8th European Workshop On Structural Health Monitoring (EWSHM 2016), 5-8 July 2016, Spain, Bilbao
www.ndt.net/app.EWSHM2016

# Multidimensional big data processing for damage detection in real pipelines using a smart pig tool

Magda Ruiz[1], Luis E. Mujica[1], Santiago Alferez[1], Mario Quintero[2], Rodolfo Villamizar[3]

[1]Department of Mathematics, Escola d'Enginyeria de Barcelona Est (EEBE), Universitat Politècnica de Catalunya (UPC) BARCELONATECH, Barcelona, Spain.
magda.ruiz@upc.edu luis.eduardo.mujica@upc.edu santiago.alferez@upc.edu
[2] Department of Electronics. Research Institute of Corrosion CIC. Colombia mquintero@corrion.uis.edu
[3] Escuela de Ingenierías Eléctrica, Electrónica y de Telecomunicaciones (E3T), Universidad Industrial de Santander (UIS). Grupo de Control Electrónica Modelado y Simulación (CEMOS), Colombia.
rovillam@uis.edu.co

**Key words:** Smart structures, Pipelines, Damage detection, Industrial applications.

**Abstract**

*The history of the hydrocarbons business in Colombia dates back to the early twentieth century where mining and energy sector has been one of the principal pillars for the its development. Thus, the pipelines currently in service have over 30 years and most of them are buried and phenomena like metal losses, corrosion, mechanical stress, strike by excavation machinery and other type of damages are presented. Since it can generate social and environmental problems, monitoring tools and programs should be developed in order to prevent catastrophic situations. However, the maintaining of these structures is very expensive and it is normally developed by foreign companies. In order to overcome this situation, recently the native research institute "Research Institute of Corrosion - CIC (Corporación para la Investigación de la Corrosión)" developed an in-line inspection tool to be operated in Colombian pipelines (especially gas) to get valuable information of their current state along of thousand kilometres. The recorded data is of big size and its processing demand a high computational cost and adequate tool analysis to determine a certain pipeline damage condition. On other hand, the author from UPC and UIS have been bringing its expertise in processing and analysing this type of big data by using mainly Principal Component Analysis (PCA) as an effective tool to detect and locate different damages. In previous papers, multidimensional data matrix was used to locate possible damages along the pipeline, however most of activated points were considered false alarms since they corresponded to weld points. Thus, in this paper it is proposed no considering piecewise weld points (tube sections) and an extension of PCA named Multiway PCA (MPCA) is applied for each each one of the tube sections that form the pipeline. Therefore, if a tube section is found outside from overall indices found by using the MPCA model, an alarm activated in that section and a precise location can be obtained by analyzing only data from that specific tube section.*

## 1 INTRODUCTION

The main objective of Structural Health Monitoring (SHM) is the verification of the condition of a structure in the incipient state guaranteeing the its integrity and hence, increasing the security, reducing costs of maintenance and repair [1]. Although it is not a recent topic, the structural fault detection based on Data Driven Models has recently started and it consists of taking measurements for assessment the current state of a structure [2]. Then, damages can be detected by comparing the current against an undamaged condition previously stored by using current sensor signals attached to the structure. A structure of special research interest is the pipelines since a damage condition represents real human lives, environmental, social and economic impact in a country. Thus, the pipeline safety and reliability is a critical monitoring aspect. Recently, an in-line inspection tool for monitoring Colombian gas pipelines denominated ITION (Inspection of Trends of Integrity and OperatioN) was developed and it is being implemented with very good performances. It is a robust engine with powerful technology that can help to gauge the health and integrity of metallic pipelines without stops in the process during its running. ITION travels through/inside the pipeline storing sensed data used to detect structural conditions along the pipeline. A very important measurement unit that can be installed in this tool consists of an arrangement of Magnetic Flux Leakage (MFL) sensors, which is a useful variable to detect along of a metallic pipe line dents, anomalous weld seams, longitudinal cracks, longitudinal grooves and corrosion. This variable sensed along a long pipeline (30 km for the experimental case analysed in this paper) contain valuable information but its millions of samples demand data compression achieve a reasonable quantity of variables. Thus this problem motivated the present research and here it is presented the use of an alternative tool (Multiway Principal Components Analysis) to locate sections of the pipeline with probable structural damages.

## 2 ITION- SMART IN-LINE INSPECTION TOOL

The Corrosion Research Institute (CIC) from Colombia is a center for developing technology and generating knowledge associated to industrial corrosion problems, especially those related with gas and oil infrastructures. On this direction this center developed a Smart Inspection Tool denominated ITION (Inspection of Trends of Integrity and OperatioN, see Figure 1) and it was used to inspect a 36 km Colombian gas pipeline, where inertial, pressure, temperature, magnetization and magnetic leakage signals were recorded along it with the purpose of detecting subnormal conditions of the structure such as weld failure, geometric deformation, corrosion, mass loss or adding, among others.

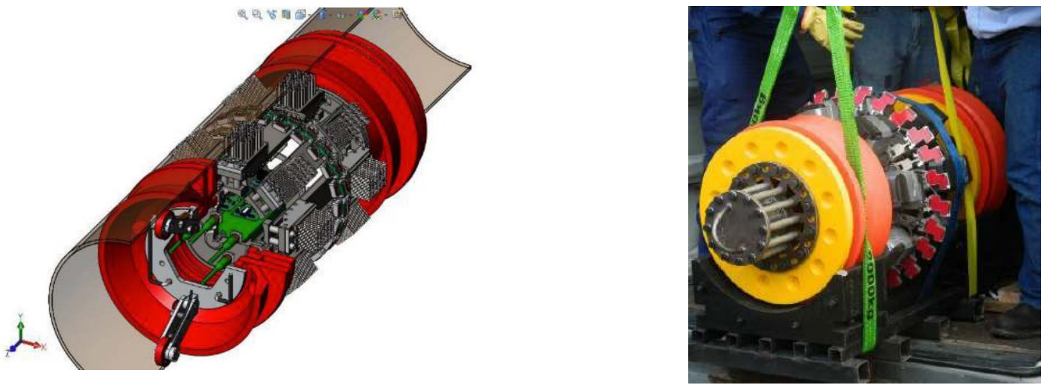

Figure 1: Inspection of Trends of Integrity and OperatioN -ITION-

## 2.1 ITION Configuration

ITION corresponds to a smart pipeline inspection tool that consists of an instrumented vehicle capable of travelling inside a pipeline by propulsion of the transported fluid. It can acquire and record along the pipeline different variables according with the inspection purpose (measuring operational conditions, locating or inspecting damages, defining geometrical profiles) by installing different sensor inside and outside the chassis. Figure 1 presents a basic configuration of this inspection tool, where all electronics are protected by a mechanical housing and designed to fit most of the conventional scraper routinely used in pipeline cleaning processes. The ITION tool includes the next measurement devices: odometer, Inertial Measurement Unit (IMU), accelerometers, calipers, pressure and temperature sensors. However, for inspecting the pipeline studied in this paper, the system has been expanded to incorporate a MFL system that consists of a prototype array of linear transducers that varies its output voltage in response to the perturbed magnetic field applied by permanent magnets. Thereby, next signals are recorded:

| Signals | Variable |
|---|---|
| 1-3 | Inertial Movement |
| 4-5 | Remanent Magnetic fields |
| 6-7 | Pressure and Temperature of the transported fluid |
| 8 | Vibration |
| 9-10 | Calibration Measurements |
| 11-18 | Magnetic Flux Leakage |

Table I: Recorded signals

By using the MFL measurements it is possible to obtain a magnetic profile of the pipeline and to identify, inspect or locate wall conditions, such as welds, mass loss or adding, among others, based on comparison of previous measurements or expert analysis. To understand better the MFL principle, it is presented in next section.

## 2.2 Magnetic Flux Leakage (MFL)

MFL is the most common in line inspection (ILI) technique used for monitoring wall thickness in long carbon steel pipelines in order to detect defects such as mass loss or adding, fitting or non well conditioned welds, associated with the presence of corrosion or other phenomena. According to [3] the MFL technique consists of detecting irregularities in the ferromagnetic pipe material under inspection (i.e. loss or adding material) when a permanent axially oriented magnetic field is applied by means of permanent magnets. Since the magnetic field is perturbed by the material defects a flux leakage outside the pipe is produced and measured by field sensors. For monitoring purpose, three important conditions for the applied field must be accomplished: strong, consistent (it should be measureable along the pipe) and spread out uniform through the pipe. Ferromagnetic materials (such as carbon steel materials) exhibit the hysteresis effect (See Figure 2), when a magnetic field is applied, which is used to detect abnormal conditions in the inspected material. This nonlinear behavior is summarized according to the operation points in the hysteresis curve of figure 2 (see [3] for more details), where zones MFL (Magnetic Flux Leakage), LFM (Low Frequency Magnetic) and RES (Residual Magnetic Field) are the common ones during a MFL based pipeline monitoring.

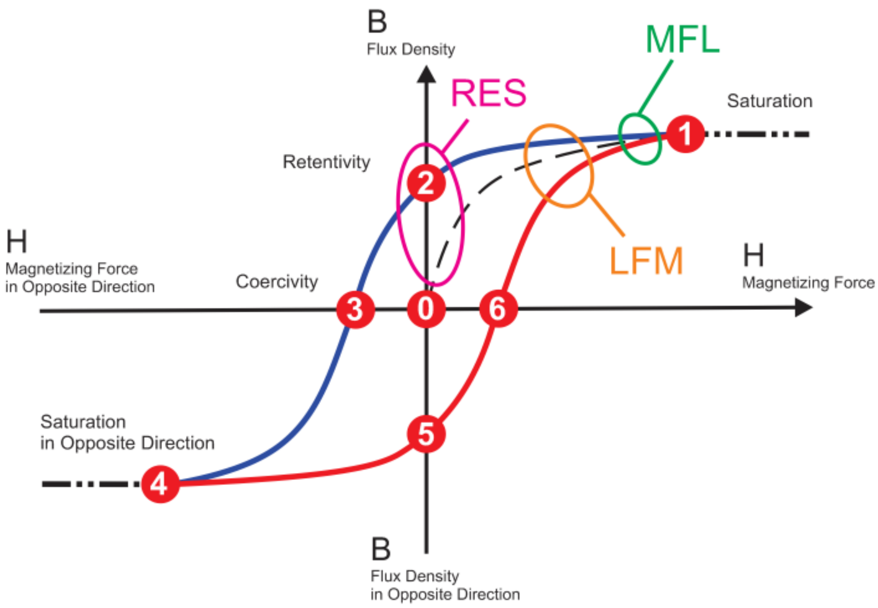


Figure 2: Typical hysteresis loop

## 3 DAMAGE DETECTION METHODOLOGY

The main purpose of an in-line inspection is to ensure an adequate detection, sizing and location of flaws and defects within the pipe wall, in order to keep the integrity of the infrastructure and to establish its current and future state. Different technologies are used such as ultrasound, eddy current and Magnetic Flux Leakage (MFL) to detect different types of defects at different precision levels. MFL is commonly used to inspect long gas pipelines, thus the ITION tool was used to record MFL measurements, among other variables, along a 30 km Colombian gas pipeline. By recording samples by steps of cm, it means to get millions of samples, which also of requiring a high capacity storage system. Therefore, it is necessary an adequate data conditioning and analysis in order to avoid false alarms associated to changes in the time signals that does not belong to material defects. Due to analyse this millions of samples is not possible by human inspection, an alternative technique used by the authors in previous works is the Multivariable Statistical Analysis by using Principal Component Analysis [4]. In [5][6] it was reported the use of PCA to detect abnormal conditions of the same pipeline studied in this paper, where the main contribution consisted on using multivariable in place of univariable statistical analysis on the whole recorded signals along the total length of the pipeline. The main conclusion of these works is that the used statistical indices can detect damages, but also a big number of false alarms are activated by the presences of welds. On other hand, in [7] an artificial neural network is used to automatically detect welds of this pipeline by using acceleration, vibration and magnetic recorded signals.

Thus, in this paper a modified methodology based on PCA is proposed to determine pipeline sections that can contain a possible material defect (to be evaluated by experts) by using the statistics indices computed by independent sections of the pipeline.

### 3.1 Principal Component Analysis

PCA is a statistical tool that allows an easy graphical representation of observations that belong to a general *m*-dimensional space in a small dimensional space (*r*) by transforming original variables (usually correlated) to new uncorrelated variables. The goal of PCA is to find a subspace of a shorter dimension than *m* that conserves the original structure, minimizes the redundancy and maximizes the variance. It means to find an orthogonal transformation matrix *P* to transform the original measurements matrix **X** into the form:

$$\mathbf{T} = \mathbf{XP}. \tag{1}$$

It has been demonstrated that the *r*-dimensional space that better represents **X**, correspond to the eigenvectors **P** associated with highest eigenvalues (diagonal values of **Λ**) of the covariance observations matrix $\mathbf{C_X}$ obtained by equations 2 and 3.

$$\mathbf{C_X} = \frac{1}{\mathrm{n}}\mathbf{X^T X}, \tag{2}$$

$$\mathbf{C_X P} = \mathbf{P\Lambda}, \tag{3}$$

where the columns of **P** are denominated the Principal Components (or loadings) and **T** the projected or transformed matrix to the principal component space (or score matrix). Since only a reduced number of *r* principal components are selected, it is not possible to fully recover ***X,*** however **T** can be projected back to the original space *m* and a new measurements matrix $\widehat{\mathbf{X}}$ is obtained as follows:

$$\widehat{\mathbf{X}} = \mathbf{TP^T}. \tag{4}$$

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

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

***Statistical Indices****:* Two well-known statistics indices are commonly used for analysis purposes: Q-statistic (or SPE-statistic) and the Hotelling's $T^2$-statistic (D-statistic). The first one represents the variability of the data projection in the residual subspace and denotes changes of events that are not explained by the principal components. The Q-statistic of the *i*-th sample or experiment (row vector $x_i$ of data matrix **X**) is defined as follows:

$$Q_i = e_i e^T = x_i(I - PP^T)x_i^T, \tag{6}$$

where $e_i$ is its projection into the residual subspace (row vector of residual data matrix $\widehat{\mathbf{X}}$). $T^2$-statistic is based on 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:

$$T_i^2 = t_{si}\Lambda^T t_{si}^T = x_i(\mathbf{P\Lambda^{-1}P}^T)x_i^T, \tag{7}$$

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

### 3.2 Methodological steps

The proposed methodology consists on using the signals recorded by the smart pig ITION and applying multivariable statistical analysis separately to each section of the pipeline instead of the total length. That is the main difference comparing to previous work. Thus individual and independent statistical indices are obtained for each zone by following the procedure depicted in Figure 3.

i) *Selection of variables to analyse:* Since eighteen variables were measured by ITION for the inspected gas pipeline (inertial (3), remnant magnetic field (2), caliper (2) and MFL (8). Temperature (1), vibration (1) and pressure (1)), the three last variables were no considered for the multivariable analysis since they are not directly associated with any kind of defect to be analysed. Although the variable directly related to material changes corresponds to MFL, it is also influenced by the dynamic behaviour of the inspection tool and the applied magnetic field, then the multivariable statistical analysis was applied on the first fifteen variables.

ii) *Sections definition and samples retrieving*. A section for the present study was defined as each pipeline portion that exists between two consecutive welds without including the last ones. Thus a section can have different geometric shapes and sizes. Welds were not considered in the statistical analysis since they generate high MFL measurements that alter the statistical indices and hide the indices of interest. To define each section, the MFL and odometer measurements joint to the welds chart given by the owner of the gas pipeline were used. It consisted of identifying each weld point reported by the owner with samples position where typical dynamical behaviour of the MFL signal associated to welds and then retrieve every one of the samples between two consecutive solds. This was necessary since the odometer contains error measurements and the reported weld position does not match with the distance measured by the odometer.

iii) *Selection of number of observations by section.* By using Principal Component Analysis, it should be ensured that the number of observations $m$ be greater than the number of correlated variables $n$. For the present analysis each sample of each recorded signal is considered as a correlated variable, thus the number of samples $l$ by sensor for each observation is obtained by the following condition: $m = L/l \geq n = sl$, where $L$ is the number of samples by sensor for each section and $s$ is the number of sensors (15 for this case). Thus the number of samples $l$ to be selected for each observation is given by: $l \leq \sqrt{L/15}$ and it can be different for each section since the length and number of recorded samples is variable.

iv) *Measurement matrix (X) organization.* For each section an independent $m \times n$ measurement matrix **X** is obtained, where a row belongs to $15l$ samples of the analysed signals. The position of each one of the samples is not important since PCA consider each sample as a correlated variable.

v) *Computation of scores, $T^2$ and $Q$ statistical indices.* For each section the PCA analysis is applied such that scores **T** and loadings **P** matrices are obtained and used to obtain the $T^2$ and $Q$ statistical indices for each section.

vi) *Section activation based on maximum Qs*. Since the goal of the multivariable statistical analysis is to create alarms on sections where greatest statistical indices are presented, it was conserved the highest Q index by section and used to compare it with every one of the maximal section Q indices. The activated maximal Qs are those greater than the mean value of the maximal Qs plus two times the standard deviation.

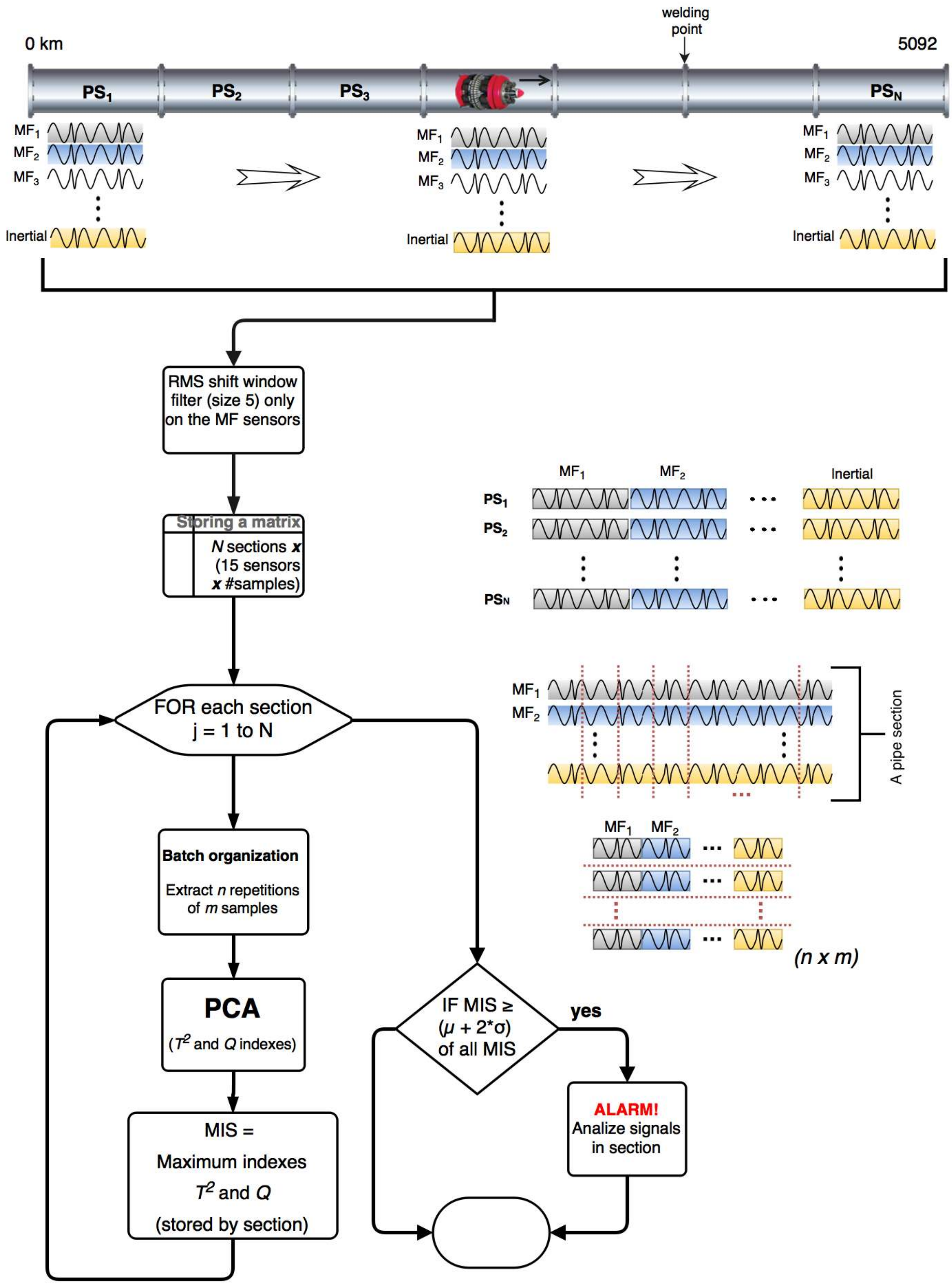


Figure 3: Methodological Steps followed to activate pipeline sections with potential changes

# 4 EXPERIMENTAL RESULTS AND DISCUSION

The main purpose of this paper was to experimentally validate the methodology on each sections of an entire pipeline and to compare maximal statistical indices between them in order to detect those sections with some defect or abnormal situation by reducing the analysis from a huge quantity of samples to only one index ($Q$ statistics).

One part of the Colombian industrial gas pipeline network was inspected on September 2012 by the ITION tool, whose recorded variables were used to experimentally validate the methodology proposed here. The pipeline section is 12" nominal diameter, 36 km inspected and 3000 welds. Around 14.000.000 samples were recorded, but 10.139.436 samples were considered (320 MB) (Initial static samples were disregarded).

A first experimental validation of the methodology consisted into analyse the first 5 of the 36 km total pipeline length (434 of 3000 sections). Then, each one of the above mentioned methodological steps were applied.

i) By selecting 15 of the 18 recorded variables (Inertial, Remand Magnetic Field, Caliper and MFL), the total number of samples by sensor for the total length of the pipeline section is 2.538.273 and the maximum number of samples by sensor in a section is 107.177.

ii) By applying this step over the 5 km of the pipeline most of the weld points were identified and located at a sample were the dynamic behaviour of the MFL is that of a weld, however some points were no totally identified and it was necessary a visual inspection of the plotted MFL signal and contrasted with the chart weld points. Each point was identified by computing the sum of the square rms value of each MFL signal (computed for a window of 5 samples) and at each weld chart point, where the sample selected corresponds to that where occurs the maximal value of the MFL between an interval of the half point of the current weld and previous one and the half point of the current weld and the next weld. A graphical representation of the type of plots used is presented in Figure 4.

iii and iv) Once updated the sample position for each weld points and based on the condition of minimum number of observations, a $m\times151$ measurement matrix was obtained for each one of the 434 sections. For example, for the section with the longest sample number (107.177) the number of observations was 1275 (it is rounded to the nearest integer toward minus infinity) and the length of each row is 15×84=1.260.

v) For each measurements matrix $\mathbf{X}$, PCA was applied and scores, $m$ $T^2$ and $Q$ indices were obtained for each section. An example of the $T^2$ indices obtained for section is presented in figure 5.

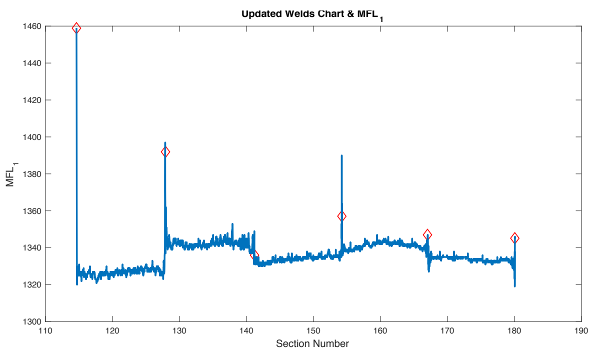


Figure 4. MFL1 and weld chart points correspondence to weld chart

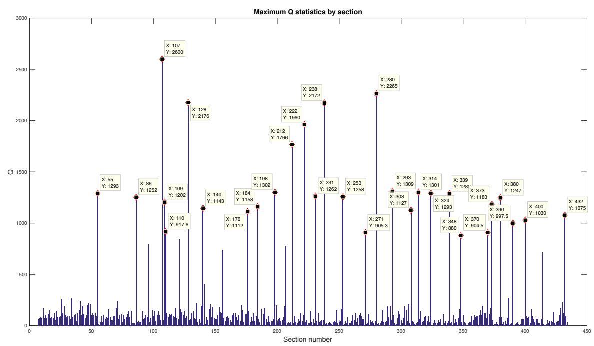


Figure 5. Activated Maximum Q statistics

vi) Once the statistical indices are computed for each section and by applying the established condition, 29 sections were activated when the maximum $Q$ indices were analysed (see figure 5). This means that in that sections a differentiated variation of one or more variables occurs and an expert should evaluate a potential variation in the operation condition. For the activated sections a similar pattern is recognized and it corresponds to abrupt changes in specific points of MFL sections that according to the inertial signals, they do no correspond to movements of magnets or tool. Figure 6 presents the dynamical behaviour of section 107 where occurs the maximum $Q$ of all sections. In contrast, figure 7 presents the dynamic behaviour of section 291 with the minimal value of $Q$ where it can be observed a significant variation of MFL signals, but it can be observed that it is associated to movements in the tool, reported by inertial variables.

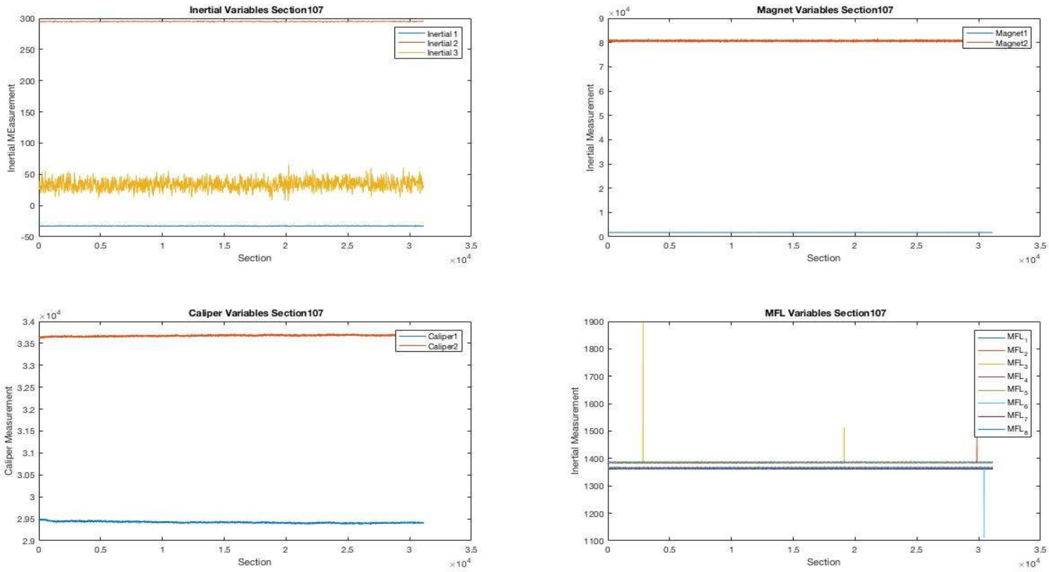


Figure 6. Dynamic behaviour of section 107 with maximum Q statistics

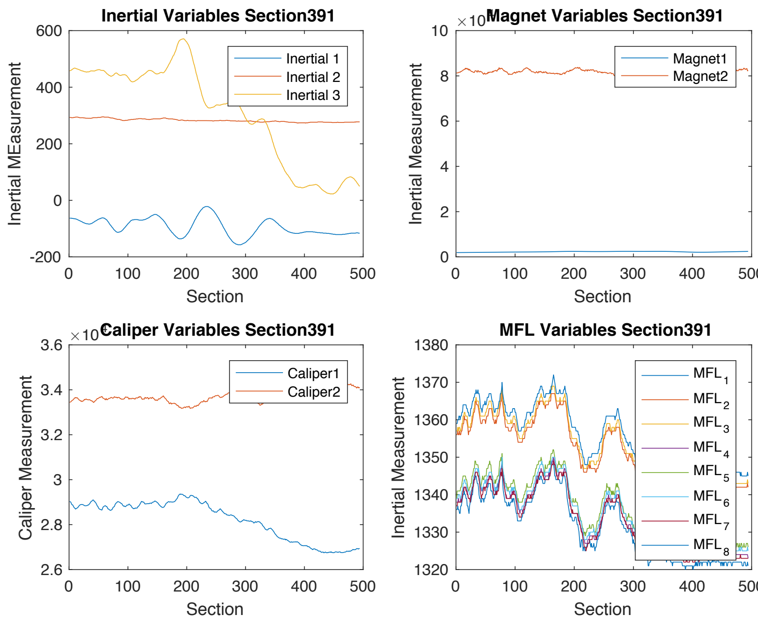


Figure 7. Dynamic behaviour of section 291 with minimum Q statistics

## 5. DISCUSION AND CONCLUSIONS

Promising results by using multivariable statistical analysis in the analysis o a big number of data collected from a in-line inspection tool executed on a Colombian gas pipeline were previously reported, however some drawbacks such as false alarms associated to the presence of welds were continuously reported. Thus, this work was focused on to solve this drawback by excluding the weld effect on the MFL signals and to enhance the sensibility of the statistical indices.

By applying multivariable statistical analysis on signals recorded along of each pipeline section, it was demonstrated that is it possible to exclude the effect of the weld on the MFL signals and to observe small and punctual events such as a fast change of some of the MFL signals, which could correspond to punctual defects such as mass loss at a specific point. Also it was demonstrated the robustness of the PCA analysis to exclude significant changes in variables such as inertial or MFL variables that do not correspond to damages but also correspond to normal dynamics behaviour of the inspection tool travelling along a pipeline that transport a fluid to high pressure.

### Acknowledge

The authors would like to thank the support from Colombian Government (Departamento Administrativo de Ciencia y Tecnología Francisco José de Caldas – COLCIENCIAS), the Catalonia Government through the project 2014SGR859 and the Colombian Research Institute of Corrosion - CIC-.

## REFERENCES

[1] C.R Farrar and K. Worden. An introduction to structural health monitoring. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, (365), 1851. Pp. 303—315, 2007.

[2] C.P. Fritzen, P. Kraemer. Self-diagnosis of smart structures based on dynamical properties. *Mechanical Systems and Signal Processing*, 23, 1830-1845 (2009).

[3] J.B Nestleroth. and R.J Davis. The effects of remanent magnetization on magnetic flux leakage signals. *Review of Progress in Quantitative Nondestructive Evaluation*. 14: 483-490. 1995.

[4] L.E Mujica, J. Rodellar, A. Fernández, and A. Güemes. Q-statistic and T2-statistic PCA-based measures for damage assessment in structures. *Structural Health Monitoring*, 10(5):539–553. 2010

[5] M.L. Ruiz, L.E. Mujica, and M. Quintero. and J.F. and S. Magnetic Flux Leakage and Principal Component Analysis for metal loss approximation in a pipeline. *Journal of Physics: Conference Series*, 628(1), 12027. Retrieved from http://stacks.iop.org/1742-6596/628/i=1/a=012027. 2015

[6] M.L. Ruiz, L.E. Mujica, M. Quintero, J. Florez. In-line inspection of pipelines by using a smart Pig (ITION) and multivariate statistical analysis. *IWSHM 10th International Workshop on Structural Health Monitoring.*: Stanford University (USA), 1-3 September, 2015.

[7] C.J. Arizmendi, W.L. Garcia and M.A. Quintero. Automatic welding detection by an intelligent tool pipe inspection. *Journal of Physics: Conference Series*, 628(1), 12082. Retrieved from http://stacks.iop.org/1742-6596/628/i=1/a=012082. 2015