Multidimensional data statistical processing of magnetic flow leakage signals from a Colombian gas pipeline

Luis Eduardo Mujica¹, Magda Ruiz¹, Rodolfo Villamizar²

Abstract
The hydrocarbons industry in Colombia is one of the principal pillars for the Colombian economy, representing around 5% of its gross domestic product. Since petroleum reserves have decreased, gas becomes in one main alternative for economical growth. However, current gas pipelines have been in service for over 30 years and some of them are buried and phenomena like metal losses, corrosion, mechanical stress, strikes by excavation machinery and another type of damages are presented. The maintenance program of these structures is typically corrective type and is very expensive. To overcome this situation, the native research institute “Research Institute of Corrosion - CIC (Corporación para la Investigación de la Corrosión)” recently developed an in-line inspection tool to be operated in Colombian gas pipelines to get valuable information of their current state along of thousands of kilometers. A huge quantity of data is recorded (including tool movement, magnet, magnetic flow leakage-MFL, caliper signals) which demand a high computational cost and an adequate tool analysis to establish the current pipeline structural health condition. In this sense, authors have shown in several works that Principal Component Analysis (PCA) is an effective tool to detect and locate abnormal operational structural conditions from multidimensional data. In a previous analysis, multidimensional data were used to locate possible damages along the pipeline. However, most of the activated points belonged to weld points. Then, in this paper it is proposed to use the Root Mean Square (RMS) value of MFL signals to separate these points and to obtain sets of signals by sections removing the welds, and then Multiway PCA (MPCA) is applied for each set of signals of each gas pipeline section. The maximum values of damage indices (Q and $T^2$ –statistics) of each section are conserved to activate the sections of the gas pipeline with more probability of damages and then, they must be evaluated by experts.

INTRODUCTION
One of the Oil & Gas industry goals is to improve its capacity of continuous monitoring their infrastructures. Its value-add is the capacity of early fault detection in order to take timely decisions that guarantee their integrity. Those structures used for hydrocarbons transport, distribution and storage, which are mainly constituted by ferromagnetic materials, are subject to environmental and operational conditions that gradually modify its original state. These changes are usually beyond the visual range until the imminent collapse. In the case of pipelines, assessment of the extent of metal-loss defects such as corrosion, cracks, and dents is not a straightforward task since the pipelines are usually buried underground, and thus inaccessible for direct inspection. In-line inspection (ILI) tools have been developed to perform an in-service inspection of pipelines to collect information about many undesirable situations as defects, sediments, or leakages among others¹.

In-line Inspection (ILI) involves the internal evaluation of pipes and pipelines using "smart pigs", which travel along of them and are equipped with sensors that utilize different nondestructive techniques according to each particular situation. Among the existing technologies, it can be found: Eddy Current (EC), Ultrasonic Testing (UT), Electromagnetic Acoustic Transducer (EMAT) and Magnetic Flux Leakage (MFL)². Eddy Current (EC) is an electromagnetic technique that lies in inducing electromagnetic eddy currents in the specimen by means of an energized coil. These currents tend to oppose the original magnetic field affecting the impedance of the coil. In the presence of flaws or material variations, this impedance is changed and easily measured. The disadvantage for ILI applications is its slow response comparing with the pig’s speed. Ultrasonic Testing (UT) uses sound waves of short wavelength and high frequency. Its main difficulty lies in coupling to the pipe wall that limits the quality of the data as well as the size of the defect to be detected. Electromagnetic Acoustic Transducer (EMAT) consists of a coil in a magnetic field at the internal surface of the pipe wall. Alternating current placed through the coil induces a current in the pipe wall, causing Lorentz forces that in turn generate ultrasound.

¹Control, Modelling, Identification and Applications (CoDAlab), Departament de Matemàtiques, Escola d’Enginyeria de Barcelona Est (EEBE), Universitat Politècnica de Catalunya (UPC), Spain.
²Universidad Industrial de Santander (UIS), Escuela de Ingenierías Eléctrica, Electrónica y de Telecomunicaciones (EET), Colombia.

Corresponding author:
Luis Eduardo Mujica, Universitat Politècnica de Catalunya (UPC), Campus Diagonal-Besòs, Eduard Maristany, 6-12, 08930 Sant Adrià de Besòs, Barcelona, Spain
Email: luis.eduardo.mujica@upc.edu
As a drawback, EMAT needs to be located very close to the specimen and high frequencies cannot be applied, as a consequence, it is highly sensitive to electronic noise. Finally, Magnetic Flux Leakage (MFL) is the technique in which powerful permanent magnets create a longitudinally moving flux field around the full circumference and throughout the pipe wall thickness. A metal loss defect tends to disturb the magnetic field causing the flux to leak from the pipe wall. The flux leakage can be detected by means of magnetic field sensors. In real applications with pipelines of hundreds of kilometers, these devices gather huge volumes of data. A typical ILI tool can generate until 80GB of raw data from a single run in a 200 km pipeline. Besides, the information provided from this data tends to be imprecise owing to the limited resolution of the technology and therefore an imperfect measurement of defect locations and dimensions. Furthermore, uncertainties associated with pipeline material properties, topography, operational loads, and the rate of corrosion growth results in an uncertain description of the pipeline condition. In order to deal with these uncertainties, postprocessing of collected data by using mathematical and statistical techniques to provide the location and estimation of defects has attracted the attention of researchers over the last several years. An adaptive filter and a wavelet-based denoising technique was used to improve SNR of the signal and therefore, to improve the ability of automated defect characterization to detect defects that would be covered up. The approach was validated using data from field tests on a pipeline gas 350cm length (section without welds). Around 115 signals from MFL sensors were recorded, but unfortunately, no information about the sample rate, speed of the pig or number of data-points is provided. From another point of view, the compression of the MFL signals is an alternative way to handle the problem. A three-stage univariate-multivariate algorithm is proposed with high compression ratios with minimal loss of information. It involves Mean Absolute Deviation, Principal Component Analysis (PCA) as a dimensional reduction tool, and Wavelet Transform (WT) as a univariate compression technique. It is presented another approach that uses pattern-adapted wavelets to detect and estimate the length of the defect and, artificial neural networks to predict their depth. This technique has been successfully (high levels of accuracy and computational efficiency) using signals from MFL simulations. No real signals from the test, which include all uncertainties related to data collection, were involved. Besides, the simulation was conducted in a small section of pipe where the defect covers more than 50% of the specimen. Keeping the same idea, it is proposed several machine learning techniques (support vector machine, kernelized principal component analysis, and kernelized partial least squares PLS) for defect detection and sizing using real MFL images recovered from actual pipeline inspections. It means, that despite that MFL data is derived from actual pipeline inspection runs, the entire pipeline has to be previously inspected and the damage localized by using other techniques. Once the portion of the pipe with damage is isolated, MFL data is processed to estimate its size and depth.

All these papers demonstrate that detecting any abnormality in a real pipeline in service by using directly the signals from a complete smart pig running, is not an easy work.

On another hand, the hydrocarbons industry in Colombia started formally at the beginning of the twentieth and currently is one of the principal pillars for the Colombian economy, representing around 5% of its gross domestic product. Since, petroleum reserves have decreased, gas becomes in one main alternative for economical growth. However, current gas pipelines have been in service for over 30 years and some of them are buried and phenomena like metal losses, corrosion, mechanical stress, strikes by excavation machinery and another type of damages are presented. Typically, the maintenance program of these structures is corrective and very expensive. Predictive maintaining is no usual since internal inspections are conducted by foreign companies. In order to overcome this situation, the native research institute “Research Institute of Corrosion - CIC (Corporación para la Investigación de la Corrosión)” developed an in-line inspection (ILI) tool to be operated in Colombian pipelines to get valuable information of their current state along of thousands of kilometers.

This tool, named ITION (Inspection of Trends of Integrity and OperatioN), was developed and run along a gas pipeline, demonstrating a good mechanical and electronic performance. 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 installed in this tool consists of an arrangement of Magnetic Flux Leakage (MFL), tool movement, magnet, and caliper sensors, among others. These variables sensed along a long pipeline contain valuable information, but their millions of samples demand data compression to obtain a new reduced and representative data to be processed.

Regarding the post-processing ILI data, authors have shown in several works that Principal Component Analysis (PCA) is an effective tool to detect and locate abnormal operational structural conditions from multidimensional big data. In a previous analysis, multidimensional data from a complete ITION running was used to locate possible damages along the pipeline, however, most of the activated points belonged to weld points. All the above motivated the present paper in which the main contribution can be focused on two lines: The development of the ILI tool which has been adapted to necessities to the Colombian pipeline network and, the achievement of a modified monitoring methodology focused on real applications: long pipelines with many years in service. This methodology is based on: Firstly to detect welds and update the owner weld chart in order to organize recorded data by sections (pipeline portion between to consecutive welds) and secondly to apply multivariable statistical analysis to each section in order to highlight sections whose recorded variables behave quite different than mean, based on different statistical indices obtained by applying Principal Component Analysis. This involves the correlation of all variables (sensors) even whether they are measuring different things.
The smart pig (ITION) was run in a real 110Km gas pipeline placed in the Colombian territory. This pipeline has been in service for more than 15 years and presents all kind of complex situations that make difficult its diagnosis such as: changes of the topography; some portions are buried, others are not; despite the original weld chart is available, it is not updated and therefore it does not match with real locations; it is formed by tube sections of different lengths; there are many curved sections, and so on.

This paper is organized as follows: Section 1 introduces the studied problem and solution exposed, in section 2 it is detailed the developed smart in-line inspection tool denominated ITION. For the sake of completeness in presentation, the formulations of multiway principal component analysis technique (MPCA), PCA, and Damage indices (DI) are discussed in Section 3. The proposed methodology is described in section 4. Section 5 presents the experimental results and the discussion of their main contributions. Finally, section 5 concludes the main research results.

ITION- SMART IN-LINE INSPECTION TOOL

ITION Configuration

A smart pipeline inspection tool is an instrumented vehicle capable of traveling inside a pipeline by means of the transported fluid propulsion, besides of acquiring and recording along a pipeline different physical variables according to the inspection purpose (measuring operational conditions, locating or inspecting damages, defining geometrical profiles) by means of different sensors inside and outside its chassis.

The Corrosion Research Institute (CIC) from Colombia developed the Smart Inspection Tool denominated Inspection of Trends of Integrity and OperationN- ITION (see Figures 1 and 2). Its basic instrumentation includes: odometer, Inertial Measurement Unit (IMU), accelerometers, calipers, pressure, and temperature sensors. Additionally, a special array of linear MFL’s transducers that varies its output voltage in response to the perturbed magnetic field applied by permanent magnets was attached to analyze the wall health of the studied gas pipeline.

This tool was run along a real and operated Colombian gas pipeline with the main purpose of establishing the current pipeline structural health condition and determining potential abnormal conditions such as weld failure, geometric deformation, corrosion, mass loss or adding, among others. Figures 3 and 4 show specials launching and reception stations constructed to run the ITION tool.
**Inspected Gas Pipeline**

The inspected gas pipeline is owned and operated by one of the biggest Colombian gas transport companies. It goes through several Colombian political departments and kinds of topographies, including horizontal and inclined terrains, straight and curved paths. It is 110Km long, built with steel carbon seamless pipe sections of 12m long and 30cm diameter. It has been in service for more than 15 years.

**Recorded Physical Variables**

This was the first time that the ITION tool was run in a pipeline in service. Although it traveled all 110Km of the pipeline, we only select a portion of them for this study, specifically the first 23.9Km. The selected set of data contains more than 10 millions of data-points by variable. On the other hand, the tool has 24 arms equipped with many sensors each one, for this study, the following variables were available: movement, pressure, temperature, calibration distances, magnetic field, and flow leakage. Although these are physical variables, the available sensors have not yet been calibrated, therefore, the values measured and represented in this work do not contain the real units. Besides, ITION included an odometer to measure the traveled distance. In total, 18 signals were recorded and labeled as follows:

- **Inertial Measurements**: By using an Inertial Movement Unit (IMU), three inertial signals were recorded and their profile along the portion of the gas pipeline is presented in Figure 5.

- **Remanent Magnetic Field (RMF)**: By magnetizing the pipeline wall with a permanent magnet unit installed in the ITION tool and by using a pair of magnetic sensors, the remanent magnetic field in the pipeline wall is sensed and recorded. Figure 6 presents the obtained profile.

- **Temperature, Pressure and Vibration**: These three signals are measured and recorded to know the fluid condition and tool vibration. Figure 7 presents the obtained profile for each one of them.

- **Calibration Signals**: By using two calipers located in the opposite way outside of the chassis, calibration signals were measured and recorded. Figure 8 presents a sample of the obtained profile.

- **MFL Signals**: MFL measurements are one of the most important variables measured by ITION for the studied gas pipeline, and it is the most common In-Line Inspection (ILI) technique used for monitoring wall thickness in long carbon steel pipelines to detect defects such as mass loss or adding, fitting, non well-conditioned welds or those associated with the presence of corrosion or any other phenomena. 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 using permanent magnets. Since the magnetic field is perturbed by material defects a flux leakage outside the pipe is produced and measured by field sensors. Figure 9 presents the profile of the 8 MFLS signals measured by a ring array uniformly distributed.

Typical MFL analysis considers that signal is acquired at quasi-static velocity conditions, however, most of the In-Line Inspection (ILI) of pipelines with MFL’s are highly affected by high velocities of the tool, thus a regulator is included to maintain a velocity under 2m/s. For the specific Colombian topographical conditions, where high downs are present, this regulator is not ever effective and greater velocities can be experienced during the run. This is the case for some parts of the gas pipeline trajectory traveled by ITION, where velocities up to 30 m/s for high downs were experienced (see Figure 10), which is not favorable for the MFLs technique performance.

**Signal profile of abnormal situations**

In previous lab tests, the company owner of the pipeline identified the typical profile of some abnormal situations and they are shown in Figures 11 to 16. Due to confidentiality agreement, it is not specified the abnormal situation type, just the following labels: MA, CC, FC, MM, TA, V.

From these figures, some particular behaviours can be highlighted. For instance, in all cases, RMF signals present changes in magnitude. Besides, MA class seems to produce small variations in MFL signals and opposite trends in calipers. CC, FC and MM do not show changes in MFL neither Inertial. It can be seen also fluctuations in MFL and Calipers but it does not in RMF signals in the case of PM. Finally, when the device seems to be moving perpendicular to the displacement, all signals show fluctuations. It is also clear that these abnormal situations do not belong to damages in the wall of the pipeline since MFL profiles are not the typical when a leakage of magnetic flux is produced.

**MULTIWAY PRINCIPAL COMPONENT ANALYSIS (MPCA)**

MPCA is a straightforward extension of conventional PCA to handle data in multi-dimensional arrays. A typical 2D data matrix can be considered as a two-way array, with experiments and variables (or discretization instant times) forming the two different ways. In some applications, it is necessary to extend this scheme to multiway arrays, e.g. in the case of different experimental trials, where several sensors are measuring at different time instants. MPCA is equivalent to performing ordinary PCA to an unfolded version of the original multiway array. For the sake of presentation, a brief explanation is given here, however the reader can be referred to [14] and [16] for details.

**Organization of the recorded data**

Let us address the analysis by measuring just one sensor (e.g. inertial 1) at a given number of samples. This measurement is repeated several times (experimental trials), considering that each measurement is an individual experiment in the data set. The collected data are arranged as follows:

\[
X = \begin{pmatrix} x_1^T \\ \vdots \\ x_T^T \\ \vdots \\ x_T^T \\ \vdots \\ x_T^T \end{pmatrix} = \begin{pmatrix} x_{i1} & x_{i2} & \cdots & x_{ik} & \cdots & x_{iK} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{j1} & x_{j2} & \cdots & x_{jk} & \cdots & x_{jK} \end{pmatrix} \quad (1)
\]
Figure 5. Inertial measurements: describe the movement of the ITION tool.

Figure 6. Remanent magnet field measured in the gas pipeline wall.
Figure 7. Registered temperature and pressure of the gas, and vibration.

Figure 8. Profile of Calibration Signals
Figure 9. Profile of MFL signals
Figure 10. Profile of ITION velocity

Figure 11. Profile of signals with situation class MA
Figure 12. Profile of signals with situation class CC
Figure 13. Profile of signals with situation class FC
Figure 14. Profile of signals with situation class MM
Figure 15. Profile of signals with situation class PM
Figure 16. Profile of signals with situation class V
This matrix $X \in M_{I \times K}(\mathbb{R})$ —where $M_{I \times K}(\mathbb{R})$ is the vector space of $I \times K$ matrices over $\mathbb{R}$— contains information from $K \in \mathbb{N}$ samples and $I \in \mathbb{N}$ experimental trials. Each row vector $(x_i^T)$ represents measurements from the sensor at a specific $i$th trial. Considering that a multisensor system is used, $J \in \mathbb{N}$ sensors are continuously gathering data by each trial experiment. The collected data are arranged as follows:

$$
X = \begin{bmatrix}
  x_{i1} & x_{i2} & \ldots & x_{iK} \\
  x_{i1} & x_{i2} & \ldots & x_{iK} \\
  \vdots & \vdots & \ddots & \vdots \\
  x_{i1} & x_{i2} & \ldots & x_{iK} \\
\end{bmatrix}
$$

This $X \in M_{I \times K \times J}(\mathbb{R})$ and each row vector $(x_i^T)$ represents the measurements from the $j$th sensor at a specific $i$th trial.

**Unfolding**

As described in 14, to apply PCA to a three-way data matrix $X$ as in equation (2), it has to be unfolded into a two-dimensional data matrix $X^{15,16}$. According to 17, there are six possible ways of unfolding this three-way data $X$ matrix.

In this work, as it has typically been used in SHM, unfolding type $E$— as in equation (3)— has been used since that the whole set of experiments have the same length and, besides, we want to preserve the information of the experiments to study the correlation between sensors and time-instants.

**Principal component analysis (PCA)**

The matrix obtained after unfolding the matrix in equation (2) as the one in equation (3), $X \in M_{I \times J \times K}(\mathbb{R})$, contains information from $J$ sensors at $K$ samples (time instants) and $I$ experimental trials. Consequently, each row vector represents a specific experimental trial, the measurements from all the sensors at a particular time instant.

The main objective of the principal component analysis is to distinguish which dynamics are more relevant in the system, which are redundant and which can be considered as a noise 10. This objective is essentially accomplished by defining a new coordinate space to re-express the original one, by maximizing the variance and minimizing the correlation between variables in the new space. In other words, the objective is to find a linear transformation orthogonal matrix $P \in M_{J \times J \times K}(\mathbb{R})$ that will be used to transform the original data matrix $X$ into the form:

$$
T = XP \in M_{I \times J \times K}(\mathbb{R}).
$$

$P$ is usually called the principal components of the data set or loading matrix and matrix $T$ is the transformed or projected matrix onto the principal component space, also called score matrix. The columns of the principal components (columns of $P$) are the eigenvectors of the covariance matrix $C_X$ organized according to its associated eigenvalue in descending order. In this way,

$$
C_X P = P \Lambda,
$$

where

$$
C_X = \frac{1}{J \cdot K - 1} X^T X \in M_{J \times J \times K}(\mathbb{R}),
$$

and the diagonal terms of matrix $\Lambda$ are the eigenvalues $\lambda_i$, $i = 1, 2, \ldots, J \cdot K$. Using all the $J \cdot K$ principal components, that is, in the full dimensional case, the orthogonality of $P$ implies $PP^T = I$. Therefore, the projection can be inverted to recover the original data as $X = TP^T$. However, PCA also seeks to reduce the dimensionality of the data set $X$ by choosing only a reduced number, $g < J \cdot K$, of principal components, that is, only the eigenvectors related to the $g$ highest eigenvalues. In this way, given the reduced matrix $\hat{P} \in M_{J \times K \times g}(\mathbb{R})$, the score matrix is defined as:

$$
\hat{T} = X\hat{P} \in M_{I \times K \times g}(\mathbb{R}).
$$

Now, it is not possible to fully recover $X$, although $\hat{T}$ can be projected back onto the original $J \cdot K$-dimensional space to obtain a reconstructed data matrix as follows:

$$
\tilde{X} = \hat{T}\hat{P}^T \in M_{I \times J \times K}(\mathbb{R}).
$$

The difference between the original and the reconstructed data matrices ($X$ and $\tilde{X}$), which describes the variability not represented in the projections, is defined as the residual error matrix $E$ as follows:

$$
E = X - \tilde{X} = X - \hat{T}\hat{P}^T = X - X\hat{P}\hat{P}^T,
$$

For the sake of simplicity, the caret is removed from the reduced matrices ($\hat{T}$ and $\hat{P}$) in the rest of the paper. If equations (7) and (10) are analyzed in terms of experimental trials, we obtain:

$$
t_i^T = x_i^T P \in \mathbb{R}^g, \; i = 1, \ldots, I,
$$

and

$$
e_i^T = x_i^T (I - PP^T) \in \mathbb{R}^{J \times K}, \; i = 1, \ldots, I,
$$

where, $x_i^T \in \mathbb{R}^{J \times K}$ denotes the vector of the $i$th experimental trial—that is, the $i$th row of the original matrix $X$; $t_i^T$ is the vector of the projection of $x_i$ onto the first $g$ principal components—that is, the $i$th row of the score matrix $T$; and, finally, $e_i^T$ is the residual error of the $i$th experimental trial— $i$th row of matrix $E$.

In the literature, several methods can be found to calculate the principal components (PCs)—matrix $P$ in equation (4)— and the projections—matrix $T$ in equation (4)–. Some of them focus on the computation of the whole set of principal components, for instance, using the singular value decomposition (SVD) of the data matrix 18. Other strategies try to compute uniquely a reduced number of PCs, such as, for instance, the power method described initially by Hotelling 19 or the QL algorithm presented by Wilkinson 20.

An alternative way of identifying the required PCs is to use the nonlinear iterative partial least square algorithm (NIPALS) 21, which is more efficient and accurate for large matrices, but slower than SVD. On the other hand, artificial neural networks (ANN) provide a way of extending PCA, including nonlinear generalizations 22.
Damage index: $Q$ and $T^2$-statistic

PCA can be used to detect abnormal behavior in a process or system. Two well-known statistics are commonly used to this aim: the $Q$-statistic and the Hotellings $T^2$-statistic. The first one is based on analyzing the residual error matrix $E$ to represent the variability of the data projection in the residual subspace. The second index is based on analyzing the score matrix $T$ to check the variability of the projected data in the new space of the principal components. These indices are based on the assumption that the underlying process follows approximately a multivariate normal distribution where the first moment vector is zero.

$Q$-statistic denotes the change of the events which are not explained by the model spanned by the principal components. In other words, it is a measure of the difference, or residual, between a sample and its projection onto the model. The $Q$-statistic of $i$-th experimental trial is defined as the sum of the squared residuals of each variable as follows:

$$
Q_i = ||e_i||^2 = e_i^T e_i = \sum_{\ell=1}^{K} e_{i,\ell}^2
$$

$$
= x_i^T (I - PP^T) (I - PP^T)^T x_i
$$

$$
= x_i^T (I - PP^T) x_i
$$

where $e_{i,\ell} \in \mathbb{R}$ denotes the $\ell$th element of the vector $e_i$, $\ell = 1, \ldots, J \cdot K$. $T^2$-statistic denotes the inner change of principal component model. $T^2$-statistic of the $i$th experimental trial is defined by the averaged sum of its projection into the new space as follows:

$$
T^2_i = \sum_{r=1}^{q} \frac{c_{r,i}}{\lambda_r} = t_i^T \Lambda^{-1} t_i
$$

$$
= x_i^T (P \Lambda^{-1} P^T) x_i \in \mathbb{R},
$$

where $c_{r,i} \in \mathbb{R}$ denotes the $r$th element of the vector $t_i$ –the projection onto the $r$th principal component or $r$th score of the $i$ experimental trial--; and $\lambda_r \in \mathbb{R}$ is the $r$th eigenvalue –see equation (5)–. $T^2$-statistic only detects variations in the plane of the first $p$ principal components which are greater than what can be explained by the common-cause variations. In other words, the $T^2$-statistic is a measure of the variation in each sample within the PCA model.

Normally, $Q$–statistic is much more sensitive than $T^2$-statistic. This is because $Q$ is very small and therefore any minor change in the system characteristic will be observable. $T^2$ has great variance and therefore requires a great change in the system characteristic to be detectable. The concept of principal components is depicted in Figure 17. A three-dimensional data set is shown, where the data lie primarily in a plane, thus the data is well described by two PC's. The first PC aligns with the greatest variation in the data while the second PC aligns with the greatest amount of variation that is orthogonal to the first PC. Besides, examples damage indices for trials with unusual variations inside ($T^2$-statistic) and outside of the model ($Q$-statistic) are shown.

![Figure 17. PCA model of three dimensional data set showing $T^2$-statistic and $Q$-statistic outliers](image)

**MONITORING METHODOLOGY**

As previously mentioned, for the inspection of the first portion of 23.9 km of a Colombian gas pipeline under study, more than 10 millions of samples were retrieved, this demands a high capacity storage system and an adequate data conditioning and analysis to avoid false alarms (associated for example to changes in the signals due to different situations that do not belong to material defects). This means that a simple eye inspection of the signals does not provide an effective diagnosis of the state of the pipeline.

An alternative technique, used by the authors in previous works, to analyze this kind of multidimensional and big data is the multivariable statistical analysis by means of principal component analysis (PCA). In $^{12 13 25}$ it was reported the use of PCA to detect abnormal conditions of the same gas pipeline studied in this paper, where the main contribution consisted on using multivariable statistical analysis (instead of univariable) on the whole set of acquired signals in a shorter gas pipeline section. The main conclusion of the cited works is that by using statistical indices, abnormal conditions are detected, but most of them are masked by false alarms activated by the presence of repetitive MFL’s high variations which are associated to the presence of normal welds. On another hand in $^{25}$, an artificial neural network is used to automatically detect welds of this pipeline by using acceleration, vibration, and magnetic measurements.

Thus, this paper is dedicated to present a modified monitoring methodology focused on: Firstly to detect welds and update the owner weld chart to organize recorded data by sections (pipeline portion between two consecutive welds, removing the welds) and secondly, to apply multivariable statistical analysis to each section of the pipeline to detect those sections whose recorded variables behave quite differently than the others. The general methodology is depicted in Figure 18 and described in the next sections.
Selection of variables

Weld detection and pipeline section definition

For $s = 1$ to $S$ (each section)

For $j = 1$ to $J$ (each variable)

Considering all measurements as Equation (2)

Unfolding as Equation (3)

PCA & $Q$-indices (Equations 4 & 13)

$MQ = \text{Maximum } Q$

If $MQ \geq (\mu + 1.96*\sigma)$

95% CI of all Maximum indices

ALARM

Analyze signals in section

Detection of potentially abnormal sections

Figure 18. General methodology
Selection of Variables

Eighteen variables were measured in the inspected gas pipeline by using the ITION tool (inertial (3), remanent magnetic field (2), temperature (1), vibration (1), pressure (1), calipers (2) and MFL (8)). MFL is the variable directly related to material changes, however, any abnormal movement of the tool that is not associated to damage can be reflected on the MFL measurements. Therefore, variables of ITION movement: inertial and caliper, and even the remanent magnetic field variable which are not directly related to damages, can be useful to avoid false alarms. On the contrary, variables related to the transported fluid: pressure and temperature, and vibration of the tool are not related with damages or changes in the MFL signal, for this reason, they are not considered in this study.

Weld detection and pipeline sections definition

In this study, a pipeline section is defined as a pipe portion between two consecutive welds, thus a section can have different geometric forms and sizes. Each pipeline section can be defined by using the welds chart given by the gas pipeline owner, but unfortunately, this chart must be updated.

To update the weld chart of the pipeline owner, measurements from magnetic sensors (MFL signals) and odometer are used. Since the odometer is an instrument that can slip, it is presumed that its measurement contains an error and the reported weld positions do not necessarily match with the distance measured by the odometer.

The detection of each weld point is conducted by computing the sum of the windowed overlapping RMS (Root Mean Square) of all MFL signals using a window length of 5. As previously mentioned, raw signals contain a lot of outliers and noise, therefore they should be cleansed, the windowed RMS filter is a good option. It is assumed that the highest value occurs in the weld points. In this way, by comparing the location of this value with the reported weld location, the odometer error is estimated and the weld chart is updated.

Since welds generate changes on MFL measurements producing false alarms, they should not be considered, therefore the measurements of 50cm around each weld are removed. It is clear that the weld size length is only a few centimeters, however, due to magnetic properties, its effect on MFL measurements is reflected in a bigger region.

Finally, the section $s$ or $s^{th}$-section is defined as the pipeline portion between the $(s-1)^{th}$ and the $s^{th}$ weld. Data gathered from the $s^{th}$-section is stored in the matrix $X_s \in \mathcal{M}_{J \times n_s(\mathbb{R})}$ which contains information from $J = 15$ measured signals (inertial 1, inertial 2, inertial 3, remanent magnetic field 1, remanent magnetic field 2, caliper 1, caliper 2, and MFL 1 to 8) and $n_s \in \mathbb{N}$ data-points by signal.

Data organization by each pipeline section

As remarked in the introduction, the proposed methodology is focused on finding the pipeline sections with damages, therefore pipeline sections are independently analyzed. Once the $s^{th}$-section is selected, its gathered data ($X_s$) need to be reorganized to apply MPCA as follows. For the sake of simplicity, the subscript $s$ is removed in the formulation.

Let us denote $x^j$ as the $j^{th}$ measured signal of the analyzed section, it means the $j^{th}$ row of $X$. Each $x^j$ is split up into $I$ observations of $K = n/I$ data-points (time instants) each. Each observation is located below the previous one. In this way, the measurements of each signal in a given section are organized as Equation 1.

To consider all signals in the MPCA, they are arranged in a 3D matrix ($\mathbf{X}$) as Equation 2. Therefore, each row vector $(x^j)^T$ represents the measurements from the $j^{th}$ variable at a specific $j^{th}$ observation. An observation can be considered as measurements from a small area of the wall pipeline, its size is variable depending on each section.

Finally, $\mathbf{X}$ is unfolded into a two-dimensional matrix as shown in Equation 3. The resulting matrix $\mathbf{X} \in \mathcal{M}_{I \times J \times K(\mathbb{R})}$, contains information from $J$ variables at $K$ time instants and $I$ experimental trials.

It is recommended that the number of observations $I$ be greater than the number of columns $K \cdot J$, therefore:

$$I = \frac{n}{K} > K \cdot J \quad \text{then} \quad K < \sqrt[3]{\frac{n}{J}}. \quad (14)$$

In Equation 3, $K$ is rounded to the nearest integer toward minus infinity of $\sqrt[3]{\frac{n}{J}}$, to guarantee the maximum number of columns.

PCA and damage indices by each section

Once all data from the given section are organized as described above, a MPCA is conducted, this means that the principal components or loading matrix $\mathbf{P}$ and the transformed or projected matrix $\mathbf{T}$ in Equation 7 are calculated using 5 principal components ($\varrho = 5$). For each observation (small area of the pipeline section) $Q$ and $T^2$-statistic are calculated according to Equations 13 and 14. The maximum values of them are retained to be compared with the maximum of the rest of the sections.

Detection of potentially abnormal sections

Previously it was stated that the goal of the multivariable statistical analysis applied in this study is to detect sections where a probable abnormal condition can be present (measurement errors, potential mass loss, or another type of defects). To achieve this, all the maximum retained indices (one by section) should be compared between them. Therefore, it is calculated a 95% confidence interval of the mean for the indices. The sections with the statistical indices outside of this interval are activated as potential abnormal sections.

Experimental Validation

The main purpose of this study is to exploit millions of samples recorded by several sensors along/inside a real Colombian gas pipeline and detect all these sections that are potential candidates to present some abnormal situation. Thus, the above methodology was proposed specifically for the first run achieved by the ITION tool under the particular conditions previously exposed. Specifically, a data portion corresponding to the record of the first 23,9Km of the Colombian gas pipeline was used to validate the proposed methodology. This data portion contains 10.139.436 samples (320 MB) and the main results obtained are here presented.
According to the highest value of the windowed overlapping RMS of all MFL signals, 2071 welds (2070 sections) were identified and compared with the welds chart reported by the owner of the pipeline. As an example, Figure 19 depicts a section of the MFL signal recorded by sensor 1 (MFL 1), the corresponding updated weld points (red diamonds labels), and the weld chart points (black triangle) in the portion between 7.3 and 7.4Km of the pipeline.

Since the speed running of the ITION tool is not constant due to the topological configuration of the pipeline and, the flow of the transported gas (e.g. the tool is faster going up a sloping hill than going down), measurements are sampled uniformly in time, but it does not in distance. Therefore, we found some pipeline sections whose signals have more than $10^5$ samples and others whose signals with less than $10^2$ samples. On the other hand, data contains a lot of outliers that cause huge values of $Q$ and $T^2$—statistics and therefore, false alarms. In conclusion, oversampling and outliers produces false alarms. Considering this particular situation, to detect the potentially abnormal sections, the methodology was validated by using resampled data (same distance between observations).

Since outliers and oversampling are associated to no well-conditioned signals, signals of each section were resampled by maintaining a constant sample distance and by using as equivalent time vector the odometer record. Then, a sample step of 1 cm was defined and a mean of 1200 samples by sensor in a 12 m typical section is obtained.

$Q$ and $T^2$—statistics were obtained for each section and those with the highest statistical indices were activated. It was observed that high values of statistical indices are associated in most of the cases to temporal changes of MFL signals or continuous movements of the ITION tool (see Figure 20). Sections with unusual changes of MFL signals are candidates to be analyzed by an expert or to take into account in a new run for contrasting purposes.

As seen in the figure, the sections with the highest damage indices are: 1707, 238, 1529, 1345, 1246, 1845, 1916. The signal profiles of this sections are shown in Figures 21 to 28. From these profiles it is clear that sections where MFL signals colorblue show changes, are easily detected. As it was expected, MFL technology are able to measure those places where leakage of magnetic flux is produced due to damages in the pipeline wall.

Some physical limitations of the ITION device do not allow infer directly which type of wall damage exists. For example, MFL sensors are mounted on arms with springs such that they are most of the time very close to the wall pipe, but magnetos are joint to the device chassis all the time, therefore the magnetization is not the same in the whole pipeline. As consequence, MFL signals are not directly correlated with magnetization signals.

Checking those sections with a high index, out of the confidence interval, it can be found some of them with similar features or profiles that the shown ones in Figures 11 to 16. For instance, profiles from section 1078 are similar to the ones from abnormal situations labeled as CC, section 1609 to MA label, and 1513 to MM label. Figures 29 to 31 shown the mentioned profiles.

Discussion and Conclusion

In general, the decision making regarding the pipeline maintenance and future inspections are complicated due to the uncertainties involved with the condition assessment. Nevertheless, ILI tools equipped with magnetic sensors are considered as a good option to evaluate the state of pipelines, however, a good interpretation of the magnetic effect is necessary to improve detection of a failure in the analyzed ferromagnetic structure and consequently to avoid more potential accidents that could occur with negative consequences (e.g. the collapse of a gas pipeline).

This paper presents a statistical analysis, based on Multiway Principal Components Analysis (MPCA), on a large number of samples that are associated with different physical variables recorded by the ITION tool along a real Colombian gas pipeline. First, an easy and adequate methodology was used to update the owner’s weld chart by taking advantage of high values of MFL at weld points and based on distances recorded by the odometer. Also, it was demonstrated the utility of the PCA analysis to obtain indices in a reduced space (principal components), where significant changes of these indices correspond to MFL changes that contain useful information for experts and pipeline owner.

On the other hand, if the position of permanent magnets varies significantly (associated with vehicle movement), MFL changes could be wrongly associated with wall defects. Thus, for future versions of ITION tool it is recommendable to keep permanent magnets as near as be possible to the pipe wall or to record the magnet movements and updating the MFL value at each sample according to the hysteresis curve and based on the estimation of the applied magnetic field. Thus, statistical indices sensitivity can be improved with additional conditioning of devices installed in the ITION tool such as permanent magnets. It is also necessary a deep analysis of all variables to identify which variables are more influential to the different abnormal situations or phenomena existing along the pipeline (feature selection).

Usually, ILI techniques are successfully in pipelines where their geometric layout is uniform. However, in the case of pipelines located on non-uniform land surface relieve (e.g. the Andeas Region, Colombia), with many years in service, where some portions are buried, others are located on the terrain surface and others are elevated, the measurements are not reliable. Therefore, is not easy to find research works that address this particular situation, so, unfortunately, our results cannot be compared with other ILI tools/MLF data processing approaches.

The methodology proposed here is successfully on detecting those pipeline sections (the portion between two welds) with an abnormal situation inside the pipeline. As future work, the authors are intended to apply machine learning methods to classify each detected abnormal situation, and even to quantify its severity. To do that, the approach should compare the obtained signals profile of each candidate pipeline section with the profiles of the known situations (MA, CC, FC, MM, TA, V). Unfortunately, to validate and test the complete approach, each case should be contrasted with the real situation, and this information is not available since the pipeline is not easily accessible and verifiable.
Figure 19. A comparison of updated vs weld chart points at five sections of the pipeline

Figure 20. Maximum $Q$ and $T^2$—statistics by section
Figure 21. Profile of signals from section 1707
Figure 22. Profile of signals from section 238
Figure 23. Profile of signals from section 1529
Figure 24. Profile of signals from section 1261
Figure 25. Profile of signals from section 1345
Figure 26. Profile of signals from section 1246
Figure 27. Profile of signals from section 1845
Figure 28. Profile of signals from section 1916
Figure 29. Profile of signals from section 1078
Figure 30. Profile of signals from section 1609
Figure 31. Profile of signals from section 1513
Acknowledgment

The author would like to thank greatly the support of the entities involved in this research: Universitat Politècnica de Catalunya (UPC), Universidad Industrial de Santander (UIS) and, specially to Eng. Mario Quintero from Corporació para la Investigació de la Corrosió (CIC), who provided the whole dataset and all information about the ITION-smart pig.

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References