predictor with the simulated measurements.

The first phase of our complete proposed methodology under development is presented in this work. We have focused on carrying out simulations of pressure along a pipeline using RTTM and applying Principal Component Analysis as a tool to recognize hided patterns which allow classify leakages in different locations and different magnitudes. As a consequence, a model of a pipeline with a total length of 80 Kilometers in horizontal topography and transporting heavy crude oil is built. 9 simulations with undamaged condition with 152 samples per 4 variables that belong to few days of operation are developed. These variables are the pressure at 10 Km, 30 Km, 50 Km and 70 Km. Besides, 9 simulations with a length of 162 samples (each one) for three leakage (1 inch, 3 inches and 5 inches) have been simulated in three different locations (20 Km, 40 Km and 60 Km), without taking into account the possible degradation of the pipeline for its use during the fault. In Table 1, it can be found a summary of the main parameters configured in the simulation.

2.3 Principal Component Analysis

Principal Component Analysis is widely used in these kind of multivariate problems since it allows represent graphically as effectively as possible observations belongs to a general m-dimensional space in a small dimensional space (r) [3]. Besides, PCA allows transform original variables, usually correlated, to new uncorrelated variables, making easier its interpretation. The original data are organized in m variables (columns) and n samples per variable (rows) in a data

Table 1. RTTM model for Heavy crude oil in horizontal topography

Fluid	Heavy crude oil
Pipe length (Km)	80
Pipe diameter (in)	22
Temperature (°F)	59
Environmetal pressure (psia)	14.7
Flow rate (MSTB/d)	60.4
Number of pressure gauges	4
Pressure gauges localization (Km)	10, 30, 50, 70
Leakages size (in)	1, 3, 5
Leakages localization (Km)	20, 40, 60
Undamaged simulations	9
Damage simulations	9

matrix named **X** as follows:

$$X = \begin{bmatrix} x_{11} & x_{12} & . & . & . & x_{1m} \\ x_{21} & x_{22} & . & . & . & x_{2m} \\ . & . & & & & . \\ . & . & . & . & . & . \\ . & . & & & & . \\ x_{n1} & x_{n2} & . & . & . & x_{nm} \end{bmatrix}. \tag{1}$$

The goal of PCA is to find a subspace with dimension lesser than m such that projecting into it, the new variables keep its structure and minimize the distortion. In other words, a linear transformation orthogonal matrix **P**, which is used to transform the original data matrix **X** into the form

$$\mathbf{T} = \mathbf{XP}. \tag{2}$$

In the literature, it can be found that the r-dimensional space ($r \leq m$) that represents better the original data is defined by the eigenvectors associated with the highest eigenvalues of the covariance matrix of the observations as follows:

$$\mathbf{C}_X = \frac{1}{n}\mathbf{X}^T\mathbf{X}, \tag{3}$$

$$\mathbf{C}_X\mathbf{P} = \mathbf{P}\Lambda, \tag{4}$$

where $\mathbf{C}_X$ is the covariance matrix of the original data **X**, the eigenvectors of $\mathbf{C}_X$ are the columns of **P**, and the eigenvalues are the diagonal terms of Λ (the off-diagonal terms are zero). The eigenvectors p_j forming the transformation matrix **P** (its columns) are sorted according to the eigenvalues by descending order, the eigenvector with the highest eigenvalue represents the most important pattern in the data with the largest quantity of information. In Equation 2, Columns of **P** are called the Principal Components (or loading matrix in other references) of the data set and **T** the projected or transformed matrix to the principal component space (or score matrix in other articles).

In the full dimension case (using all the n principal components), this projection is invertible (since $\mathbf{PP}^T = \mathbf{I}$) and the original data can be recovered as $\mathbf{X} = \mathbf{TP}^T$. But, PCA also seeks to reduce the dimensionality of the data set **X** by choosing only a reduced number r of principal components ($r < n$). Now, with **T** given by the reduced matrix **P**, it is not possible to fully recover **X**, but **T** can be projected back onto the original m-dimensional space and obtain another data matrix as follows:

$$\hat{\mathbf{X}} = \mathbf{TP}^T. \tag{5}$$

Therefore, the original data matrix **X** can be decomposed by the projected back data $\hat{\mathbf{X}}$ and the residual error matrix **E**, which describes the variability not described by the model as follows:

$$\mathbf{X} = \mathbf{TP}^T + \mathbf{E}. \tag{6}$$

Two well-known statistics are commonly used to this aim: the Q-statistic (or SPE-statistic) and the Hotelling's T^2-statistic (D-statistic). Q-statistic is based on analyzing the residual data matrix **E** to represent the variability of the data projection in the residual subspace. It denotes the change of the events that are not explained by the model of principal components. The Q-statistic of the i-th sample or experiment (row vector x_i of data matrix **X**) is defined as follows:

$$Q_i = e_i e_i^T = x_i \left(\mathbf{I}\text{-}\mathbf{PP}^T\right) x_i^T. \tag{7}$$

where e_i is its projection into the residual subspace (row vector of residual data matrix **X**).

T^2-statistic is based in analyzing the score matrix **T** to check the variability of the projected data in the new space of the principal components. The T^2-statistic of the i-th sample (or experiment) is defined in the form:

$$T_i^2 = t_{si} \Lambda^{-1} t_{si}^T = x_i \left(\mathbf{P} \Lambda^{-1} \mathbf{P}^T\right) x_i^T. \tag{8}$$

where t_{si} is its projection into the new space (row vector of the score matrix **T**) [4][7].

3. METHODOLOGY

A leak changes the hydraulics of the pipeline, and therefore changes the pressure or flow readings after some time. Local monitoring of pressure or flow at only one point can therefore provide simple leak detection. It is only useful in steady-state conditions, however, with the objective of classifying, locating or even identifying different kind of leaks, an uni-variate monitoring is not sufficient.

The methodology that has been previously used by the authors for a multivariate analysis always include information related with the undamaged structure (baseline) to create a statistical model based on PCA. Afterwards, data collected by sensors when structure need to be assessed are projected into the new space given by the PCA model. These projections provide information about how these new measurements are different to the baseline, therefore to know whether the structure still keep in pristine condition or not (damaged), how it has changed and, potentially to distinguish among different kind of damages [4] [5] [6][7].

In the current work, information of the "healthy" structure (no leaks) is given by simulation of the pressure in the different points previously mentioned when the system is operating in normal conditions considering all parameters included in the RTTM. This initial information is organized and arranged in the matrix X as shown in Equation 1 where the number of variables is given by the number of sensors $(m = 4)$. The number of samples is given by number by samples per simulation times the number of simulations $(n = 152 \times 9 = 1368)$. According to Equations 3 and 4 the statistical PCA model is calculated (Transformation matrix or loadings denoted by **P**).

Data acquired by simulations of the structure by each leak (defined in Section 2.2) are organized in the same way, the new matrices $X_{l1}, X_{l2}, ..., X_{l9}$, with dimension 162 samples $\times$ 4 sensors, are projected into the PCA model previously calculated (Equation 2). Besides, statistical indices Q and $T^2 - statistics$ are also determined (Equations 7 and 8). Finally, scatter plots of the first two scores and indices are depicted.

4. Results

Firstly, the simulated measurements of the different sensors are analyzed to verify whether the leakage can be detected and identified. It can be seen from Figure 1 the measurements by each sensor in normal operation (no leaks). On the other hand, Figure 2 shows the profile by each sensor at the different simulated leaks. Color and line belong to the location of the leaks (red dotted line to 20Km, blue dash dot line to 40Km and, green dashed line to 60Km). Shape belongs to the dimension of the leakage (square to 1in, circle to 3in and, diamond to 5in). The first samples represent the transportation in normal operation. After that, the pressure is reduced according to the leakage. However, these profiles did not yield any relevant information about the location and dimension of the leakage. Only it can be observed the instant time when the leakages started.

After applying PCA to the baseline data (Pressure in the four sensors in normal operation at 1368 samples) the following variance is captured by the PCA model.

Percent Variance Captured by PCA Model

Principal Component Number	Eigenvalue of Cov(X)	% Variance Captured This PC	% Variance Captured Total
-----	-----	-----	-----
1	2.55e+00	63.64	63.64
2	1.22e+00	30.58	94.22
3	2.31e-01	5.78	100.00

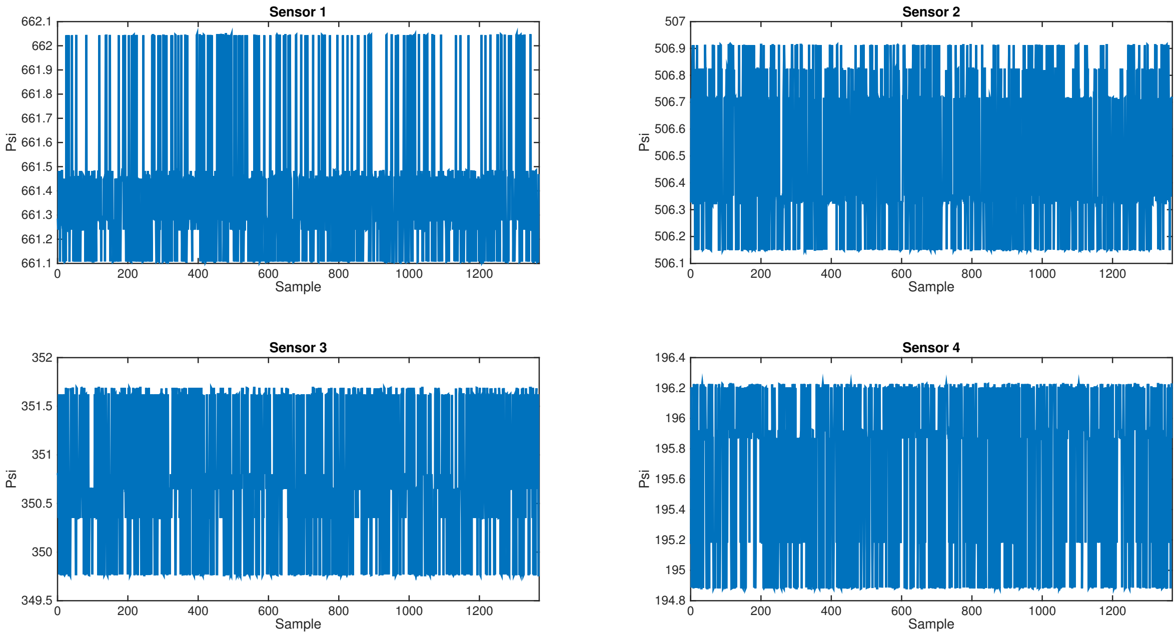

Figure 1. Pressure in the four sensors in normal operation

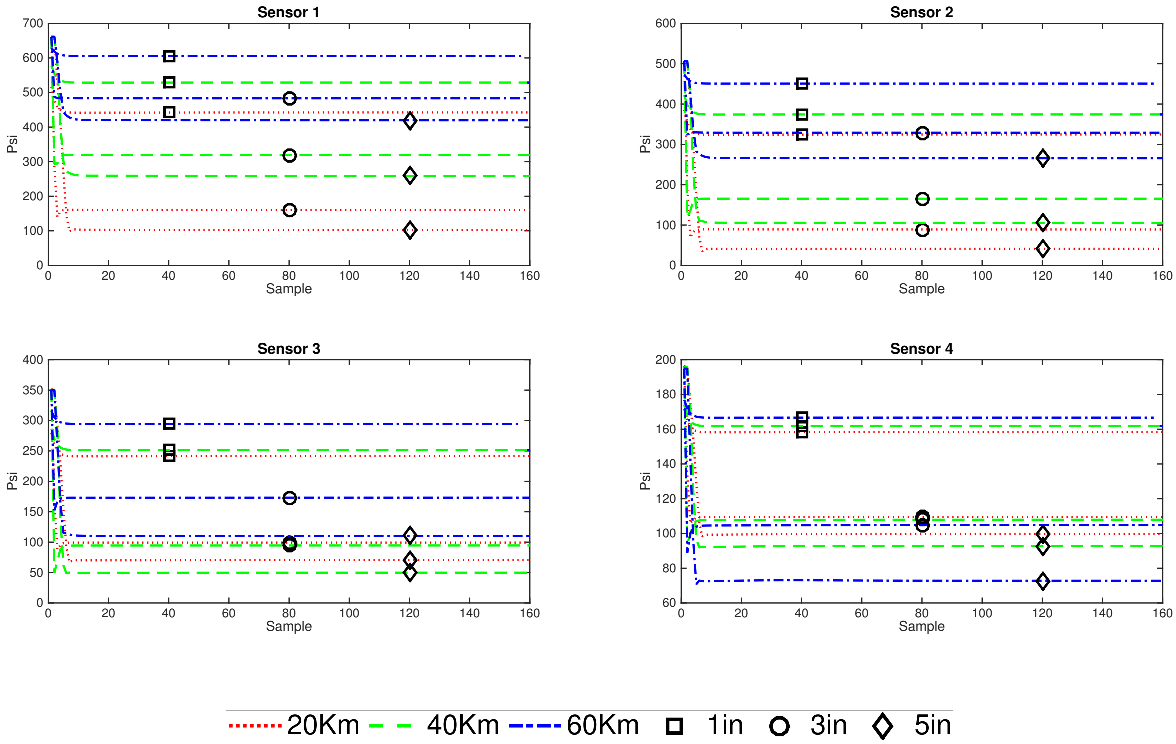

Figure 2. Pressure in the four sensors in presence of leaks

Two principal components are selected and the confidence limit of 95% is determined. Projection of the data from simulations without leaks (pristine pipeline) are depicted in Figure 3. It can be seen that all samples are assorted in six-nine groups, probably they are defined by the random variables includes in each of the nine simulations. All new data projected in the PCA model and located within the confidence limit is considered to be resulting from pipeline in normal operation.

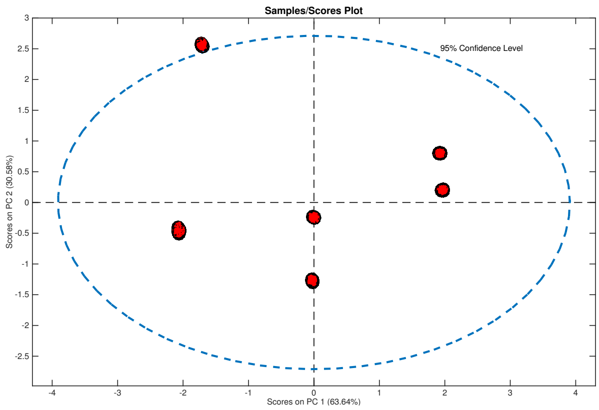

Figure 3. Projection of the measurements resulting from pipeline in normal operation into the first two principal components

Projections into the first and second PC of the data simulated on the pristine pipeline and the different scenarios are depicted in Figures 4 and 5. In both plots, the projection of each sample is represented by different shape and color as shown in the following table:

Table 2. marks used for different pipeline conditions

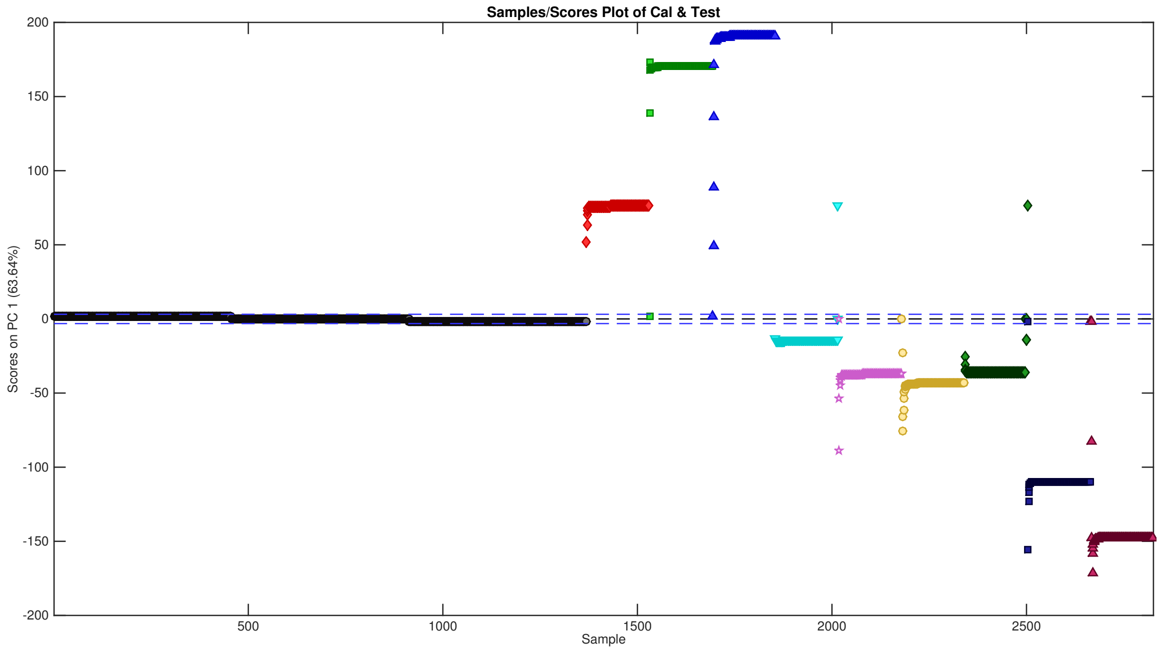

Figure 4. Projection of the measurements resulting from pipeline in different scenarios into the first principal component

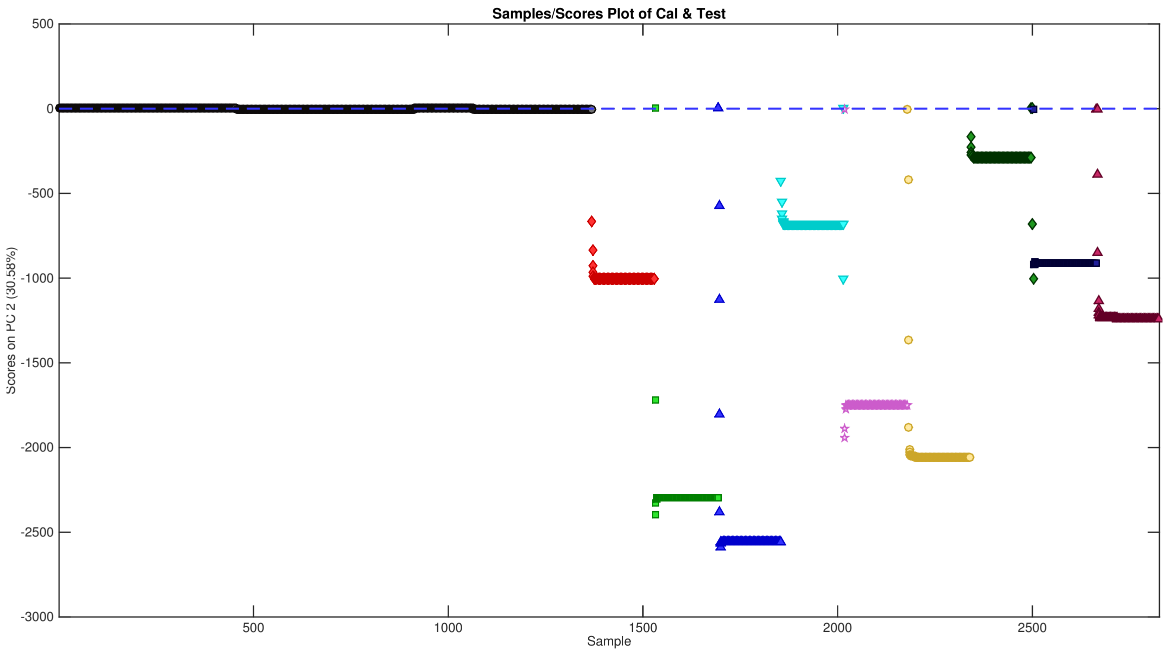

Figure 5. Projection of the measurements resulting from pipeline in different scenarios into the second principal component

In both projections, it can be seen that the presence of leaks is clearly detected, even and the first instant of time (when the leak starts). Besides, it is clearly distinguishable leaks at 20Km and different magnitudes there. Analyzing the plots by separated, leaks at 40Km and 60Km can be confused. A scatter plot of the projection into the first two principal components are shown in Figure 6. When the the pipeline is still operating without leaks, projections fall down into the limit of confidence (see Figure 7, zoom of Figure 6 around the origin of coordinates). If we should be able to show how the projections are changing as the leaks are starting, we can see how the projections are moving away of the origin in a specific direction, once the leak is stable, all the projections fall down in a specific region. Leaks in the same location take the same direction, but the stabilizing region is given by the magnitude (size of the crack). In this way it is clearly feasible a classification or even a localization of any leak considering simulation of some few leak scenarios.

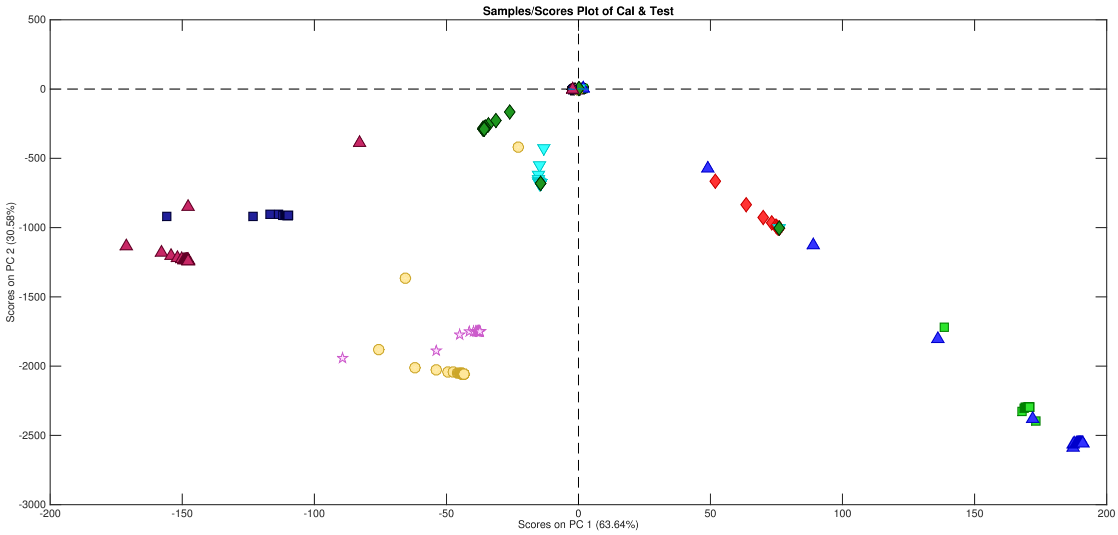

Figure 6. Projection of the measurements resulting from pipeline in different scenarios into the the first two principal components

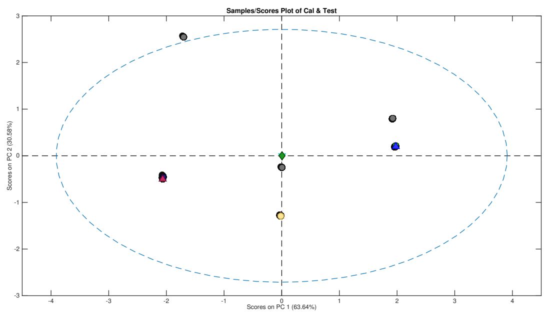

Figure 7. Projection of the measurements resulting from pipeline in different scenarios into the the first two principal components inside of confidence limits

Finally, statistical indices Q and $T^2 - statistics$ are also depicted in Figures 6 and 7 respectively. Information provided by these indices is almost similar to the provided by scores, however, using a log scale, transients in the measurements when the leaks start are more clear.

5. CONCLUSIONS

The first phase of our complete proposed methodology under development is presented in this work. We have focused on carrying out simulations of pressure along a pipeline using RTTM and applying Principal Component Analysis as a tool to recognize hided patterns which allow classify leakages in different locations and different magnitudes.

It is well known that PCA is more than only the principal components (indices Q and $T^2 - statistics$, contribution plots). In this work, only two principal components are selected in order to develop the statistical PCA model.

Observed results show a promising future by applying this methodology. Projecting measurements into the PCA model, it can be seen that leaks in the same location take the same direction, but the stabilizing region is given by the magnitude (size of the crack). In this way it is clearly feasible a classification or even a localization of any leak considering simulation of some few leak scenarios.

The next phase of the whole methodology will be focused on: Simulation of pipeline on different topographies and transporting different kind of hydrocarbons: Besides, validating the approach with data obtained from a real reservoir.

ACKNOWLEDGMENT

The authors thank to the “Escola Universitària d’Enginyeria Tècnica Industrial de Barcelona”, “Consorci Escola Industrial de Barcelona”, “Universidad Nacional de Colombia” and “Ministerio de Ciencia e Innovación” in Spain through the coordinated research project DPI2011-28033-C03-03

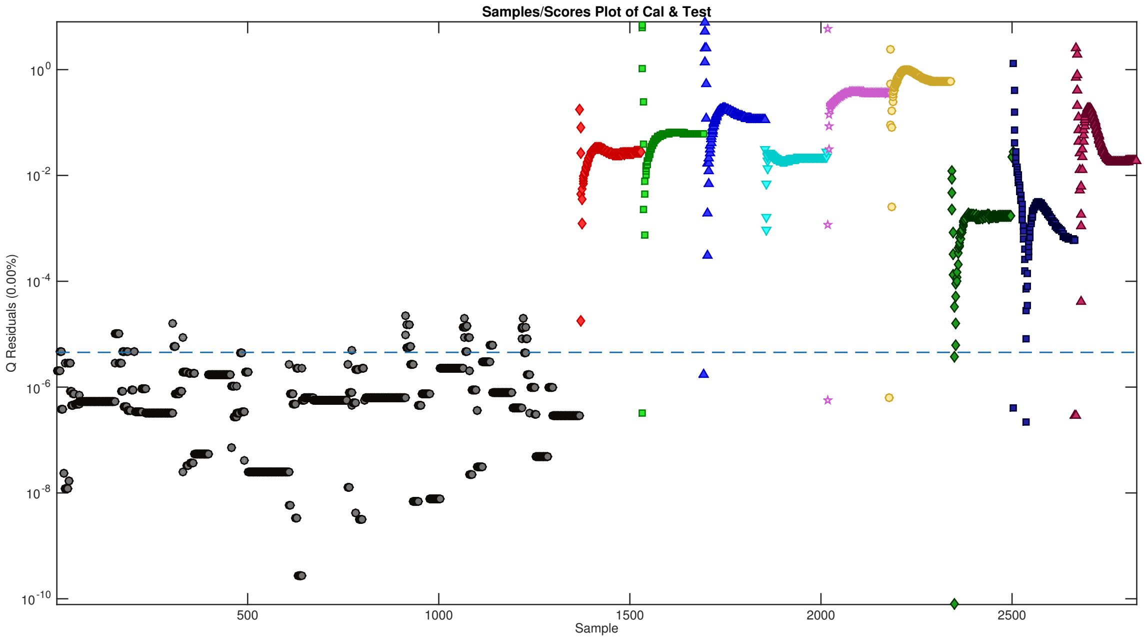

Figure 8. Qtesting2

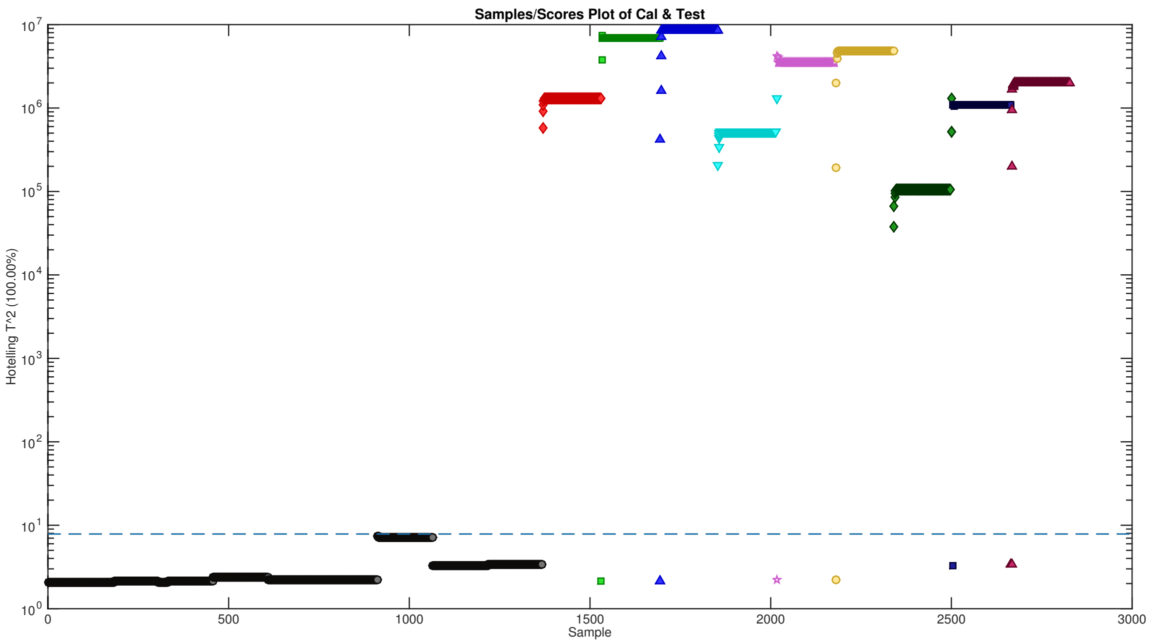

Figure 9. Ttesting2

References

[1] D.C. Cafaro. Programación óptima de operaciones en sistemas de transporte de combustibles múltiples a través de poliductos. *Universidad Nacional del Litoral, Facultrad de Ingeniería Química, Santa Fe - Argentina*, Tesis doctoral, 2009.

[2] E.L. Russell, L.H. Chiang and R.D. Braatz. Data-driven techniques for fault detection and diagnosis in chemical processes. Advances in Industrial Control. *ISBN 1-85233-258-1*, London, 2000.

[3] I.T. Jolliffe. *Principal Component Analysis*. Springer Series is Statistics. Springer-Verlag, second edi edition, 2002.

[4] L.E. Mujica, J. Rodellar, A. Fernández, and A. Güemes. Q-statistic and T2-statistic PCA-based measures for damage assessment in structures. *Struc. Health Monitoring*, 10(5):539–553, 2011.

[5] L.E Mujica, M. Ruiz, F. Pozo, J. Rodellar and A. Güemes. A structural damage detection indicator based on principal component analysis and statistical hypothesis testing. *Smart materials ans structures*, 23(2):025014-1–025014-21, 2014.

[6] L.E Mujica, J. Vehí, M. Ruiz, M. Verleysen, W.J. Staszewski, and K. Worden. Multivariate Statistics Process Control for Dimensionality Reduction in Structural Assessment. *Mechanical Systems and Signal Processing*, 22:155–171, 2008.

[7] M. Ruiz, L.E. Mujica, X. Berjaga and J. Rodellar. Partial least square projection to latent structures (PLS) regression to estimate impact localization in structures. *Smart materials ans structures*, 22(2):025028, 2013.