Journal of Physics: Conference Series **628** (2015) 012128 doi:10.1088/1742-6596/628/1/012128

# Pipe Leaks Classification by Using a Data-driven Approach Based on Features from Cross-Correlated Piezo-vibration Signals.

**Jhonatan Camacho-Navarro[1-2], Magda Ruiz[1], Oscar Perez[2], Rodolfo Villamizar[2], Luis Mujica[1].**

[1]Department of Applied Mathematics III, Escola Universitària d'Enginyeria Tècnica Industrial de Barcelona (EUETIB), Universitat Politécnica de Catalunya (UPC) BARCELONATECH, Comte d'Urgell, 187, E-08036, Barcelona, Spain.

[2]Escuela de Ingenierías Eléctrica, Electrónica y de Telecomunicaciones (E3T), Universidad Industrial de Santander (UIS). Grupo de Control Electrónica Modelado y Simulación (CEMOS). Santander, Colombia.

jhonatan.camacho@estudiant.upc.edu, magda.ruiz@upc.edu, oscar.perez.gamboa@gmail.com, rovillam@uis.edu.co, luis.eduardo.mujica@upc.edu

**Abstract.** This work presents a data driven approach for pipe leaks classification, validated on a steel carbon pipe section conditioned with leaks of different sizes and locations in order to emulate abnormal conditions. The tested structure was instrumented with piezoelectric devices attached at different locations over the surface, in order to induce guided waves and to record its behaviour along the structure. For each experiment, one piezo device is excited by means of a high frequency burst type signal and the other ones are used as sensors. A blind bridle is connected to one of the extremes and an air source is coupled to the other extreme to emulate operational conditions. Statistical indices of correlated piezoelectric signals are obtained by using principal component analysis to distinguish different leak scenarios. Next, a self-organizing map is used to classify them. The experimental results show an improvement of the classification-learning rate when correlated signals are used instead of uncorrelated ones

## 1. Introduction

If pipeline damages could be early detected, adverse environmental and economic effects can be avoided improving human safety. In this sense, pipe leaks is one of the most important type of damage with great economical and environmental impact [1]. In this point, the formulation of methodologies for pipeline leakage identification has been widely documented in the state of the art [2]. Common methods for pipelines damage identification are based on physical models and analysis of measurements obtained from available instrumentation in the structure [3], [4]. However, the big amount of false alarms is one of the drawbacks of these methods, which is a not desirable condition for continuous monitoring systems.

In recent years, the use of guided waves for damage assessment in pipeline structures has been reported as successful [5]. Thus, by taking advantage of piezoelectric properties and guided waves, piezoelectric devices are a cheap technology with promising results for pipeline damage assessment. Experimental results for detecting cuts with different depths in a pipe-like structure using piezoelectric technology are discussed in Gharibnezhad et al. [6]. Other application involves the analysis of piezo-electrical measurements from macro fiber-composite as part of guided wave methods to detect partial circumferential crack and corrosion on the pipe's surface of a pipeline structure located at Los Alamos

National Laboratory [7]. Also, piezo-ceramic transducer has been implemented to identify the position and severity of leaks in the pipeline network at Sheffield University [8]. In addition, the effectiveness of principal component analysis and self-organizing maps for structural damage detection algorithms have been widely demonstrated (for example in pipeline crack detection algorithms [9], [10]).

In this paper, a non-intrusive pipeline leaks detection and classification approach is experimentally validated. Which consists of processing induced guided waves, measured along the structure, by using cross-correlation functions and principal component analysis for detection purposes, and self-organized maps for classification task. The main contribution of this approach is the analysis of the positive influence of cross-correlation analysis as pre-processing tool in a data driven approach used for pipeline leaks detection and classification.

## 2. Structural damage detection methodology

Figure 1 summarizes the procedure for pipelines leaks detection and classification. The approach involves three main modules: 1. Piezo-electric instrumentation; 2. Statistical processing; and 3. Supervised classification.

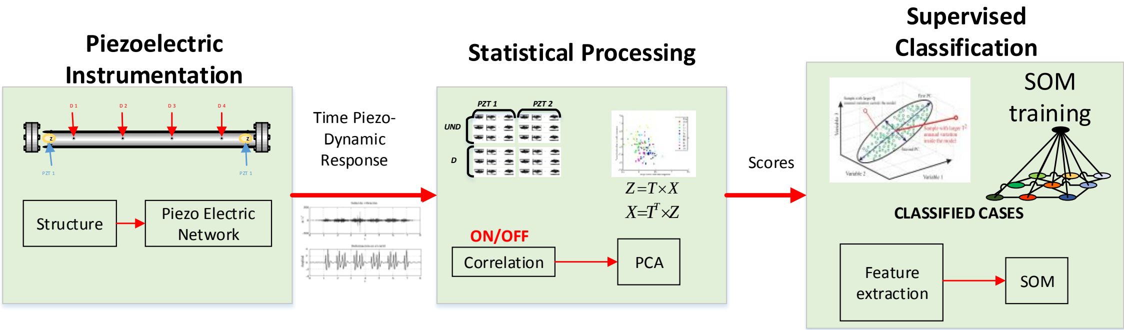


**Figure 1** Damage assessment methodology

### *2.1. Piezoelectric instrumentation*

The approach for damage identification developed in this work is based on piezo-diagnostics principle, which refers to damage identification based on the phenomenon of elastic waves propagation generated by piezo-actuators [11]. For instance, in [12] implements impedance-based structural health monitoring methods by means of piezoelectric materials analysis. Similarly, in [13] takes advantage of lamb waves in order to find patterns with high sensitivity to structural damages.

In this way, a piezoelectric active scheme is used, where several piezo devices are installed in the structure to induce and record a guided wave response at different locations of the structure. One of the piezoelectric devices is excited with a periodic high frequency burst type signal in order to induce a guide wave and the remaining piezo-devices are used as sensors. The instrumentation consists of fine-tuning filters, high wide-band amplifiers and acquisition systems, among others (Figure 2).

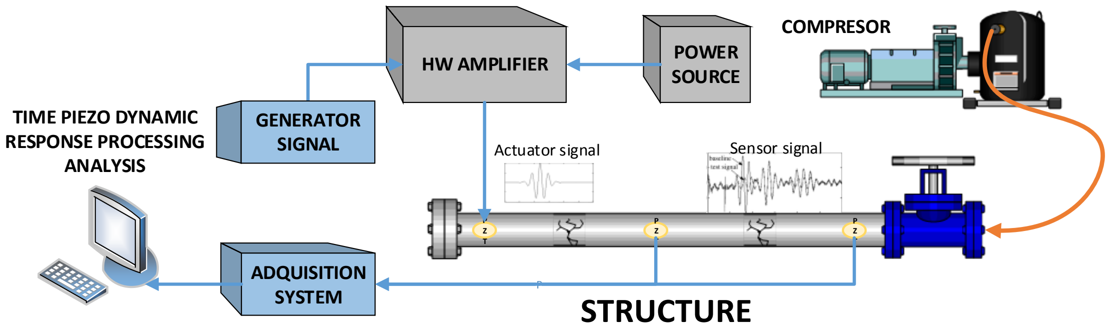


**Figure 2** General components for damage identification based on piezo dynamic response analysis.

### *2.2. Statistical processing*

Three steps are necessary for statistical processing: *i.)* Uncorrelated data matrix building; *ii.)*Cross-correlated data matrix building; and *iii.)* Model building. An important piece of the statistical processing module is the data matrix, which contains information about different types of leaks to be used for validating the classification process. The data matrix is organized as follows: types of damages studied and the number of experiments for each damage; number of piezo device sensors used and finally, number of sample times recorded from each piezo sensor. This matrix has concatenated row vectors representing several experiment repetitions for undamaged and damage conditions (**Figure 3**). The cross-correlated and uncorrelated data matrices are intended to facilitate results comparison regarding to the cross-correlation performance in the piezo-diagnostic approach. Thus, two distinct statistical models are built with both data matrices.

*2.2.1. Uncorrelated data matrix building.*
A first data matrix of recorded signals from piezo sensors are obtained. It contains time measurements as response to the propagation of guided waves along the structure. Thus, time piezo-electrical response for damage and undamaged conditions are folded in the data matrix (**Figure 3**).

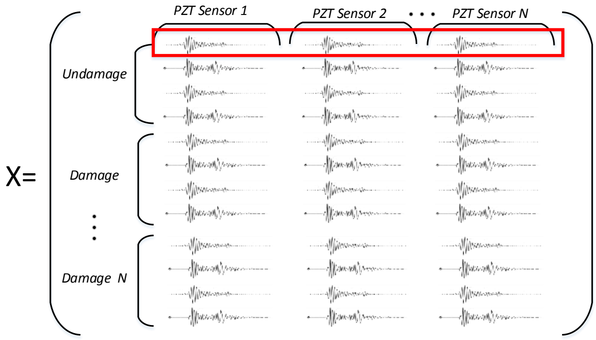


**Figure 3** Data matrix

*2.2.2. Cross-correlated data matrix building.*
A second data matrix is built using cross-correlated functions between piezo actuation and sensing signals, since several applications for structural damage assessment have demonstrated the effectiveness of using cross-correlation functions [14]. For example, in [15] damage identification methods based on natural excitation Technique (NeXT) employs cross-correlation functions for modal analysis. It had been useful for damage identification in civil structures. Other proposal include the estimation of the time of flight of wave packages by means of cross correlation functions determining the location of the defects within a large area of a thin-plate specimen [16].

*2.2.3. Model building.*
The last step is modelling process. This applies Principal Component Analysis (PCA) in the data matrices. PCA is a statistical tool widely used in structural damage detection algorithms, for example in applications related to delamination detection in a composite beam [17]. The objective in PCA is to obtain a reduced space representation for multidimensional data. As is illustrated in **Figure 4**, the original data matrix **X** with ***n*** samples of ***m*** statistical variables is projected onto the orthogonal space, defined by the principal components. The new variables **Z** consists of uncorrelated data organized by down variance, but preserving the cumulative original variance.

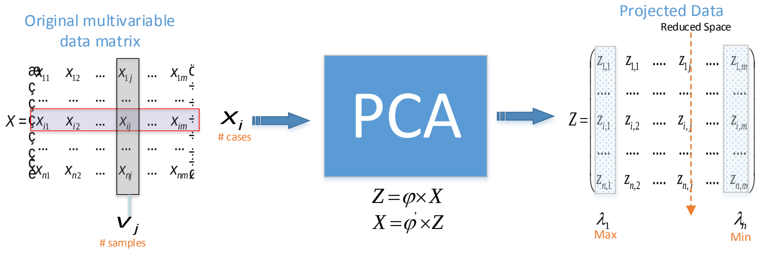


**Figure 4** Principal component analysis scheme

The general procedure used to obtain the reduced space is the next [18]:

**i.** Normalize the original variables by means of standard deviations ($\hat{\sigma}_j$) and mean values ($\hat{\mu}_j$):

$$\hat{X} = \frac{X-\hat{\mu}_j}{\hat{\sigma}_j} \tag{1}$$

Where $\hat{X}$ correspond to normalized data.

**ii.** Compute the covariance matrix of centered data matrix:

$$C_{\hat{X}} = \frac{1}{m-1}\hat{X}'\hat{X} \tag{2}$$

**iii.** Estimate the singular value decomposition for covariance matrix:

$$C_{\hat{X}} = \varphi\Sigma\varphi' \tag{3}$$

**iv.** Transforming the original variables onto the orthogonal space defined by the Eigenvectors ($\boldsymbol{\varphi}$) of the covariance matriz:

$$Z = \varphi \cdot \hat{X} \tag{4}$$

**v.** Keep only the first $\boldsymbol{r}$ components in order to obtain a reduced representation for original variables. The variance for each new variable is equal to its respective $\boldsymbol{r}$ eigenvalues ($\boldsymbol{\Sigma}$):

$$Z_r = \varphi_r \cdot \hat{X} \tag{5}$$

Despite, the above procedure is a common strategy for data reduction, in this approach it is used as statistical model by using the inverse transformation:

$$X = \varphi_r{}' \cdot Z_r + \hat{X} \tag{6}$$

Some considerations should be take into account to build a statistical base-line model of the structure [19]:

*a.) Training stage*

I. Only the undamaged records of the data matrix is processed.

II. The Group-Scaling (GS) normalization method [20] is applied instead of common normalization methods. GS procedure considers the nature of data by estimating standard deviation for each block of piezo measurements. Thus, a standard deviation per piezo sensor is obtained.

III. The $\boldsymbol{r}$ principal components are estimated using an iterative algorithm as NIPALS [21], because the size of the data is big.

IV. The statistical model is obtained only from undamaged records and it consists of four elements: mean values ($\hat{\mu}_j$), standard deviations ($\hat{\sigma}_j$), $\boldsymbol{r}$ eigenvectors ($\varphi_{und,r}$) and eigenvalues ($\boldsymbol{\Sigma_r}$).

*b.) Validation stage*

I. Damage experiments are normalized using the mean values ($\hat{\mu}_j$) and standard deviations ($\hat{\sigma}_j$) from the statistical model according to GS method.

II. Projections onto the reduced space of Undamaged and Damage experiments are computed using the statistical model in order to differentiate damaged from undamaged states.

$$Z_{\text{damage}} = \varphi_{und} * \hat{X}_{\text{Damage}} \tag{7}$$

### *2.3. Supervised Classification*

Damage detection is achieved by means of statistical indexes with high sensitivity to model deviations. In this case, two widely used indexes ($T^2$ $and\ Q - statistic$) are computed. $T^2$ statistic is a variation mesurement of each experiment within the statistical PCA model (undamaged model) and Q-index is the squared 2-norm that measures deviations of the experiment respect to the lower-dimensional PCA representation [19]. Thus, $T^2$ and Q statistics are used to measure deviations of each experiment respect to the PCA model:

$$T^2 = X^T \varphi_{und} \left(\Sigma^T \Sigma\right)^{-1} \left(\varphi_{und,}\right)^T X \tag{8}$$

$$Q_i = \tilde{x}_i{}^T \tilde{x}_i, \quad \tilde{x}_i = [I - \varphi_{und} * (\varphi_{und})^T] X_i \tag{9}$$

where $\tilde{x}_i$ is the residual projection for each experiment.

In addition, the above statistical indexes are used as inputs for a self-organizing map (SOM) for cluster cases in similar damage types. This clustering is achieved by means of competitive learning and preserving topology. Accordingly, nearby data in the input space are mapped into neighbor clusters [22]. Thus, SOM network facilitates classification tasks and graphical interpretation. **Figure 5** deploys how SOM network operates over the input space, specified by $T^2$ and Q-indexes.

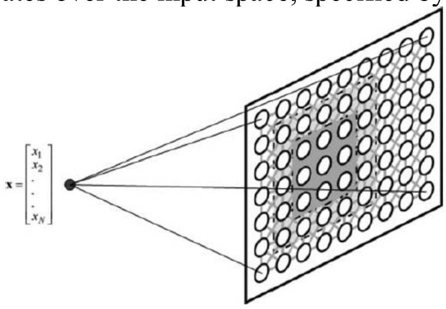

**Figure 5** SOM clustering for damage classification.

From **Figure 5**, the SOM consists of **N** clusters, characterized by a prototype vector (Codebook) or cluster center, and grouping several labelled cases. Then similar cases are labelled in clusters, where each label keeps only one instance and the number of stored cases. Similarly, the validation cases are ticked assigning the label with most instances and with the most similar cluster to find the best matching units (BMU). In consequence, the classification error can be estimated by majority voting.

Finally, the SOM quality is evaluated with the quantization and topographic errors. The first one is the average distance between each experiment and its BMU. The second one corresponds to the proportion between data vectors whose first and second BMUs are not adjacent clusters and the total number of experiments [23].

## 3. Experimental setup

The methodology detailed in this paper was validated for leak detection in a pipeline as part of a non-intrusive damage monitoring system. The specimen used as test structure is a carbon-steel pipe section of dimensions 1 m x 0,0254 m x 0,003 m (length, diameter, thickness). The pipe section has bridles at the ends and a valve sets the air pressure from a compressor in 80 psi at one of the ends (**Figure 6**).

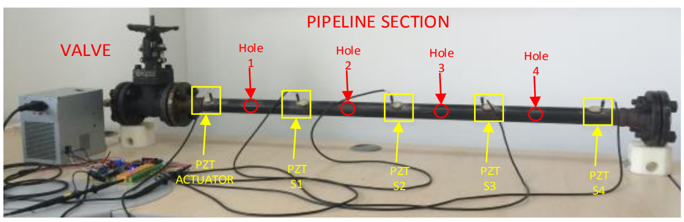


**Figure 6** Experiment configuration

In order to induce leaks in the test structure, four ¼-inch holes were drilled along the pipe section wall. Graduable screws are used to control where the leak is produced. **Table 1** details the damages studied in this work. The proposed damage configuration allows concluding if classification and location of different sizes leaks is possible. For each damage type, 100 experiment repetitions were conducted with 1-second periodic excitation signal, where undamaged experiments are tagged with label 'UN'.

**Table 1** Leak Damage specification

| **Label** | Open Hole | **Label** | Open Hole |
|---|---|---|---|
| **1** | H1 | **5** | H4 ,H3 |
| **2** | H2 | **6** | H4, H3, H2 |
| **3** | H3 | **7** | H4, H3, H2, H1 |
| **4** | H4 | | |

Five piezoelectric devices (PZT) were attached along the structure, (previously explained in **Figure 1**). One of the PZT is used as actuator and the remaining devices as sensors. The PZT actuator was excited by an 80 KHz burst signal generated by an AWG PicoScope series 2000 and then amplified to ±10 V. The piezo electrical response is recorded with a picoscope and multiplexor board. **Figure 7** deploys the time piezo electrical response for each piezo device, where it is possible to visualize time of flights for one of the experiments under undamaged state.

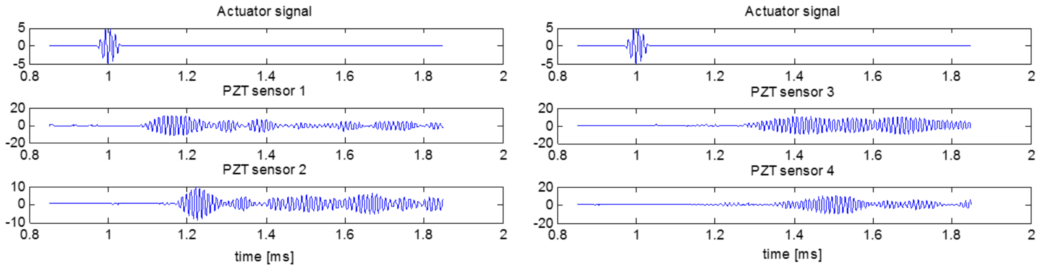


**Figure 7**Time piezo dynamic structural response

The **Figure 8** presents the respective cross-correlation functions.

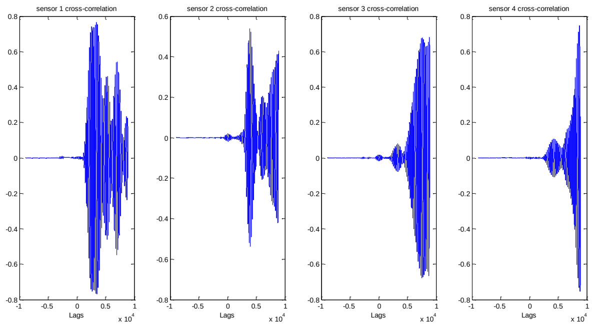

**Figure 8**. Evolution of the Cross-correlation function for each PZT sensor

## 4. Results and Discussion

**Figure 9** presents the statistical indexes for the studied experiments, which were estimated preserving 15 principal components. In addition, the SOM codebooks location and their respective labels (assigned by majority voting) are shown in the right scattered plot. Empty SOM codebooks are necessary to describe the data distribution. It is observed that by the cross-correlation, better groups for studied damages are obtained.

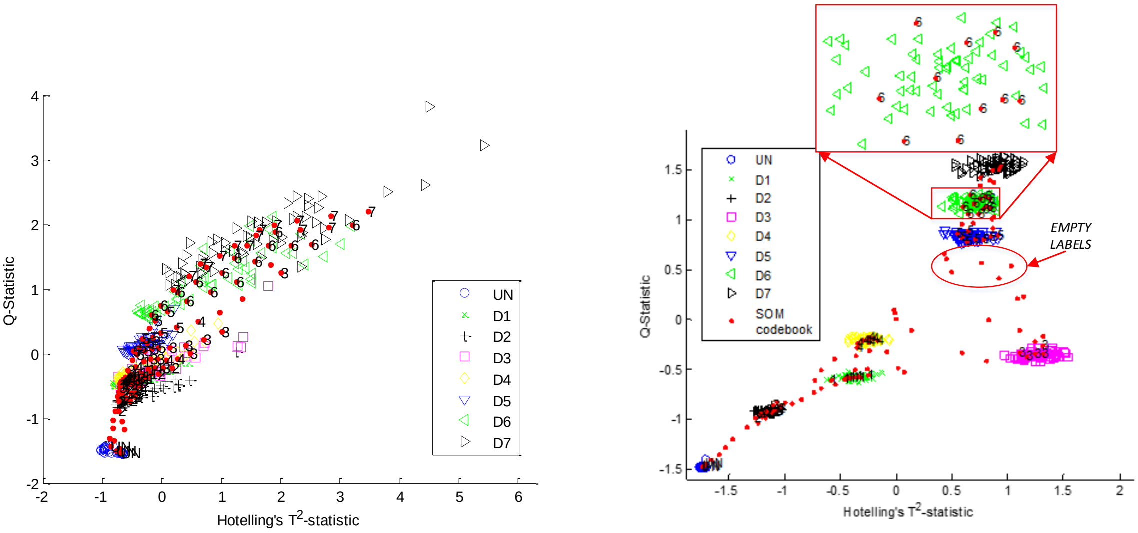


**Figure 9** Distribution of damage indexes. *Left:* Without cross-correlation analysis. *Rigth:* By including cross-correlation processing.

The clusters obtained in the Self Organizing Map are depicted in **Figure 10**. A major differentiation between different damage types with boundaries clearly defined by empty clusters and BMU distance matrix (U-matrix) is observed when the cross-correlation as preprocessing stage is applied. In addition, the cases distribution avoids damages combination in one similar cluster, which allows a better classification.

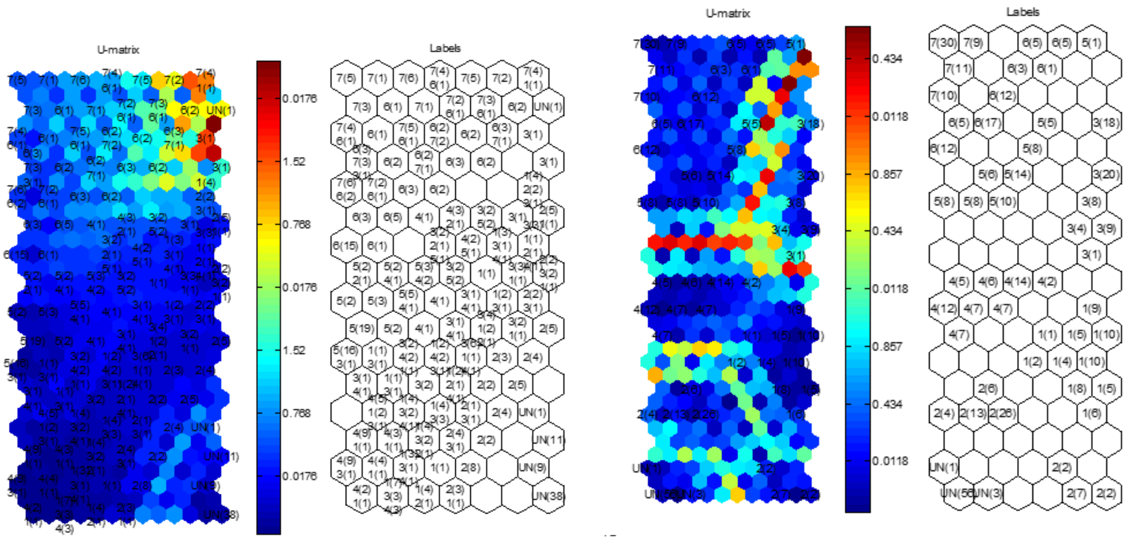


**Figure 10.** Training SOM network. *Left:* Without cross-correlation analysis. *Right:* By including cross-correlation processing.

The SOM quality indexes are summarized in **Table 2,** which are calculated by using 70% of cases for training purposes and 30% of cases for validation.

**Table 2** . SOM quality indexes

| **Index** | **Uncorrelated Signals** | **Cross-correlated signals** |
|---|---|---|
| Quantization error | 0,1083 | 0,0336 |
| Topographical error | 0,0083 | 0,0833 |
| Distortion measure | 1,3828 | 0,4327 |

| | | |
|---|---|---|
| Training Error | 18,5417 | 0 |
| Empty labels in Training data | 0 | 0 |
| Empty Clusters | 14 | 44 |
| Validation Error | 36,5625 | 1,5625 |
| Empty labels in validation data | 17 | 5 |

## 5. Conclusion
In this paper is shown, by means of experimental results, that damage classification is highly dependent of preprocessing stage. Thus, by using correlation of piezoelectric signals a better behavior is obtained. However, for future works normalization strategy could be considered as a preprocessing stage, also influence of correlation and normalization on damage location and quantification should be studied. Further studies are required to evaluate the influence of using cross-correlated features from piezo-electric measurements for damage detection in buried pipes.

## 6. Acknowledgments
This work has been developed as part of the research project: “Monitorización y Detección de Defectos en Estructuras usando Algoritmos Expertos Embebidos”, financed by the Departamento Administrativo de Ciencia y Tecnología Francisco José de Caldas – COLCIENCIAS and Banca Mundial in the Universidad Industrial de Santander, Colombia. Agosto 2012-Junio 2015.

## References
[1] G. S. Mavi. Feasibility of using ultrasonic guided waves for underwater pipe inspection. Thapar University, Patiala, 201.
[2] Billmann, L., & Isermann, R. (1987). Leak detection methods for pipelines. Automatica, 23(3), 381-385.
[3] Black, P. (1992). A review of pipeline leak detection technology. In Pipeline systems (pp. 287-298). Springer Netherlands
[4] Geiger, G. (2006). State-of-the-art in leak detection and localization. Oil Gas European Magazine, 32(4), 193.
[5] Lowe, M. J., Alleyne, D. N., & Cawley, P. (1998). Defect detection in pipes using guided waves. Ultrasonics, 36(1), 147-154.
[6] Gharibnezhad, F., Mujica Delgado, L. E., Rodellar Benedé, J., & Fritzen, C. P. (2012). Damage detection using robust fuzzy principal component analysis.
[7] Thien, A. B. (2006). Pipeline structural health monitoring using macro-fiber composite active sensors (Doctoral dissertation, University of Cincinnati).
[8] Taghvaei M, Beck SBM & Staszewski WJ (2007) Leak detection in pipeline networks using low-profile piezoceramic transducers. STRUCT CONTROL HLTH, 14(8), 1063-1082
[9] Meléndez, J., Colomer, J., López, B., Vehí, J., Pous, C., Ruiz, M. L., & Mújica, L. E. Experiences in exploiting data at the Universitat de Girona: Selected problems in application domains.
[10] Buethe, I., Torres-Arredondo, M. A., Mujica Delgado, L. E., Rodellar Benedé, J., & Fritzen, C. P. (2012). Damage detection in piping systems using pattern recognition techniques.
[11] Kolakowski, P. Damage Identification in Beams via Piezodiagnostics. 2nd European Workshop on Structural Health Monitoring July 7-9, 2004. Amazeum Conference Centre at Deutsches Museum Munich, Germany.
[12] Park, G., Sohn, H., Farrar, C. R., & Inman, D. J. (2003). Overview of piezoelectric impedance-based health monitoring and path forward. Shock and Vibration Digest, 35(6), 451-464.
[13] Ong, W. H., & Chiu, W. K. (2012). Redirection of Lamb waves for structural health monitoring. Smart Materials Research, 2012.

[14] Muyu Zhang, Rüdiger Schmidt, A Comparative Study of the correlation function based structural damage detection methods under sinusoidal excitation.
[15] Caicedo, J. M. (2011). Practical guidelines for the natural excitation technique (NexT) and the eigensystem realization algorithm (ERA) for modal identification using ambient vibration. Experimental Techniques, 35(4), 52-58.
[16] Yu, L., Bao, J., & Giurgiutiu, V. (2004, July). Signal processing techniques for damage detection with piezoelectric wafer active sensors and embedded ultrasonic structural radar. In Smart Structures and Materials (pp. 492-503). International Society for Optics and Photonics.
[17] Garcia, D., & Trendafilova, I. (2014). A multivariate data analysis approach towards vibration analysis and vibration-based damage assessment: Application for delamination detection in a composite beam. Journal of Sound and Vibration, 333(25), 7036-7050.
[18] Jolliffe, I. (2005). Principal component analysis. John Wiley & Sons, Ltd.
[19] Mujica, L. E., Rodellar, J., Fernandez, A., & Guemes, A. (2010). Q-statistic and T2-statistic PCA-based measures for damage assessment in structures. Structural Health Monitoring, 1475921710388972.
[20] Mujica, L. E., Vehi, J., Ruiz, M., Verleysen, M., Staszewski, W., & Worden, K. (2008). Multivariate statistics process control for dimensionality reduction in structural assessment. Mechanical Systems and Signal Processing, 22(1), 155-171.
[21] Nomikos, P., & MacGregor, J. F. (1994). Monitoring batch processes using multiway principal component analysis. AIChE Journal, 40(8), 1361-1375.
[22] Hyötyniemi, H. (2001). Multivariate regression: Techniques and tools. Helsinki University of Technology.
[23] Kohonen, T., & Maps, S. O. (1995). Springer series in information sciences. Self-organizing maps, 30.