ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2015

Identification of Concurrent Control Chart Patterns in Time Series

Shilpy Sharma [1], David A Swayne [1] Charlie Obimbo[1]

Department of Mathematics& Statistics, University of Guelph, 50 Stone Road East, Canada[1]

ABSTRACT: Control chart patterns (CCPs) can be associated with certain assignable causes, and recognition of such patterns can assist the diagnostic search for those causes. Variations could be one or more instances of trend, cyclic, hugging, sudden shift or some other variations over time. Each pattern has special statistical characteristics which differentiate one pattern from another. In a time series, presence of more than one pattern may exist and identification of concurrent pattern is important. In this paper, we will utilize a new approach, RobustICA, for identification of concurrent patterns which is efficient when compared to traditional approaches being used for feature extraction. It will identify independent components hidden in mixture patterns and input those independent components to decision trees for recognition of as many as eight separate control chart patterns.

KEYWORDS: Control chart patterns.

I. INTRODUCTION

Control charts are used as a tool to detect unnatural variations in a time series. In this paper, we are using control charts to detect unnatural variations in hydrological data which can help to improve the water quality. Typically, in water quality data, control charts help to measure if the concentration of chemicals, metals etc. increases or decreases with time. Based on control chart characteristics, a process is considered to have unnatural variations if the points lie outside the $\pm 3\sigma$. There are eight basic control chart patterns: natural, stratification, systematic, cyclic, trend, and shift as shown in Figure 1.

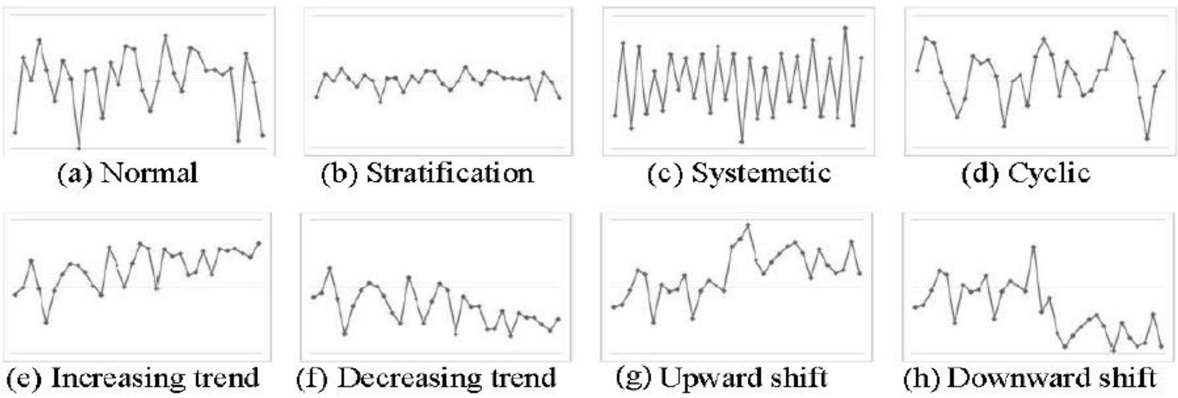

Figure 1: Basic Control Chart Patterns

Most of the existing control chart patterns (CCPs) methods focus on the recognition of one abnormal pattern as shown in Figure 1. However, in hydrological data, there may be situations when more than one unnatural pattern exists, as shown in Figure 2.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2015

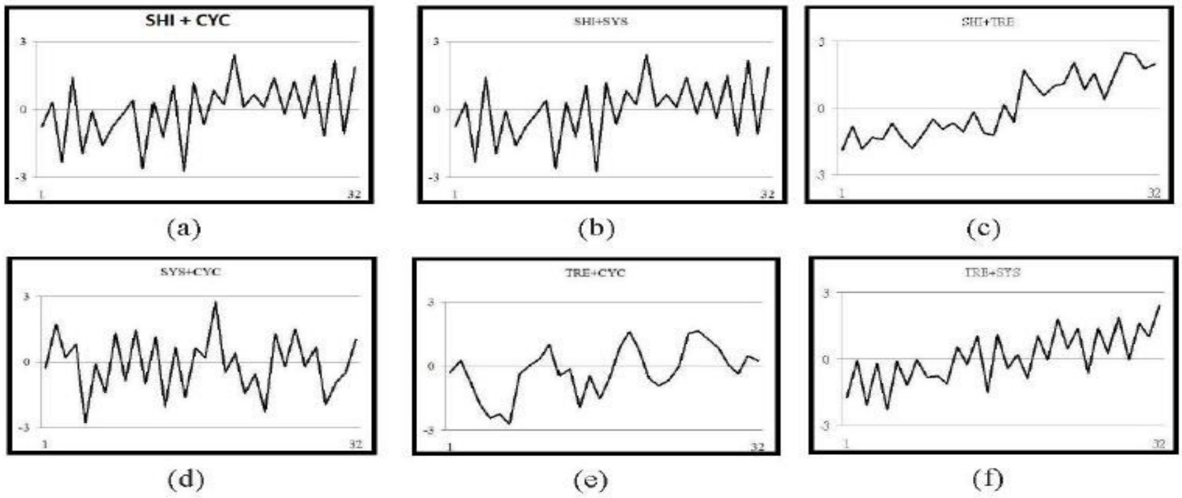

Figure 2: Mixture Patterns

In water quality data situations such as heavy rainfall, snowfall or drought results in inaccurate identification of unnatural patterns. Classifying mixture patterns is a difficult and challenging task. There are very few papers in the literature which discuss identification of concurrent patterns [10,13,14,15,16,22,25]. Accurate identification of concurrent patterns helps in correctly recognizes if the concentration of dissolved parameter is increasing or decreasing over time.

Traditionally, Shewart's rules, Nelson rules, AIAG rules, Boeing AQS rules, and Trietsch rules have been used to interpret control chart patterns manually [17, 18]. However, manual analysis of control chart patterns requires expert knowledge and experience. Otherwise, it can lead to false or inaccurate analysis. Within the last two decades different approaches (artificial intelligence (AI), artificial neural network (ANN), expert systems (ES), support vector machines (SVM), decision trees (DT), hybrid techniques and Bayesian networks (BN)) have been used for automatic detection of control chart patterns [3,4, 5, 6, 7, 8, 11,12,21]. ES were created to overcome the problem of manual techniques but they require explicit rules for pattern recognition. Moreover, ES has problem for false recognition of shift and trend. ANN is considered as a solution to ES problems and majority of work has been done using ANN but it still has drawbacks. Typically, they have complex network topologies and training processes, which are very time-consuming [23]. Although BNs are superior to ANN in the way that they use prior knowledge for analysis, they considered as decision theoretic and do not require a training phase, they require belief updating after every step thereby increasing the complexity of a system. Different authors followed varying approaches based on the model and type of patterns to be discovered. Some authors emphasize raw-based, whereas others focus on feature based approach [11]. Feature based approach is more efficient compared to raw data-based approach [7].

In this paper, we are using a feature based approach for extraction. The feature based approach first extracts features and then uses those features for control chart pattern detection [23]. Different patterns have following characteristics:

1. **Natural/ normal pattern (N):** this pattern shows that behaviour of the natural process and points lies within $\pm 3\sigma$. In this pattern, points fluctuate at random, most of the points are close to the mean and a few points fall near the control limits. The distribution is smooth, and unimodal, neither flat nor skewed.

2. **Sudden Shift (SD or SU):** a sudden shift in the mean level is manifested by a positive or negative change in one direction, where parameter values start to appear on one side of the mean only. Under the new assumed process mean, the process still behaves randomly or naturally. If the two periods are plotted separately, their distribution will be distinct.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2015

3. **Stratification (SR):** it is a form of stable mixture characterized by unnaturally small variations inside the control limits. This pattern appears on the chart as a set of points that closely hug the mean line with absence of points near the limits. The control limits based on the estimated standard deviation are very wide that is the variations from sample to sample are minimal compared to the control limits. The net effect is seemingly reduced variance.

4. **Cyclic (C):** The mixture pattern is formed by points falling near the high and low ends of the chart, with absence of observations near the process mean.

5. **Systematic (SY):** The systematic pattern is formed by points alternating above and below the process mean. It is relatively flat-topped, and typically, has large variations compared to a natural pattern.

6. **Trend up/down (TD or TU):** constant rise (positive direction) or fall (negative direction) in successive points.

The purpose of this pater is to develop a model that can efficiently identify mixture patterns. In this paper, the proposed method recognizes eight basic patterns and mixture patterns with combination of eight basic patterns using feature based approach. We use RobustICA with decision trees to identify mixture patterns. RobustICA generates independent components and decision tree is to identify patterns based on the statistical characteristics.

The paper is organized as follows. Section 2 focuses on method used in this pater. Section 3 focuses methodology for the proposed system. Section 3.4 outlines experimental results and finally, Section 4 focuses on conclusions.

II. MATERIAL AND METHODS

To demonstrate the performance and effectiveness of proposed method, we generated synthetic data with 500 data points. For simplicity of the experiment, we generated 120 data points of mixed patterns; 60 points for stratification pattern and 60 points for increasing trend. This example explains the steps to detect the concurrent patterns from the proposed method. The control chart patterns are generated using the equations in Table 1.

Table 1: [6] Parameters for Control Chart Patterns

Pattern type	Parameters / Values	Pattern Equation
Normal (N)	Mean (μ) = 0 Standard deviation (σ) =1	$y_i = \mu + r_i\sigma$
Stratification (SR)	Random noise(σ') = 0.2σ to 0.4σ	$y_i = \mu + r_i\sigma'$
Systematic(SY)	Systematic Departure(d) = 1σ to 3σ	$y_i = \mu + r_i\sigma + d * (-1)^i$
Cyclic (C)	Amplitude (a)= 1.5σ to 2.5σ Period (T) = 8 and 16	$y_i = \mu + r_i\sigma + a\sin(\frac{2\pi i}{T})$
Trend Up (TU)	Gradient (g)=0.05(σ) to 0.1(σ)	$y_i = \mu + r_i\sigma + ig$
Trend Down(TD)	Gradient (g) = 0.05(σ) to 0.1(σ)	$y_i = \mu + r_i\sigma - ig$
Shift Up (SU)	Shift magnitude(s)=1.5(σ) to 2.5(σ) Shift position (P) = 9, 17,25	$y_i = \mu + r_i\sigma + ks$
Shift Down (SD)	Shift magnitude(s)=1.5(σ) to 2.5(σ) Shift position (P) = 9, 17,25	$y_i = \mu + r_i\sigma - ks$

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2015

III. METHODOLOGY

3.1. RobustICA (Robust Independent Component Analysis)

Independent Component Analysis (ICA) is a novel feature extraction technique and aims at recovering independent sources from their mixtures, without knowing the mixing procedure or any specific knowledge of the sources. The independent components generated from mixture patterns are then served as the independent sources of the mixture patterns. The hidden basic patterns of the mixture patterns could be discovered in these independent components (ICs). An ICA has been successfully applied in various fields of multivariate data procession from image processing, face recognition to time series prediction [14]. The statistical independence of ICs can be measured in terms of their non-Gaussian properties: kurtosis and negentropy [14]. Different version of ICA: ICA, fastICA, and RobustICA are available for analysis. From literature, we found that RobustICA outperforms ICA and FastICA in a way that it eliminates the pre-whitening step and has optimized kurtosis contrast function [26, 27]. Moreover, this method shows high convergence speed. So, we will use RobustICA for our analysis. RobustICA is based on the kurtosis which is optimized by computationally efficient technique based on an optimal step size. Any IC with non-zero kurtosis can be extracted in this manner. No simplifying assumptions concerning specific source type are involved in the application of the algorithm. The methodology behind RobustICA is exact line search, used in numerical optimization field. When compared to other kurtosis based algorithms such as the original FastICA and its variants, the method presents a number of advantages such as: elimination of pre - whitening step which in turn increases the performance. In addition, it is very efficient in real-world problems.

3.2. Decision Tree for pattern recognition

Decision trees are the simplest and most successful method for decision problems [19, 20]. They have been extensively used in different area for analysis due to their various characteristics: simplicity, comprehensibility, ability to handle mixed-type data and using few parameters [19]. A decision tree describes formally the decisions to be made, the events that may occur, and the outcomes associated with combinations of decisions and events. Decision tree models take as input an object or situation described by a set of properties, and as outputs provides yes/no decision. They use “divide and conquer” and “top down” approaches for analysis that can handle a large amount of data in a cost effective way of classification [20]. They start from the root node that represent the classification problem, and split the tree into branches, which represent the discrete value classifier. According to Othman's experimental results [19], DT has more detection accuracy when compared to support vector machine and multilayer perceptron neural networks.

3.3 Proposed Model

Figure 3 shows the proposed model for mixture pattern. The ICs generated by RobustICA are provided to decision trees (DT) for the recognition of independent patterns. The proposed system has following two steps: Firstly, mixture data is applied to RobustICA model to estimate ICs. Secondly, features are extracted from control chart patterns

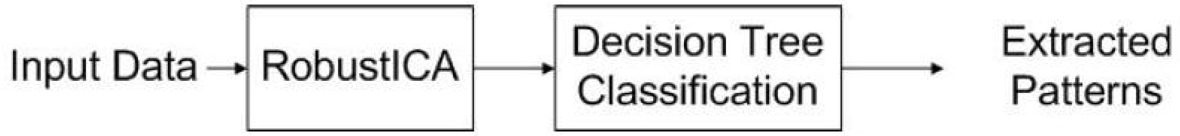

Figure 3: Proposed Model: Concurrent Pattern Recognition

and classification is done using a decision tree. We will skip the correlation analysis, used in the literature, since we are using feature selection methods which have very low correlation among themselves [7]. According to Gauri [7], the performance of feature selection technique decreases if feature selection methods are highly correlated. Based

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2015

on feature characteristics of control chart patterns, four feature candidates for pattern identification are listed below:

- Slope of least square regression and RVE [8] (Ratio between variance of the observation): If the value of slope of least square regression line is 0 then it signifies normal, stratification, systematic, and cyclic pattern otherwise, if greater than 0 then trend up, shift up, trend down and shift down. The positive and negative sign signifies whether the trend/shift is upward or downward. According to Gauri [7], RVE is more powerful compared absolute value and we will use it for our analysis. RVE is able to differentiate Trend up/trend down and shift up/shift down by positive and negative values. A positive value indicates that the trend or shift is upward whereas negative values shows shift or trend is downward.

- Sum of mean regression error [22, 23]: In order to differentiate trend from shift, the trend pattern will have an intermediate mean error while shift pattern will have high mean error. Stratification can be used to differentiate upward trend from upward shift and downward trend from downward shift. Similarly, this rule can be used to differentiate a systematic from a cyclic pattern. For systematic pattern the value is high whereas for cyclic it is low.

- ACLPI defined as Area between the pattern and mean line per interval in terms of SD [8]: to differentiate normal, stratification, cyclic and systematic pattern from each other this method can be applied. If the value is lowest then systematic, intermediate for stratification, normal and Highest for cyclic.

- ALSPI defined as Area between the overall pattern and the LS line per interval in terms of SD2 [8]: to differentiate systematic, normal, and stratification pattern from each other this method can be used. If the value is lowest then systematic, intermediate for normal and highest for stratification.

IV. EXPERIMENTAL RESULTS

For simplicity of the experiment, 120 data points of mixed pattern were generated, 60 points for stratification pattern and 60 data points for increasing trend. This example demonstrates the steps to detect the concurrent patterns from the proposed method. The artificially generated mixed pattern is input to the RobustICA for the generation of ICs. Figure 4 shows the ICs generated by the RobustICA from the input data.

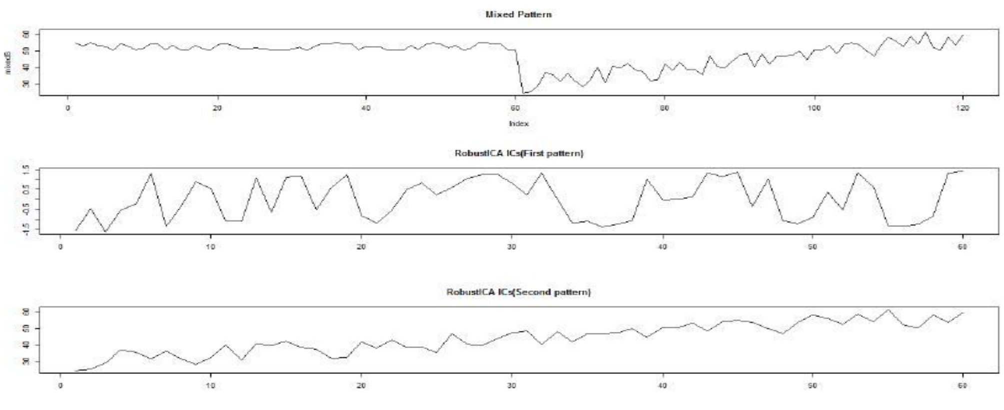

Figure 4: Mixed patterns separated using RobustICA

Based on feature characteristics of the patterns, we started from the root node as shown in Figure 5. In order to differentiate stratification from increasing trend, RVE is used. For stratification, the value for RVE is -0.039. Similarly, the computed value for RVE for increasing trend is 1.648. Inputting the computed values to the DT, RVE values is less than 0 signifies that the pattern is stratification and RVE value for increasing trend pattern is greater than 0.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2015

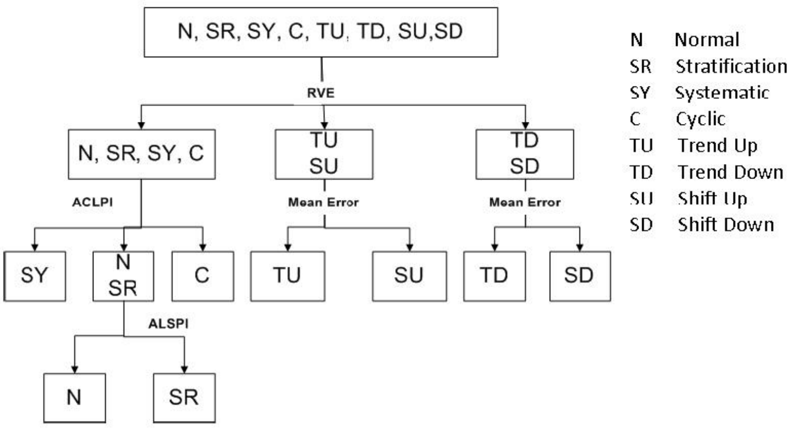

Figure 5: Decision Tree

The pattern recognition experimental success results are shown in Table 2. These are the average results of 100 experiments conducted on 500 data points. The classification tree for feature-based decision tree is shown in Figure 5. Compared to other methods, the proposed method is efficient to detect the ICs ad detect the patterns with accuracy.

Table 2: Pattern Identification, percent success from 100 experiments

Concurrent Pattern	Accuracy Rate	Concurrent Pattern	Accuracy Rate
N + SR	99.50	N + SY	98.20
N + TU	98.16	N+TD	97.78
N+SU	97.80	N+SD	98.95
N+C	100	SR+TU	96.07
SR +TD	97.29	SR +SU	98.54
TU+C	98.99	TU+SU	98.61
TU+SD	98.78	TU+TD	96.48
SY+C	97.23	SY+SD	100
SY+SU	99.92	SY+TD	98.12
SY+TU	96.53	SR+C	97.29
SR+SY	99.22	SR+SD	99.69

V. CONCLUSION

Effective recognition of mixture CCPs in a process is an important and challenging task. In this research, RobustICA is applied along with a decision tree based approach to detect eight possible concurrent control chart patterns. The method first uses RobustICA to generate ICs and then uses feature based decision tree for pattern recognition. From our experimental results, it can be concluded that proposed scheme may efficiently analyzemixture patterns in time-series of medical, financial and any other applications.

REFERENCES

[1] How Fast is FastICA?, Florence, Italy, 2006.

[2] M. Bag, S K. Gauri, S. Chakraborty. “Recognition of control chart patterns using discriminant analysis of shape features”. In Proceedings of the 2010 International Conference on Industrial Engineering and Operations Management Dhaka, Bangladesh. 88-93, 2010.

[3] S. K. Gauri. “Control chart pattern recognition using feature-based learning vector quantization”. Int J AdvManuf Technology, 48:1061-1073, 2010.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2015

[4] S. K. Gauri and S. Chakraborty. “Feature-based recognition of control chart patterns”. Computers & Industrial Engineering, 51:726-742, 2006.

[5] S. K. Gauri and S. Chakraborty. “A study on the various features for effective control chart pattern recognition”. Int J AdvM anuf Technology, 34:385-398, 2007.

[6] S. K. Gauri and S. Chakraborty. “Feature-based recognition of control chart patterns: a generalized approach”. Quality Technology & Quantitative Management, 5:203-220, 2008.

[7] S. K. Gauri and S. Chakraborty. “Improved recognition of control chart patterns using artificial neural networks”. The international Journal of Advanced Manufacturing Technology, 36:1191-1201, 2008.

[8] S. K. Gauri and S. Chakraborty. “Recognition of control chart patterns using improved selection of features”. Computers & Ind ustrial Engineering, 56:1577- 1588, 2009.

[9] N. Golyandina and A. Korobeynikov. Basic singular spectrum analysis and forecasting with r. Computational Statistics & Data Analysis, 71:934-954, 2014.

[10] N. Gu, Z. Cao, L. Xie, D. C. Creighton, M. Tan, and S. “Nahavandi. Identification of concurrent control chart patterns with singular spectrum analysis and learning vector quantization”. J. Intelligent Manufacturing, 24:1241-1252, 2013.

[11] W. Hachicha and A. Ghorbel. “A survey of control-chart pattern-recognition literature (1991-2010) based on a new conceptual classification scheme”. Comput. Ind. Eng., 63:204-222, 2012.

[12] K. Lavangnananda and A. Piyatumrong. “Image processing approach to features extraction in classification of control chart patterns”. IEEE Mid Summer Workshop, pages 85-90, 2005.

[13] C.-J. Lu, Y. E. Shao, and C.-L. Chang. “Applying ica and svm to mixture control chart patterns recognition in a process”. In ISNN (2), volume 6676 of Lecture Notes in Computer Science, 278-287, 2011.

[14] C.-J. Lu, Y. E. Shao, and C.-C. Li. “Recognition of concurrent control chart patterns by integrating ICA and SVM”. Applied Mathematics & Information Sciences, 8:681-689, 2014.

[15] C.-J. Lu, Y. E. Shao, and P.-H. Li. “Mixture control chart patterns recognition using independent component analysis and support vector machine”. Neuro- comput., 74:1908-1914, 2011.

[16] C.-J. Lu, Y. E. Shao, P.-H. Li, and Y.-C. Wang. “Recognizing mixture control chart patterns with independent component analysis and support vector machine”. In L. Zhang, B.-L. Lu, and J. T.-Y. Kwok, editors, ISNN (2), volume 6064 of Lecture Notes in Computer Science, 426-431. 2010.

[17] H. R. Neave and D. J. Wheeler. “Shewhart's charts and the probability approach”. Ninth Annual Conference of the British Demi ng Association, 1-6, 1996.

[18] D. NOSKIEVIÄ.OVÃ . “Analysis of non-random patterns in control chart and its software support”. Metal, 5:23-25, 2012.

[19] Z. Othman and H. F. Eshames. “Abnormal patterns detection in control charts using classification techniques”. International Journal of Advancements in Computing Technology, 4:61-70, 2012.

[20] L. Rokach and O. Maimon.” Data Mining with Decision Trees: Theory and Applications”, chapter 6,. World Scientific Publishing Co., Inc., River Edge, NJ, USA, 71-83,2008.

[21] A. Shaban, M. Shalaby, E. Abdelhafiez, and A. S. Youssef. “Automated identification of basic control charts patterns using neural networks”. Journal Software Engineering and Applications, 3: 208-220, 2010.

[22] C.-H. Wang, T.-P. Dong, and W. Kuo. “A hybrid approach for identification of concurrent control chart patterns”. J. Intelligent Manufacturing, 20: 409- 419, 2009.

[23] C.-H. Wang, R.-S. Guo, M.-H. Chiang, and J.-Y. Wong. “Decision tree based control chart pattern recognition”. International Journal of Production Research, 46: 4889-4901, 2008.

[24] Z. Wu, M. Yao, H. G. Ma, and W. M. Jia. “De-noising mems inertial sensors for low-cost vehicular attitude estimation based on singular spectrum analysis and independent component analysis”. Electronic Letters, 49: 892-893, 2013.

[25] J. Yu. “Pattern recognition of manufacturing process signals using Gaussian mixture models-based recognition systems”. Comput. Ind. Eng., 61: 881- 890, 2011.

[26] V. Zarzoso and P. Comon. “Comparative speed analysis of FastICA”. In M. E. Davies, C. J. James, S. A. Abdallah, and M. D. Plumbley, editors, ICA, volume 4666 of Lecture Notes in Computer Science, 293-300. 2007.

[27] V. Zarzoso and P. Comon. “Robust independent component analysis by iterative maximization of the kurtosis contrast with algeb raic optimal step size”. IEEE Transactions on Neural Networks, 21: 248-261, 2010.