

FAULT DIAGNOSIS USING GENETIC ALGORITHMS AND PRINCIPAL CURVES

Tawfik Najeh¹, Lotfi Nabli²

¹ PHD student, Laboratory of Automatic Control, Signal and Image Processing, National School of Engineers of Monastir, University of Monastir, 5019 Tunisia, najehtawfik@gmail.com,

² Master of conference, Laboratory of Automatic Control, Signal and Image Processing, National School of Engineers of Monastir, University of Monastir, 5019 Tunisia, lotfi.nabli@anim.rnu.tn

Abstract

Several applications of nonlinear principal component analysis (NPCA) have appeared recently in process monitoring and fault diagnosis. In this paper a new approach is proposed for fault detection based on principal curves and genetic algorithms. The principal curve is a generation of linear principal component (PCA) introduced by Hastie as a parametric curve passes satisfactorily through the middle of data. The existing principal curves algorithms employ the first component of the data as an initial estimation of principal curve. However the dependence on initial line leads to a lack of flexibility and the final curve is only satisfactory for specific problems. In this paper we extend this work in two ways. First, we propose a new method based on genetic algorithms to find the principal curve. Here, lines are fitted and connected to form polygonal lines (PL). Second, potential application of principal curves is discussed. An example is used to illustrate fault diagnosis of nonlinear process using the proposed approach.

Index Terms: Principal curve, Genetic Algorithm, Nonlinear principal component analysis, Fault detection.

1. Introduction

Although the principal component analysis (PCA) can be considered as the oldest and simplest technique in multifarious analysis [2], this method is a topical subject in several research areas such process monitoring and fault detection [2]. It belongs to the most effective techniques for linear data analysis [3]. PCA reduces the dimension of the representation space of data with minimal loss of information. The transformation of the whole original variables gives new ones called principal components. The first major components contain the most variations of initial variables. The last major components called primary residues contain the measurement noise. Despite the effectiveness of PCA in various fields, this method is no longer approved for nonlinear problems that present most real systems [4]. Using linear PCA to analyze nonlinear data gives a residual space with significant information part rather than containing the system noise. This will lead to the loss of major data variations. In order to adapt the concept of PCA to nonlinear problems, an inspired approach to PCA called principal curves or nonlinear principal component analysis (NPCA) is introduced by Hastie et al [5]. This approach can be seen as a generalization and the natural extension of linear PCA. It is defined by curves that pass through the middle of a cloud of points.

In spite of the progress reported during recent years in resolving the main problem of principal curves, existing approaches have multiple limitations. Due to the inadequate initialization of the algorithm or the predefined strategy of constructing local models, these methods present weaknesses for the data with self-intersecting

characteristics, high curvature, and significant dispersion [6]. The contribution of this paper is new practical concepts that use genetic algorithms (GAs) to estimate curve passing through the middle of probability distribution. To support the complex data structures and have a flexible algorithm that covers the important curvature of the principal curves, a new approach that differs from conventional mentioned approaches is used. This serves to demonstrate the effectiveness of using revolutionary techniques for solving the problem of the principal curves. Likewise, the use of algorithms is easily extended to complex data structures that present a great challenge for conventional approaches. The second section is an overview of different definitions of principal curve. The third section briefly outlines the genetic approach and how it differs from conventional techniques. The fourth section is the core of paper; this section describes the developed method based on genetic algorithms for constructing the principal curves and the technical details of the proposed approach. In the fifth section the results obtained on two examples and the perspectives for future research.

2. Overview of principal curves

In the literature various definitions of principal curves are proposed. The first one is based on the self-consistency property [5]. A different approach is introduced by Kégal et al [7] and regards the principal curves as curves with bounded length. Other approaches are based on the mixture model [8]. The Hastie approach does not support either closed curves or those with intersections. However many improvements are added to this approach. The fixes

developed by Tibershirani [9] of the original version of the algorithm of Hastie allowed it to overcome the bias. Later Chang et al.[10] proposed a hybrid approach based on Hastie and Banifield methods to improve obtained results on finding principal curves.

A different approach based on a model of the principal semi-parametric curve was proposed by Tibshirani [8]. But lack of flexibility in the case of curves with large curvatures this method has the same weaknesses of the theory of Hastie. In [11] authors introduced an algorithm that relies on the polygonal lines to achieve a satisfactory principal curve, but with an important calculation cost. All this algorithms are characterized by their top-down analysis. They start research for a straight line, which by default is the first principal component of the data. Besides, another group of definitions called bottom-up approach describes the same topic, but, from another side. Such approaches are based on local data analysis to construct the principal curve [6] rather than starting with a line that represents the entire cloud of points, these algorithms consider just a set of neighboring points in each step. The approach introduced by Delicado [12] for the construction of principal curves is directly inspired by this principle. The second group seems to be more effective for complex data structures such as closed curves or spiral-shaped [7]. It should be noted that considering all the data cloud at once make the result completely dependent on the initial line. This has a direct impact on the flexibility of the algorithm and the final curve is only satisfactory for specific problems [13]. Principal curves are considered as a nonlinear way to summarize a multidimensional data by a one-dimensional curve. They are defined as curves that pass through the most dense regions of the data cloud [14]. This one-dimensional representation catches shape depending on the data distributions (Figure 1).

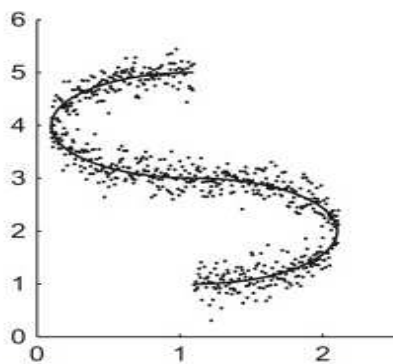


Figure.1. Graphical representation of a curve that passes from the middle of a cloud of points

Hastie and Stuetz introduce the first approach based on the self-consistency property of principal components.

Given a random vector $X \in \mathbb{R}^d$ if $f(\lambda) = (f_1(\lambda) \dots f_d(\lambda))$, $\lambda \in \mathbb{R}$ is the curve parameterized with $\lambda \in \mathbb{R}$, then for any $X \in \mathbb{R}^d$ we have $\lambda_f(X)$ the projection index and $f(\lambda)$ is a principal curve. Mathematically the projection index is defined by:

$$\lambda_f(X) = \sup\{\lambda : \|X - f(\lambda)\| = \inf_{\tau} \|X - f(\tau)\|\} \quad (1)$$

For $X \in \mathbb{R}^d$, the projection index $\lambda_f(X)$ is the largest value giving the minimum of $\|X - f(\lambda)\|$, as shown in (Figure 2).

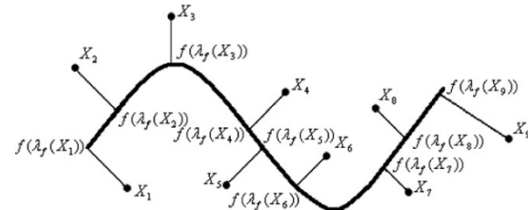


Figure.2. Projecting points on a curve

self-contained, the basic terms on distance from a point to a curve and the point's position along a line segment on \mathbb{R}^2 are reviewed.

Let the point $C(C_x, C_y)$ and the line segment AB from $A(A_x, A_y)$ to $B(B_x, B_y)$. Let P be the point of perpendicular projection of C on AB . The parameter r , which indicates P 's position along AB , is computed by the dot product of AC and AB divided by the square of the length of AB :

$$r = \frac{AC \text{ dot } AB}{\|AB\|^2} \quad (2)$$

The parameter r has the following meaning:

if $r = 0$ then $P = A$

if $r = 1$ then $P = B$

if $r < 0$ then P is on the backward extension of AB

if $r > 1$ then P is on the forward extension of AB

if $0 < r < 1$ then P is interior to AB

point P can then be found:

$$\begin{aligned} P_x &= A_x + r \cdot (B_x - A_x) \\ P_y &= A_y + r \cdot (B_y - A_y) \end{aligned} \quad (3)$$

And the distance from A to P is:

$$d = r \times \|AB\| \quad (4)$$

3. Genetic algorithms-some preliminaries

Gas are classified as evolutionary optimization methods [15]. They allow defining an optimum of a function defined on a data space. These algorithms are applied to a wide variety of problems [16]. Simplicity and efficiency are the two advantages of GAs [17]. After having fixed the expression of the objective function to be optimized, probabilistic steps are involved to create an initial population of individuals [18]. The initial population will submit some genetic operations (such as evaluation, mutation, crossover ...) to converge to the optimal solution of the problem. Optimization steps with GAs are as follows:

- **Initialization:** It is usually random and it is often advantageous to include the maximum knowledge about the problem.
- **Evaluation:** This step is to compute the quality of individuals by the allocation a positive value called "ability or fitness" to each one. The highest is assigned to the individual that minimizes (or maximizes) the objective function.
- **The selection:** This step selects a definite number of individuals of the current population. The selection is probabilistic and characterized by a parameter called selection pressure (P_s) fixing the number of selected individuals.
- **The crossover:** The genetic crossover operator creates new individuals. From two randomly selected parents, crossover produces two descendants. This step affects only a limited number of individuals established by the crossover rate (P_c) number.
- **The mutation:** The mutation consists on providing a small disruption to a number (P_m) of individuals. The effect of this operator is to counteract the attraction exerted by the best individuals; this allows us to explore other areas of the search space.

This has been a brief overview of GAs. For further details on the processing power and the convergence properties of GAs, reference should be made to [15].

4. Principal curves by genetic algorithms

It has been shown in Section 2 that obtaining the principal curves leads back to an optimization problem. This principal curve minimizes the orthogonal deviations between the data and the curve subject to smoothness constraints. To resolve this computation, several methods have been established; all of them were deterministic. The use of these methods can lead to a local optimum providing less faithful principal curves to the data distribution. The resolution of such problems by GAs avoids all local optima and converges to the global optimum of the problem. The proposed approach considers the principal curves as an ensemble of connected line segments. In each step new segment is inserted to form polygonal lines. GAs are used to find the optimal segments one by one. In each step, only data in a local neighbourhood of the segment are considered. The use of GAs in order to find the principal curves requires the development of an objective function. This function determines how to evaluate the fitness of any segment pass through the data cloud.

Let X_n be a two-dimensional random vector and $x_i \in \square^2, i = 1, \dots, n$ is a local neighbourhood for a fix point x_0 . For any segment $s=[x_0, b]$ pass through the neighborhood points let d_i denotes the distance between the considered point x_i and its orthogonal projection on the segment given by equation (4). The objective function must take into account the quadratic sum of the distances d_i .

$$\Phi = \sum_i^n d(x_i, s)^2 = \sum_i^n (\min \|x_i - f\|)^2 \quad (5)$$

The objective is to find the segment that minimizes the total squared distance of all neighbour's points. We can project orthogonally any data point in the neighbourhood cloud onto the segment and the deviations between the data and the projection on considered segment are computed by equation (5). Hence, for any segment s ; an associated fitness is calculated by Φ .

The size of the neighbourhood is an important parameter done by a fraction of all data points that are considered to be in the neighbourhood. Increasing or decreasing this parameter tends respectively to the smoothness of the curve or to interpolate the data cloud. Then the neighbourhood controls the trade-off between smoothness and fidelity to the data. To determine the size of neighbourhood one can use the cross-validation.

The new genetic curve algorithm is constructed following the strategy outlined as follows:

Algorithm 1

01: Start with a random point x_0 from X_n .

02: L_n = the local neighborhood around x_0 .

03: Repeat until number generation (NB) = $Gmax$.

04: Generate at random a population of K segments.

05: for every segment x_k do

06: compute the Euclidean distance:

$$\phi_k = \sum_i^n d(x_i, s)^2 \text{ for each segment.}$$

07: end for

08: Apply crossover and mutation.

09: Save the better line segment and return to step 3.

10: $NB = NB + 1$

11: Delete used local neighbourhood points from the original distribution.

12: The end point of previous segment is served as a starting point.

13: Connect with previous segment (Except first one).

14: Return 02 and repeat until all data points are achieved.

The series of best saved segments gives the global principal curve for the data distribution. In the next section these steps will be explained in detail:

(1) *Selection of the starting point:* In this paper the starting point can be imposed by hand or if this is not possible it might be chosen at random.

(2) *Search the local neighborhood:* For each segment in the current population the algorithm identifies distances from all points to them orthogonally projections and a fixed number of the nearest neighbor's points might be chosen.

(3) *Stop after Gmax iteration:* Following the execution of a specified number of Gmax generations the genetic algorithm converges to the global optimum. The result is a satisfactory

segment passes through the middle of neighborhood of the starting points.

(4) *Generate an initial population:* The point x_0 is taken as the end of a set of K segments having all the same length. These segments randomly generated present the initial population of individuals. Each segment represents a possible solution for the local neighbourhood (Figure 3).

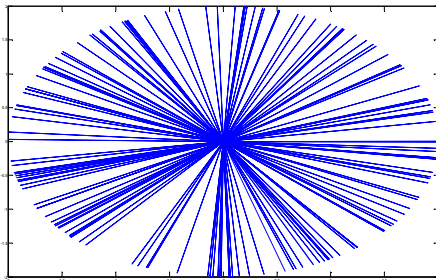


Figure 3. Initial Population

(5,6,7) *The assignment of fitness value:* The equation (5) is taken as an objective function. After having generated an initial population, the abilities of individuals are evaluated by this function. Individuals with the highest fitness are selected to be subject to different genetic operators (crossover, mutation and selection).

(8) *Apply crossover and mutation:* The fitness value assigned in the previous step is used to direct the application of two operations, crossover and mutation, which produce the new populations of segments. This new generation will contain better solutions. Crossover affords a means for each segment to mix and match their desirable qualities through a random process. The second operation, mutation, enhances the ability of the GAs to find near-optimal segments. Mutation is the occasional and partial modification of a particular end point of a given segment.

(10) *Save the better line segments:* After saving best segments return to step 3 for fixed number of generations (G_{max})

(11) *Delete used local neighbourhood:* Before moving to the next segment calculation, all points previously used are removed from the distribution. This avoids the algorithm to dangle between two points and never escaping.

(12,13,14) *Found the global curve:* The end point of the previous segment serves as a new starting point for a new algorithm. Steps (2) to (14) are repeated times until no point in the distribution can be considered to be in the neighbourhood. Then return the global principal curve those summaries the data cloud.

5. Experimental results and analyses

In this section, two examples are used to prove the performance of the proposed approach to find the principal curve in case of synthetic dataset and for fault detection. But before, we first discuss parameters of GAs.

5.1 Setting parameters

For all runs of GAs, the population size is 80 and the maximum number of generations to be run is $G_{max}=100$. Two genetic operations are used: crossover and mutation with the probability 0.25 and 0.01, respectively. The maximum number of permissible segments inserted in the new generation is 90%. The method of selection of segments for crossover is tournament selection.

5.2 Synthetic dataset

The new algorithm has been tested with synthetic datasets in case of circle of radius=1. Data points are disturbed along the circle with Gaussian noise independently imposed on different dimensions of the given curve. We start with 100 data point contaminated with small noise (0.03). The results obtained by applying the algorithm are shown in Figure (4). One can see that the circle is reconstructed quite well.

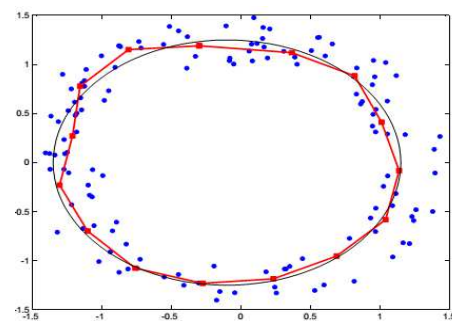


Figure 4. Principal curve: synthetic datasets

5.3 Tennessee Eastman

The Tennessee Eastman Process is a test problem published by the Eastman Company for use in evaluating process control methods [19]. This process produces two products G and H from four reactants A, C, D, and E.

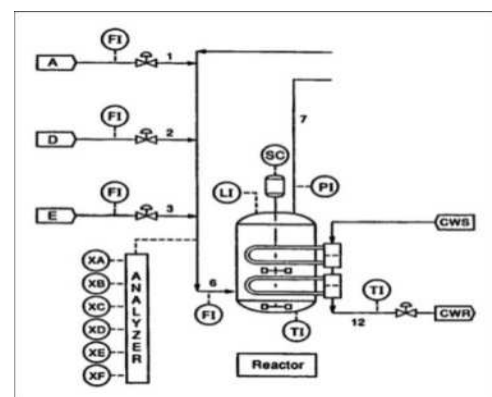


Figure 5. Tennessee Eastman reactor

The process also produces an inert B and a byproduct F. The major units of the process are: a reactor, a product condenser, a vapor/liquid separator, a recycle compressor, and a product stripper. The process gives 41 measurements with 12 manipulated variables [20]. All gaseous reactants are transmitted to the reactor to produce other liquid products. A catalyst dissolved in the liquid phase catalyzes the gas-phase reactions. Then, the catalyst remains in the reactor and then products leave the reactor as vapors. The process has a total of six operating states which comprise a basic operating state with 50/50 mass ratios between product components G and H.

In this paper we considered only the reactor unit with 2 variables: percent of component C (Mol %) and reactor coolant temperature (Deg C). The principal curve was trained with a data set of 300 samples. The detection was performed with data containing one fault at a time with two variables. The test was designed for safe and failed operating mode.

The objective of this section is to demonstrate the application of the principal curves technique to the problem of faults detection of the Tennessee Eastman Process. Supposing that the TEP have two operating mode in which the principal curve of the first normal mode was already defined: the predefined operating mode is called M_0 , the failed operating mode is noted M_1 . To identify a change of the system's operating mode by the proposed method, we try to get the principal curve noted C_0 corresponding to normal operating mode in the absence of defect. This curve (Figure 6) is obtained by the calculating algorithm of the principal curve detailed in the previous sections.

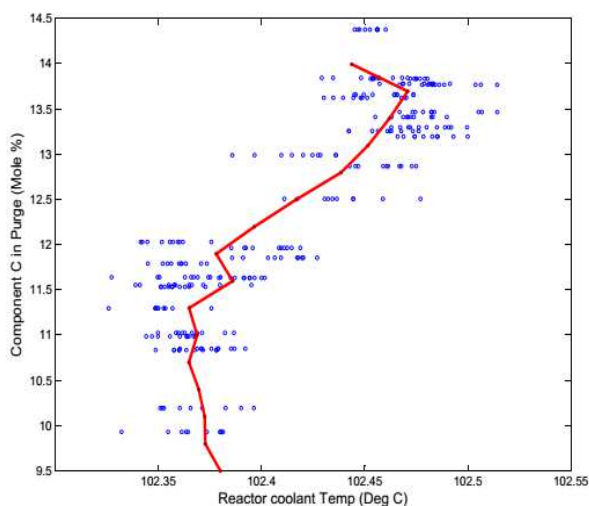


Figure 6. Principal curve

After having defined the principal curve from data, we propose, now, its use for change detection of the operating mode.

One technique for monitoring of processes uses the Squared Prediction Error (SPE) [21]. The idea consists on constructing the principal curve of safe operating mode, then check the SPE between the estimated curve and data from the system at the present time. The index below its detection threshold corresponds to the failed operating mode.

Given a set of data points x_n for each data point x_i let $p(x_i)$ being its projection point on the principal curve f . The Euclidean squared distance between x_i and $p(x_i)$ is calculated for the all data set. Then the SPE can be defined as:

$$SPE = \sum_{i=1}^n d(x_i, f)^2 \quad (6)$$

Usually, the process is considered abnormal operating if:

$$I_f > \delta \quad (7)$$

With δ is the threshold of detection.

From the constructed curve and the cloud of points, we can construct an indicator of change SPE through the equation (6). The process is simulated for 600 samples in the following manner; the first 300 samples correspond to the mode M_0 and the second 300 ones correspond to mode M_1 .

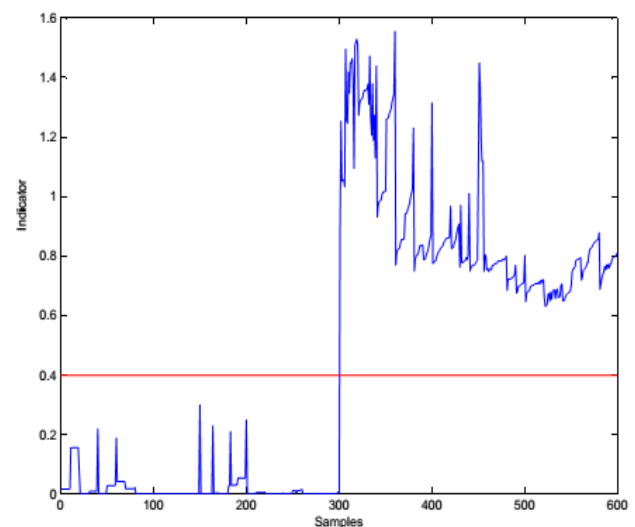


Figure 7. Variation of indicator value SPE

The evolution of indicators SPE and threshold among this simulation is shown in (Figure 7). We note the presence of several false alarms of two indicators; this is mainly due to measurement noise affecting variables. In the interval $[0, 300]$, the SPE is below the threshold corresponding to the M_0 mode and above its threshold in the interval $[301, 600]$ corresponding to failed operating mode M_1 .

6. Conclusion

In this paper genetic programming approach was applied to find principal curves. These curves are nonlinear generation of principal components, and many methods are inspired from this concept and employ principal component of the data to reconstruct the principal curves. However in many cases of high dispersion or noisy data, those conventional methods may be ineffective. To solve this problem and in connection with researching existing approaches for computing principal curves a new concept based on GAs is proposed. GAs are search algorithms that are able to find principal curves nearly indistinguishable from the true curves. It is enough flexible to be effective in a wide range

of complex data with self-intersecting characteristics, high curvature, and significant dispersion.

By using a simple GAs made up of crossover and mutation, a new approach was used to perform the principal curves. The algorithm had been applied on synthetic dataset and to the problem of fault detection on the Tennessee Eastman Process. The results demonstrate the improved performance of the proposed method in real problems. This fact leads us to conduct that GAs are a very powerful method for the construction of principal curves with higher dimension or even principal surfaces. This will be our main goal in future work.

REFERENCES

- [1] Anderson .T.W. Asymptotic theory for principal component an analysis. *Annals of Mathematics and statistics*, 2009; 45:1593-1600.
- [2] Gertler J, Formiguera R, Vicenç I. Leak detection and isolation in water distribution networks using principal components analysis and structured residuals. *Conference on Control and Fault Tolerant Systems*. 2010; 191-196.
- [3] Mina J, Verde C, Sanchez-Parra M, Ortega F. Fault isolation with principal components structured models for a gas turbine. In : 2008 American Control conference; 4268-4273.
- [4] Tharrault Y. Diagnostic de fonctionnement par analyse en composantes principales : Application à une station de traitement des eaux usées. Institut National Polytechnique de Lorraine, Nancy. France, 2008.
- [5] Hastie T, Stuetzel W. Principal curves. *Journal of American statistical Association* 1989; 84: 502-512.
- [6] Delicado, P. and Huerta, M. Principal curves of oriented points: Theoretical and computational improvements. *Computational Statistics* 2003; 18: 293-315.
- [7] Kègl B, Krzyzak A. Piecewise linear skeletonization using principal curves. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 2002; 24: 59-74.
- [8] Tibshirani R. Principal curves revisited, *Statistics and Computing* 1992; 2:183–190.
- [9] Banfield J.D , Raftery A.E. Ice floe identification in satellite images using mathematical morphology and clustering about principal curves. *J. Amer. Statist Ass* 1992; 87:7-16.
- [10] Chang K, Ghosh J. A Unified Model for Probabilistic Principal Surfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 2001; Vol. 23, No. 1.
- [11] Kègl B, Krzyzak A, Linder. T. Learning and design of principal curves. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 2000; 22: 281–297.
- [12] Delicado P. Another look at principal curves and surface. *Journal of Multi-variate analysis* 2001; 7: 84–116.
- [13] Einbeck J, Tutz G, Evers L. Local principal curves. *Statistics and Computing* 2005 15:301–313.
- [14] Verbeek J J, Vlassis N, Krose B. A k-segments algorithm for finding principal curves. *Pattern Recognition Letter* 2002, 23:1009–1017.
- [15] Goldberg D, Korb B, Deb K. Messy genetic algorithms: Motivation, analysis, and first results. *Journal of Complex Systems* 1989; 3:493-530.
- [16] Cha S H, Tappert C. A Genetic Algorithm for Constructing Compact Binary Decision Trees. *Journal of Pattern Recognition Research* 2009; 4:1-13.
- [17] Akbari Z. A multilevel evolutionary algorithm for optimizing numerical functions. *International Journal of Industrial Engineering Computations* 2011; 2:419–430.
- [18] Zhang J, Chung H, Lo W. Clustering-Based Adaptive Crossover and Mutation Probabilities for Genetic Algorithms. *IEEE Transactions on Evolutionary Computation* 2007; 11: 326-335.
- [19] Downs J J , Vogel E. A plant-wide industrial process control problem. *Computers and Chemical Engineering* 1993; 17: 245–255.
- [20] Downs J J , Vogel E. A plant-wide industrial process control problem. *Computers and Chemical Engineering* 1993; 17: 245–255.
- [21] Kresta J. V , MacGregor J F, Marlin T E. Multivariate statistical monitoring of process operating performance. *Can. J. Chem. Engng* 1991; 69:34-47.