Automatic Tuning of a Pipeline Faults Detection Algorithm

Rodolfo Villamizar, Oscar Eduardo Perez, Jhonatan Camacho Navarro

AUTOMATIC TUNING OF A PIPELINE FAULTS DETECTION ALGORITHM

Rodolfo Villamizar1, Oscar Eduardo Perez1, Jhonatan Camacho Navarro1

[1]*Escuela de Ingenierías Eléctrica, Electrónica y de Telecomunicaciones (E3T), Universidad Industrial de Santander (UIS), Grupo de Control Electrónica Modelado y Simulación (CEMOS), Carrera 27 Calle 9 Ciudad Universitaria, LP 112A, Santander, Colombia.*

rovillam@uis.edu.co, oscar.perez.gamboa@gmail.com, camacho.navarro.jhonatan@gmail.com

ABSTRACT

This paper discusses the experimental results obtained by using a principal component analysis based algorithm in joint with Self Organizing Maps (SOM) Neural Networks for detection of damages in pipeline structures. Also, a differential evolutive algorithm is used for tunig the neural network parameters. A pipeline section test structure was instrumented with an active piezoelectric system in order to apply a high known frequency signal and to determine the base-line structural dynamical performance. Several piezoelectric sensors were located along the structure surface and damage features are obtained by processing the time vibrational dynamical response through principal component analysis. Q-statistic and Hotteling T^2 indexes obtained from the PCA analysis are used to detect deviations of the current vibrational response respect to the undamaged one. Algorithm validation was achieved by using experimental data, obtained from a carbon steel tubing section, where damages were induced by adding masses to the structure. The obtained results indicate that is possible to identify and locate faults in pipeline structures and by using evolutive differential genetic technique improve the performance of the studied algorithm.

KEYWORDS: *Pipeline fault detection, Genetic algorithms, Self Organizing Maps, principal component analysis, vibrational data.*

INTRODUCTION

Structural Health Monitoring (SHM) involves methodologies with the capability to detect, locate and discriminate damages in mechanical, civil or aerospatial structures among others [1]. In recent years, several proposed methods for damage identification allow the assessment for cracks, mass aggregation or impact detection [8, 9]. Most of these approaches are implemented using signal processing techniques like Fourier Transform [3] or Wavelet Analysis [11]. The objective of these techniques is to characterize dynamical response through vibrational modes or to obtain time-frequency features in order to identify deviations from the nominal performance of the undamaged structure. However, by using Principal Component Analysis (PCA) it has been shown that it is possible to solve SHM tasks taking into account only the time vibrational response of the structure [2-5]. PCA based methods correlate time vibrational data recorded from sensors attached at different locations of the structure and then, statistical indexes are calculated by using the projection of time vibrational data onto the reduced principal component space. In these sense, computational cost is lower than techniques based on frequency-time processing. In addition, resulting information is very simple to interpret by means of graphical tools.

It is possible to detect and locate damages using PCA based method [9]. For damage classification, it is necessary to add pattern recognition tools such as neural networks, support vector machines or bayessian methods. Some works are focused on using Self Organizing Maps (SOM) because its capability to operate in an unsupervised way and the possibility for learning of new damages cases [12]. However, the training stage requires appropriate tuning of several parameters that influence meanly on the success of the damage identification strategy. For this reason, it is recommended to use metaheuristics methods for an automatic tuning in order to avoid poor performance of the methodology. In this sense, genetic algorithms can be used because its simplicity and flexibility, identification errors in the fitness function can be included.

Thus, this paper discusses an approach for automatically tuning parameters of a structural assessment algortihm experimented by CoDalab research group [12]. The algorithm includes PCA statistics and SOM network, and the automatic tuning is obtained by using differential evolutive algorithms.

1 ALGORITHM DESCRIPTION AND TECHNICAL FORMULATION

Figure 1 shows the structural damage assessment algorithm based on PCA method. The procedure is focused on damage detection and classification.

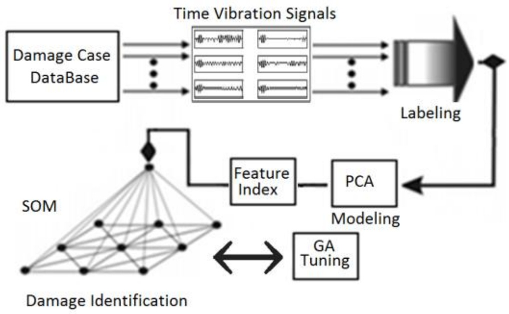


Figure 1. Studied algorithm for structural fault detection

The first step is to organize the experimental data from sensors for diferent damage cases into a vibrational time data matrix. The structure is excited by the senoidal high frequency signal presented in Figure 2. An active Piezo-Electric System with N nodes (i.e. N PZT attached to the structure at different locations), is used to analize the time vibrational response. One of the N nodes is selected as an actuator point in order to excite the structure and the reaminning $N-1$ PZT nodes operate as sensors.

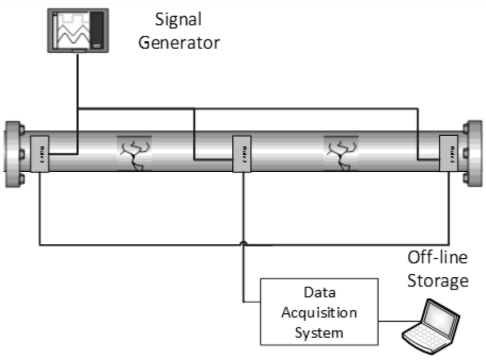


Figure 2. Active Piezo-Electric System

The experiment is repeated for non-destructive damages induced over the structure. In this sense, r experiments are stored for all M damage cases by using t_s time samples. Figure 3 illustrates the unfolded matriz of damage cases.

$$
\begin{array}{l}
Undamaged\begin{cases}
\overbrace{X_{10}(t_1)^{s_1}\ldots X_{10}(t_s)^{s_1}}^{PZT\ Sensor\ 1} & \cdots & \overbrace{X_{10}(t_i)^{s_i}}^{PZT\ Sensor\ i} & & \overbrace{X_{10}(t_s)^{SN} \quad \cdots \quad X_{10}(t_s)^{SN}}^{PZT\ Sensor\ N} \\
\vdots \quad \ldots \quad \vdots & \cdots & \vdots & \cdots & \vdots \quad \cdots \quad \vdots \\
X_{r0}(t_1)^{s_1}\ldots X_{r0}(t_s)^{s_1} & \cdots & X_{r0}(t_i)^{s_i} & & X_{r0}(t_s)^{SN} \quad \cdots \quad X_{r0}(t_s)^{SN}
\end{cases} \\
Damage\ 1\begin{cases}
X_{11}(t_1)^{s_1}\ldots X_{11}(t_s)^{s_1} & \cdots & X_{11}(t_i)^{s_i} & & X_{11}(t_s)^{SN} \quad \cdots \quad X_{11}(t_s)^{SN} \\
\vdots \quad \ldots \quad \vdots & \cdots & X_{21}(t_i)^{s_i} & \cdots & \vdots \quad \cdots \quad \vdots \\
X_{r1}(t_1)^{s_1}\ldots X_{r1}(t_s)^{s_1} & \cdots & X_{r1}(t_i)^{s_i} & & X_{r1}(t_s)^{SN} \quad \cdots \quad X_{r1}(t_s)^{SN}
\end{cases} \\
Damage\ 2\begin{cases}
X_{12}(t_1)^{s_1}\ldots X_{12}(t_s)^{s_1} & \cdots & X_{12}(t_i)^{s_i} & & X_{12}(t_s)^{SN} \quad \cdots \quad X_{12}(t_s)^{SN} \\
\vdots \quad \ldots \quad \vdots & \cdots & \vdots & \cdots & \vdots \quad \cdots \quad \vdots \\
X_{r2}(t_1)^{s_1}\ldots X_{r2}(t_s)^{s_1} & \cdots & X_{r2}(t_i)^{s_i} & & X_{r2}(t_s)^{SN} \quad \cdots \quad X_{r2}(t_s)^{SN}
\end{cases} \\
\qquad\qquad\qquad \vdots \qquad\qquad\qquad\qquad \ddots \qquad\qquad \vdots \\
Damage\ M\{X_{Mi}(t_x)^{s_1} \qquad\qquad \cdots \qquad\qquad X_{Mi}(t_x)^{SN}
\end{array}
$$

Figure 3. Damage – Data Case Matrix

Only the undamaged data of the unfolded matrix are used to construct a PCA model. The undamaged principal component space is used to project vibrational analysis cases onto the reduced representation. Then, the multivariate T^2 and Q- statistics are estimated in order to obtain indexes with the capability of distinguishing the projections from the Undamage-Damage states. Then, the interpretation of indexes computed is carried out by using graphical tools.

1.1. PCA Model Construction: Damage Detection and Clasification.

The nndamaged model for the healthy structure is obtained according to the PCA procedure [9]. In this sense, the steps followed to obtain a representation for undamaged state are:

- Normalize the undamaged DataCase Matriz X_{und} by applying GroupScaling (GS) method [8].
- Determine the most significant loading vectors by using the NIPALS algorithm [9], which is an iterative procedure to obtain the first l elements of the Eigenvalue-Eigenvector decomposition

After previous steps are accomplished, the PCA model for the undamaged state is defined by the following elements:

- Normalization components: means ($\hat{\mu}_i$) and deviations ($\hat{\sigma}_i$). According to GS procedure are obtained $N * t_s$ mean values and N deviations.
- The $\boldsymbol{l}$ loading vectors (principal components): $\varphi_i]^{Und} = [\phi_1 \quad \ldots \quad \phi_l]$
- The variance of principal components (Eigenvalues):$\lambda_i = [\lambda_1 \quad \ldots \quad \lambda_l]$

PCA model is used to obtain the projection (scores) of remaining Normalize Damage Cases onto the Undamaged PCA reduced space:

$$Z_{\mathrm{damage}} = \varphi_{\mathrm{i}}]^{\mathrm{Und}} * \left(\frac{X_{\mathrm{Damage}} - \hat{\mu}_{\mathrm{i}}}{\hat{\sigma}_{\mathrm{i}}}\right) \tag{1}$$

Because the T^2 and Q statistics are commonly used in tasks involving fault detections [8], they are estimated in order to take into account indexes with the ability for measure deviations of new data respect to the PCA model:

$$T^2 = X^T\varphi]^{\mathrm{Und}}\left(\Sigma^T\Sigma\right)^{-1}\left(\varphi]^{\mathrm{Und}}\right)^T X \tag{2}$$

$$Q = r^T r, \quad r = \left[I - \varphi]^{\mathrm{Und}} * \left(\varphi]^{\mathrm{Und}}\right)^T\right] X \tag{4}$$

Where Σ is a diagonal matrix comprising λ_i values and r is the projection onto the residual left components. The T^2 statistic allows measures for variation inside the PCA model; instead, the Q statistic is a squared 2-norm measuring the deviation of the observation to the lower-dimensional PCA representation avoiding overly sensitive to inaccuracies in the PCA space corresponding to the smaller singular values.

The scores and PCA indexes are used as features to train a SOM Network to facilitate visualization of diferent damage types and to assist classification tasks. The SOM Network groups the Damage Cases into clusters using distance-topology similarity measures through an unsupervised learning algorithm [14].

1.2. SOM Automatic tuning by using Differential Evolutive Algorithm

The success of the damage classification methodology is highly dependent of a proper SOM training tuning stage. The SOM algorithm requires selecting the following parameters [15]:

- **Normalization method:** It is necesary to normalize the data input between different options, which include variance, linear range, logarithm and logist. Data normalization avoids false dominant clusters.
- **Output neurons number:** It is the *clusters* number
- **Grid structure**: Local topology map → Rectangular or Hexagonal.
- **Map shape**: Local topology map→Laminate, Cylindrical or Toroidal.
- **Neighborhood function:** it is the interation between reference vectors and affects the precision and generalization of the SOM network. It can be *Gaussian*, cut *Gaussian* and Bubble.

SOM quality is commonly described by the following indexes **[17], [18]**:

- **Topographical error**: It is a measurement of topology preservation **[17]**. It should be near to zero.
- **Distortion**: Shows how well each neuron represents the input data **[19].**
- **Histogram uniformity**: It is measurement of the cases distribution in the clusters. Ideally, each cluster should be containing cases of the same type and there is not be empty clusters.

In order to apply a methodology for automatic tuning of the SOM network, it is proposed to minimize the following fitness function:

$$f(\vec{x}_{i,j}) = \sum_{i=1}^{n} w_i * q_i + \sum_{j=1}^{M} w_j * e_j \quad (5)$$

Where $\vec{x}_{i,j}$ is a vector containing a combination of SOM parameters; w_i, w_j are weighting factors, e_j the classification error for each damage type and q_i are the SOM quality indexes. The fitness function is minimized by applying the Differential Evolutive Algorithm (DE).

DE is a stochastic algorithm based on population focused in a scheme for generating trial parameter vectors. Thus, the main operator consist of adding the weighted difference between two population vectors to a third vector. DE algorithm has been tested on several optimization tasks getting good success rates [19]. Figure 4 summarizes the operation mode for DE algorithm, which consist of seven main steps:

- Generate a random population according to the SOM parameters.
- Choose randomly two population members as target vector and base vector.
- Compute the weigthed difference vector between two random population members.
- Add the resulting difference to the base vector and apply mutant operator.
- Apply crossover operator between target vector and the result of step $\text{iv.})$ in order to obtain a trial vector.
- Select the trial vector or target vector as member of the next population according to fitness evolution.
- Repeat steps $\text{ii.}) - \text{iv.})$ in order to obtain all members of the new generation.

2 EXPERIMENTAL RESULTS AND DISCUSSION

The test structure used is a section of carbon steel pipeline (similar to those used in the local industry) flanged at the extremes, whose dimensions are: 1m x 1 x 3mm (Large x diameter x thickness). Time vibrational response was recorded by using an active piezoelectric system with three PZT nodes distributed along the surface of the structure (see Figure 4). Every PZT was operated both sensor and actuator, thus if PZT n is used to excite the structure, the remaining PZT recorded signals are used to build the PCA model. The obtained model is named Model n (n=1,2,3) and this configuration is repeated for all every PZT, thus 3 different PCA models are built. The structure was excited during one period of a 350 KHz sinusoidal signal and time vibrational response, corresponding to the difference between actuation signal and sensed signal (Ch1 – Ch2), were recorded and stored by using a scope with sample period equal to 40ns.

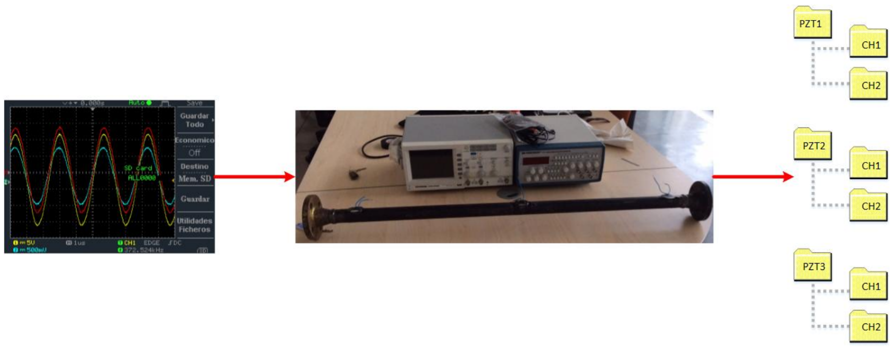
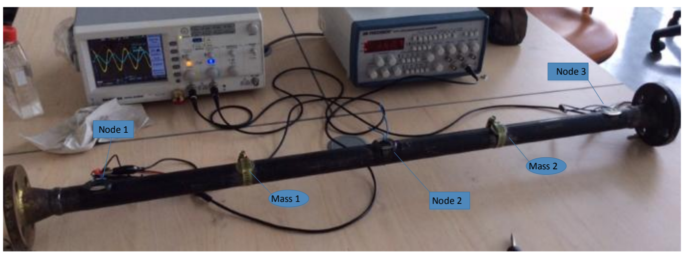


Figure 4. Tubing section and active PZT system

Structural damages were induced by adding two masses between nodes 1 and 2 (see Figure 5). In order to building the PCA model, seven undamaged cases were used (five for training purposes – orig, and two for validation purposes - und). In order to evaluate the performance of the fault detection algorithm, three different damage types (D1, D2, and D3) were considered, which are described in

table of Figure 5. Seven repetitions for every damage case were conducted and PCA models were built by retaining four principal components.



Damage type	Description
D_1	Mass between nodes 1 and 2
D_2	Mass between nodes 2 and 3
D_3	D_1 and D_2 together

Figure 5. Locations of generated damages and description

Figures 6 and 7 depict the two most significant scores and the damage detection idexes (T^2 and Q) in a scatter plot for the PCA-Models. Shapes and colours represent different damage types. Original data corresponds to the undamaged cases used to build the PCA model labeled with tag orig.

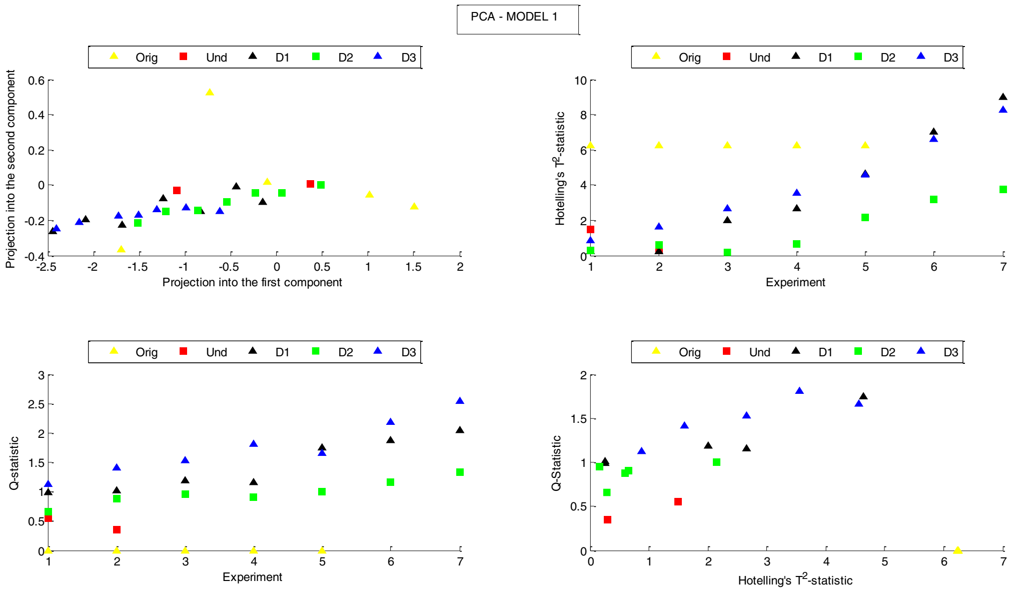


Figure 6. Results for PCA-model 1.

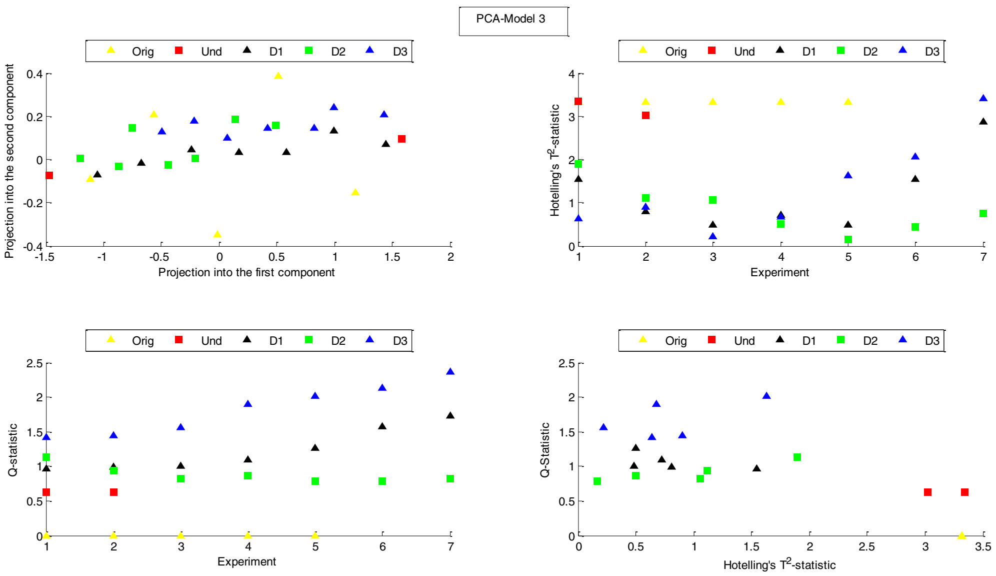


Figure 7. Results for PCA-model 3.

In Figures 6 and 7 can be observed that undamaged cases (Orig, Und) are clearly separated from damaged cases (D1-D3) by using the satatistical indexes. Results are similar for PCA model 2. Thus, the PCA model allows detecting easily the presence or absence of damages. Discrimination of damages is hardly for some groups because they appear quite overlapped. By using PCA-Model 3, damage 3 is most distinguishable than others, which take into account quantification performance.

In order to take into account possible nonlinear interations between the features used to detect damages, a SOM network was built in order to map the inputs onto a 2D cluster representation. Figure 8 depicts the resulting Map. The SOM network was trained taking into account all PCA-modelsand using parametric values by default: map size: [6 5]; lattice: hexa; shape: sheet; norm method: var; neigh: gaussian. The Final quantization error: 1.014 and Final topographic error: 0.000.

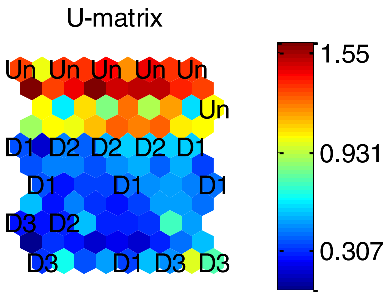


Figure 8. SOM network using default-training values

By training group labels using the SOM network it was obtained a high identification error, where only 4/30 cases were properly labelled. Then a differential genetic algorithm (GA) was programmed in order to obtain the optimal tuning parameter of SOM neural network. Figure 9 shows the evolution of the fitness function, where the weigthed sum of identification errors and SOM quality indexes were normalized to one.

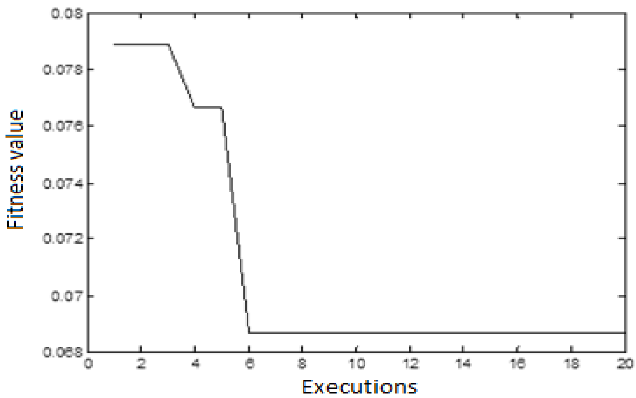


Figure 9. Fitness function evolution for the GA

After applyng the optimization it was found that the SOM Neural Network automatically tuned by using the differencial genetic algorithm, improves the identification error at rate of at least 50 percent. Because only a few of experiments were used to bult the PCA model, it is expected its improving if this number is increased.

CONCLUSION

In this paper a fault detection algoritm based on PCA analysis and a SOM neural network automatically tuned by means of a differential genetic algorithm, was experimentaly validated on a section of pipeline. It was observed that Q-statistics is a better index to separate damage cases of undagmaged cases, than scores or T2-statistics. On other hadn, by using a SOM neural network to separate damage cases, it was found that by using default parameters is not adequated and if optimal values are found by using differential GA, performance algorithm is improved. However, because only a few undamaged cases were used to build the PCA model, a better conclusion will be obtained by using more of them.

ACKNOWLEDGMENTS

This work has been developed as part of the *Monitorización y Detección de Defectos en Estructuras usando Algoritmos Expertos Embebidos* research project, 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-*Febrero* 2015.

REFERENCES

[1] A. Quintero and R. Villamizar. Estado del arte en monitorización de salud estructural: Un enfoque basado en agentes inteligentes. *Ciencia e Ingeniería Neogranadina*, Vol. 20-1, pp. 117-132. Bogotá, Junio de 2010.

[2] D . Giraldo. A Structural Health Monitoring Framework for Civil Structures. *Doctoral Dissertation thesis. Washington*, University. 2006.

[3] Caicedo, J.M. and Johnson, E.A. Structural Health Monitoring of Flexible Civil Structures, *Doctoral Dissertation thesis*. Washington, University. 2003.

[4] Structural Control and Earthquak Engineering Laboratory. Washington University. Available at: http://mase.wustl.edu/wusceel/quake/. Last Access: July 2013.

[5] Marulanda, J., Thomson, P., Caicedo, J. M., Dyke, and Shirley J. Monitoreo de Salud Estructural de Puentes Metálicos. *Proceedings of the First Colombian Workshop on Steel Structures, Cartagena, Colombia*, Sep. 26-28, 2001.

[6] Camacho, J., and Villamizar, R. Structural Health Monitoring by Using Pattern Recognition. *International Journal of Artifitial intelligence – IJAI.* Volume 5 number A10. ISSN 0974-0635. Autum 2010.

[7] T. Clarke, Guided Wave Health Monitoring Of Complex Structures, *Imperial Collage London*, 182, 2009.

[8] J. R. A. F. A. A. G. Le Mujica, Q-Statistic And T2-Statistic Pca-Based Measures For Damage Assessment In Structures, *Structural Health Monitoring*, 16, 2011.

[9] L. M. J. R. D.A Tibaduiza, Comparison of Several Methods For Damage Localization Using and Contributions Based On Pca, *9th International Conference On Damage Assessment Of Structures*, 11, 2011.

[10] L. E. M, X. B. A. J. R. M Ruiz, Partial Least Square/Projection To Latent Structures (Pls) Regression To Estimate Impact Localization In Structures, *Smart Mater. Struct*, 12, 2013.

[11] T. W. W. T. L. G. B. A. P. T. P.Malinowski, Signal Processing System For Guided Wave-Based Shm Techniue, *Proceeding of the 9th International Workshop on Structural Health Monitoring*, Stanford University, California, USA, 2011.

[12] Burgos, D. A.. Design and validation of a structural health monitoring system for aeronautical structures. *Departament de Matematica Aplicada III Universitat Politecnica de Catalunya* , 158, 2012.

[13] Storn, R., and Price, K, Differential evolution – a simple and efficient heuristic for global optimization overcontinuous spaces. *J. of Global Optimization,* 11:4, 1997

[14] Price, K.V., Storn, R.M., Lampinen, J.A. Differential Evolution A Practical Approach to Global Optimization Natural Computing Series. *Springer-Verlag*, Berlin, Germany, 2005.

[15] Kenneth V. P., Rainer M.S., and Jouni A. Lampinen.DifferentialEvolution A Practical Approach to Global Optimization. Natural Computing Series. *Springer-Verlag, Berlin*, Germany, 2005.

[16] T. Kohonen, Self-Organizing Maps. 3, 3rd ed. *Springer*, 2001.

[17] Laboratory of Computer and Information Science (CIS) Adaptive Informatics Research Centre. *Department of Computer Science and Engineering at the Helsinki University of Technology*. http://www.cis.hut.fi/somtoolbox/. Citado Julio de 2013.

[18] Santana, L.V. Un Algoritmo Basado en Evolución Diferencial para Resolver Problemas Multiobjetivo. Master Thesis. *Centro de Investigación y de Estudios Avanzadosdel Instituto Politécnico Nacional Departamento de Ingeniería Eléctrica Sección de Computación*, México DF. Noviembre de 2004.

[19] Monterrosa, C. Analizando la Relación entre elFactor de Escalamiento y losTamaños de Población en elAlgoritmo de Evolución Diferencial al Resolver Problemasde Optimización con Restricciones. *Dirección General de Educación Superior Tecnológica Instituto Tecnológico de Tuxtla Gutiérrez. Departamento de Ingeniería en Sistemas Computacionales*. México, Enero de 2010.