

INTEGRATED PROJECTION AND REGRESSION MODELS FOR MONITORING
MULTIVARIATE AUTOCORRELATED CASCADE PROCESSES

A Dissertation
Submitted to the Graduate Faculty
of the
North Dakota State University
Of Agriculture and Applied Science

By

Anakaorn Khan

In Partial Fulfillment
for the Degree of
DOCTOR OF PHILOSOPHY

Major Department:
Industrial and Manufacturing Engineering

April 2014

Fargo, North Dakota

Title

Integrated Projection and Regression Models for Monitoring
Multivariate Autocorrelated Cascade Processes

By

Anakaorn Khan

The Supervisory Committee certifies that this *disquisition* complies with
North Dakota State University's regulations and meets the accepted
standards for the degree of

DOCTOR OF PHILOSOPHY

SUPERVISORY COMMITTEE:

Dr. Canan Bilen-Green
Chair

Dr. Om Prakash Yadav

Dr. Wichai Chattinnawat

Dr. Rodney D. Traub

Approved:

May 20, 2014

Date

Dr. Canan Bilen-Green

Department Chair

ABSTRACT

This dissertation presents a comprehensive methodology of dual monitoring for the multivariate autocorrelated cascade processes using principal component analysis and regression. Principle Components Analysis is used to alleviate the multicollinearity among input process variables and reduce the dimension of the variables. An integrated principal components selection rule is proposed to reduce the number of input variables. An autoregressive time series model is used and imposed on the time correlated output variable which depends on many multicorrelated process input variables. A generalized least squares principal component regression is used to describe the relationship between product and process variables under the autoregressive regression error model. The combined residual based EWMA control chart, applied to the product characteristics, and the MEWMA control charts applied to the multivariate autocorrelated cascade process characteristics, are proposed.

The dual EWMA and MEWMA control chart has advantage and capability over the conventional residual type control chart applied to the residuals of the principal component regression by monitoring both product and the process characteristics simultaneously. The EWMA control chart is used to increase the detection performance, especially in the case of small mean shifts. The MEWMA is applied to the selected set of variables from the first principal component with the aim of increasing the sensitivity in detecting process failures. The dual implementation control chart for product and process characteristics enhances both the detection and the prediction performance of the monitoring system of the multivariate autocorrelated cascade processes. The proposed methodology is demonstrated through an example of the sugar-beet pulp drying process. A general guideline for controlling multivariate autocorrelated processes is also developed.

ACKNOWLEDGEMENTS

I would like to truly express gratitude to my advisor, Dr. Canan Bilen-Green, for her guidance, patience, and encouragement during my study at NDSU. Without her supervision this work would not be accomplished. I also would like to also thank Dr. Wichai Chattinnawat his support and assistance in overcoming difficulties in the timely completion of this dissertation.

I would like to acknowledge the rest of my committee members, Dr. Om Prakash Yadav and Dr. Rodney Traub, for their comments and suggestions. I would like to thank the faculty and staff of Industrial and Manufacturing Engineering during my entire doctoral study at NDSU.

Finally, I am most thankful to unconditional love from my family. It was their help, support and understanding that allowed me to be able to complete my study.

TABLE OF CONTENTS

ABSTRACT.....	iii
ACKNOWLEDGEMENT	iv
LIST OF TABLES	viii
LIST OF FIGURES	ix
LIST OF APPENDIX TABLES	x
LIST OF APPENDIX FIGURES.....	xi
CHAPTER 1. INTRODUCTION	1
1.1. Background.....	1
1.2. Significant of the Problem	2
1.3. Problem Statement	3
1.4. Overview of Dissertation	4
CHAPTER 2. LITERATURE REVIEW	6
2.1. Multivariate Process Monitoring and Control	6
2.1.1. Shewhart-type control chart.....	6
2.1.2. Cumulative sum control chart (CUSUM)	7
2.1.3. Exponentially weighted moving average based control chart (EWMA)	8
2.1.4. Multivariate control using principal components analysis (PCA).....	9
2.1.5. Alternative methods	10
2.2. Multivariate Autocorrelated Process Monitoring and Control	11
2.2.1. Vector autoregressive (VAR) and support vector regression (SVR) model.....	11
2.2.2. Multivariate autocorrelated using principal components analysis.....	12
2.2.3. Multivariate autocorrelated using partial least square regression.....	13
2.2.4. State-space model and neural networks model.....	14
2.2.5. Other methods	15

CHAPTER 3. PAPER 1 PRINCIPAL COMPONENTS REGRESSION CONTROL FOR MULTIVARIATE AUTOCORRELATED CASECADE PROCESSES.....	17
3.1. Introduction.....	17
3.2. Principal Components Autoregressive Multiple Regression	20
3.2.1. Multiple regression and autoregressive error correction model	21
3.2.2. Principal components and selection rule.....	22
3.2.3. Interpretation of principal components	25
3.3. Application in Sugar Beets Pulp Drying Process	27
3.4. Discussion	36
CHAPTER 4. PAPER 2 DUAL MONITORING SCHEME FOR MULTIVARIATE AUTOCORRELAED CASCADE PROCESSES WITH EWMA AND MEWMA CONTROL CHARTS	38
4.1. Introduction.....	38
4.2. Principal Components Regression Based on Generalized Least Square under Autoregressive Errors	42
4.2.1. Principal components regression	42
4.2.2. Selecting significant principal components	44
4.2.3. Multiple generalized least squares regression on significant principal components under autoregressive error correction model	45
4.2.4. Multivariate exponential weighted moving average (MEWMA) control chart on process inputs.....	48
4.3. Dual EWMA and MEWMA Control Chart Diagram	52
4.4. Implementation of the Dual Control Scheme	53
4.5. Performance of the Proposed Dual Scheme.....	58
4.6. Summary and Discussion.....	67
CHAPTER 5. CONCLUSIONS AND RECOMMENDATIONS	70
5.1. Conclusion	70
5.2. Discussion and Future Research Direction	71

REFERENCES	74
APPENDIX A. DRY PULP PROCESS DATA	81
APPENDIX B. SIMULATION PROGRAM (MATLAB PROGRAM).....	88
B.1. Main Simulation.....	88
B.2. Average Run Length (ARL) of EWMA and MEWMA.....	93
B.3. Theoretical Covariance of Y	95
B.4. Generate AR(3)	96
APPENDIX C. MULTIPLE REGRESSION CONTROL CHART INTEGRATED WITH PRINCIPAL COMPONENT ANALYSIS	97
C.1. Introduction	97
C.2. Methodology	97
C.2.1. Multiple regression model.....	98
C.2.2. Standardizing.....	98
C.2.3. Principal components analysis	99
C.3. Application.....	100
C.4. Conclusioin.....	105
C.5. References	105
APPENDIX D. MULTIVARIATE AUTOCORRELATED PROCESS MONITORING WITH MULTIPLE REGRESSION AND PRINCIPAL COMPONENTS.....	107
D.1. Introduction.....	107
D.2. Description of the Method	109
D.3. Example	113
D.4. Conclusion	117
D.5. References.....	118

LIST OF TABLES

<u>Table</u>	<u>Page</u>
3.1. Eigen values of the correlation matrix	32
3.2. Correlation between principal components and standardized variables	33
3.3. Simplified correlation between principal components and standardized variables	34
3.4. Adjusted weight of each variable (X) standardized variables for each principal component and overall weight of each variable (X) to reponse variables (Y)	35
4.1. Eigenvalues of the correlation matrix	56
4.2. Correlation between principal components and standardized variables	57
4.3. Scenarios for studying the dual control scheme performances.....	59
4.4. Performances of the dual scheme under scenario 1	60
4.5. Performances of the dual scheme under scenario 2	61
4.6. Performances of the dual scheme under scenario 3	62
4.7. Performances of the dual scheme under scenario 4	63
4.8. Performances of the dual scheme under scenario 5	64
4.9. Performances of the dual scheme under scenario 6	65
4.10. Performances of the dual scheme under scenario 7	66

LIST OF FIGURES

<u>Figure</u>	<u>Page</u>
3.1. Flow chart for analyzing multivariate autocorrelated processes.....	27
3.2. Pulp drying process in general	29
3.3. Pulp drying process according to time.....	29
3.4. Shewhart control chart for dry pulp moisture residuals.....	36
4.1. Pulp drying process in general	38
4.2. Pulp drying process according to time.....	39
4.3. Autoregressive error correction parameters estimation procedure	50
4.4. The general configuration of dual EWMA and MEWMA control chart scheme.....	51
4.5. Time series plot of moisture content (output) for the pulp drying process.....	57

LIST OF APPENDIX TABLES

<u>Table</u>	<u>Page</u>
C.1. Eigenvalues of the correlation matrix	103
C.2. Correlation between principal components and standardized variables	104
D.1. Simplified correlation between principal components and standardized variables.....	115

LIST OF APPENDIX FIGURES

<u>Figure</u>	<u>Page</u>
C.1. Pulp drying process according to time	100
C.2. Flow chart for analyzing multivariate processes	106
D.1. Process description.....	109
D.2. Shewhart control chart for dry pulp moisture residuals	116

CHAPTER 1. INTRODUCTION

1.1. Background

Recent technical advancements in production industry make it possible to collect several process measurements at the same time. The need to simultaneously monitor several quality characteristics in order to capture the multivariate nature of production processes is gaining importance for improving such processes (Bersimis et al., 2007; Jackson, 1985; Woodall and Montgomery, 2012). Multivariate statistical process control involves simultaneously monitoring processes with two or more quality characteristics that interact with one another. In multivariate processes, change in one quality characteristic can affect other quality characteristics. Therefore, multivariate control chart procedures have to take correlation of the quality characteristics into consideration. Observations from such processes are assumed to be independent over time.

Standard multivariate process control procedures are not effective when the data is serially correlated. This type of process is called “multivariate autocorrelated processes” and is especially common in chemical and agricultural industries. The higher false alarm rate due to autocorrelation effect encountered when monitoring such processes has been discussed extensively (Lowry and Montgomery, 1995; Mason and Young, 2007). Holgersson (2004) discussed several techniques to determine the time dependency of the multivariate processes. The cause of autocorrelation effect can vary according to the data collection system and the nature of the process. For example, in the chemical industry, the process data is usually autocorrelated due to the process inertia. Because of the complexity of monitoring multivariate autocorrelated processes, development of simple, robust, effective procedures is crucial and has gained interest of researchers in statistical process control.

1.2. Significance of the Problem

The challenge of monitoring multivariate and autocorrelated processes can be seen in the pulp drying process in the sugar factory, the pulp drying process produces the dry pulp which is the by-product of sugar beet processing. The dry pulp is sold as a cattle food to the farmers. The moisture content of the dry pulp is one important quality variable that needs to be monitored constantly. If the moisture content of dry pulp does not meet the target value, it affects the price and quality of the dry pulp plus the cost of the transportation to the buyers. The pulp drying process consists of a large number of input variables which are correlated over time with cross-correlation among variables. These time series input variables are combination of both controllable and uncontrollable variables. Process operators monitor the moisture content of the dry pulp and when needed adjust controllable input variables to achieve the target value. The simple pulp drying process monitoring is neither effective, nor accurate as it does not take into account the autocorrelated and multicollinear nature of the input variables. These input variables directly affect the quality of the dry pulp which is also serially correlated over time according to process inertia. This type of process is sometimes referred to as cascade or model-fixed process in the literature (Hawkins, 1993; Loredó et al., 2002). Such processes are common in other types of production industries. A crucial assumption of traditional multivariate process control procedures is that the data is independent and identically distributed with constant mean vectors μ_x and covariance matrix Σ_x . Autocorrelated observations violate this assumption and generate far more false alarms in multivariate process control (Mastrangelo and Forrest, 2002; Noorossana and Vaghefi, 2006). Therefore, traditional multivariate control charts are not effective in monitoring multivariate autocorrelated processes.

Furthermore, in some cases, it is useful to use the relationship between input variables and the output variable in monitoring processes (Woodall et al. 2004). Normally, the simple

linear regression method can be applied to identify the relationship of the process, but not in a multivariate framework. The multivariate process control typically exhibits multicollinearity due to large number of variables in the process. Multicollinearity has an effect on the least-square estimation of the regression coefficients. Thus, direct linear regression analysis is not applicable and generates more false alarms on the control chart. Another major concern is monitoring a high dimensional system. As more variables are involved in multivariate processes, the use of traditional multivariate control chart schemes are not practical to any further extent (Bersimis et al. 2007). Both the multivariate and autocorrelated nature should be tackled at the same time for effective monitoring method such process.

1.3. Problem Statement

This dissertation addresses the problem of monitoring and controlling multivariate autocorrelated cascade processes. The traditional multivariate control chart is not able to handle the multivariate autocorrelated cascade processes due to autocorrelation effect, collinearity and high dimensionality of the system. Autocorrelation effect violates the assumption of independent observations in the traditional multivariate control chart. Multicollinearity is the situation where two or more variables in a multiple regression model are highly correlated, thus the simple linear regression is not capable of obtaining the correct model. Besides, the traditional multivariate control chart is not efficient in dealing with high dimension processes. In this dissertation, multivariate autocorrelated processes with multiple times series input variables and single time series output variable is modeled using the multiple regression method based on principal components. This model is used to the predict output variable, affected by several input variables, and is also used to develop a multivariate control chart for process monitoring.

The Principal Components Analysis (PCA) is applied to alleviate the multicollinearity problem among input variables and reduce the dimension of the input variables before fitting

the multiple regression model. The parameters of the multiple regression model with selected principal components are estimated based on the both ordinary least squares (OLS) method and generalized least squares (GLS) method. An integrated selection rule is developed to select important principal components. The residuals of the multiple regression model on the time series structure data are usually in time series structure violating the independent errors assumption. Thus the autoregressive error correction model is used to address the issue of dependent residuals. The interpretation of out-of-control signals of the process output on the control chart is also discussed. Additionally, this dissertation studied the performance of two types of control charts utilized in the proposed method: (1) Exponential Weighted Moving Average (EWMA) control chart in monitoring the residuals of the regression model; (2) Multivariate Weighted Moving Average (MEWMA) control chart on significant variables from the first Principal Component (PC). A simulation study based on average run length (ARL) for various sizes of mean and covariance vector shifts along with the change in regression parameters is carried out to investigate the effectiveness of the proposed method.

1.4. Overview of the Dissertation

This dissertation developed an overall approach to monitor multivariate autocorrelated cascade processes. The work in this presentation was presented in a number of conferences and papers:

- 1) Bilen, C., Chen, X., Khan, A. and Yadav, O.P. (2007) ‘Multiple regression control chart integrated with principal component analysis’, *Proceedings of the Industrial Engineering Research Conference*, 19–23 May, Nashville, Tennessee.
- 2) Bilen, C., Khan, A. and Yadav, O.P. (2009) ‘Multivariate autocorrelated process monitoring with multiple regression and principal components’, *Proceedings of the Industrial Engineering Research Conference*, 30 May–3 June, Miami, Florida.

- 3) Bilen, C., Khan, A., and Yadav, O. P. (2010) 'Principal components regression control for multivariate autocorrelated cascade processes', International Journal of Quality Engineering and Technology, Vol. 1, No. 3, pp.301-315.
- 4) Dual SPC schemes for multivariate autocorrelated cascade processes with combined residual-based EWMA and MEWMA. (Under reviewed)

This dissertation is divided into five chapters. After this introduction, Chapter 2 gives an extensive literature review of multivariate statistical process control framework this dissertation. Chapter 3 presents an overall complete methodology for monitoring the multivariate autocorrelated process using Shewhart-type residual-based control chart in conjunctions with the principal component regression. Dual EWMA and MEWMA control chart performance based on proposed model are displayed and discussed in Chapter 4 along with simulation studies based on average run length (ARL) for various sizes of mean vector shifts for input variables are investigated. The sugar beet pulp drying process is used as an example for the proposed method application. The conclusion and recommendations of future research is provided in the last Chapter 5.

CHAPTER 2. LITERATURE REVIEW

2.1. Multivariate Process Monitoring and Control

2.1.1. Shewhart-type control chart

Multivariate quality control method was originally introduced by Hotelling (1947). The method was developed by using the generalized Student ratio T under the assumption of multivariate normal distribution and was applied to the air testing of sample bombsights problem. The Hotelling T^2 multivariate Shewhart control chart is used for unknown mean vector and unknown variance-covariance matrix and X^2 control chart is use in the case of known parameters. Aparis (1996) proposed using Hotelling T^2 control chart with adaptive sample sizes to improve its performance in detecting small process shifts. Average run lengths (ARL) were calculated to compare the power of T^2 control chart with fixed sample size and adaptive sample size for different sizes of process mean shift. The result showed that the T^2 control chart with adaptive sample size works well when the shift in the mean is small. In the case of individual observation, T^2 control chart is not effective in detecting the mean shift because the sample covariance matrix is used to estimate the covariance matrix (Sullivan and Woodall, 1996). They compared the control chart using several techniques for estimating the covariance matrix when there are step and trend shifts in the mean vector. The T^2 control chart with the estimation of covariance matrix from the vector differences between successive observations is more effective than standard T^2 control chart in detecting step and trend shifts.

Nedumaran and Pignatiello (2000) investigated the problem of T^2 control chart during Phase I retrospective testing and described the exact method of constructing T^2 control chart limits to control overall probability of a false alarm at a specified value. Mason et al. (2001) applied Hotelling T^2 control chart with some adaptations to monitor the mean vector of the batch processes in both Phase I and Phase II stages. They emphasized the importance of outlier removal, parameter estimation and the location of significant batch mean differences in

implementing Hotelling T^2 control chart for batch processes. See Nomikos and MacGregor (1995) for more details on batch processes. Kim et al. (2003) proposed the quality control chart method to monitor a linear function process or product. The linear function process or product described by a relationship between two or more variables instead of the distribution of single quality characteristic. They recommended the use of bivariate T^2 control chart in conjunction with a univariate Shewhart control chart for Phase I and the use of three univariate control charts in Phase II for a linear function process.

2.1.2. Cumulative sum control chart (CUSUM)

Cumulative sum (CUSUM) control chart was developed to overcome the issue of Shewhart-type control chart in detecting small and moderate shifts of the mean vector. Woodall and Ncube (1985) introduced a multiple univariate cumulative sum which is known as MCUSUM to use in the multivariate case. They explored the MCUSUM in the situations where the quality characteristics are independent and dependent to each other. It was shown that the MCUSUM is favored over Hotelling T^2 control chart in the bivariate case as the correlation among quality characteristics does not affect the performance of MCUSUM as much as Hotelling T^2 control chart. Healy (1987) described the procedure of detecting the mean shift in univariate CUSUM as sequential probability ratio tests to extend to the multivariate case. He also applied the concept of CUSUM in detecting the shifts of mean vector and covariance matrix for multivariate normal distribution. Hawkins (1991) proposed the CUSUM for monitoring the residual vector from regression-adjusted variables because Hotelling T^2 does not function well when there are structure mean shifts and is not easily interpretable. Later, he also recommended using the same CUSUM procedure for cascade process control by monitoring individual residual from linear regression of any variables on the rest of other variables (Hawkins, 1993). These charts of individual residual directly identify the variable that causes the out-of-control signal. According to many proposed MCUSUM methods, Runger

and Testik (2004) discussed pros and cons of several CUSUM extension methods for the multivariate case and evaluated their performances. He classified each method in term of their geometric shapes and the names were given to the method based on these geometric characteristics.

2.1.3. Exponentially weighted moving average based control charts (EWMA)

The Exponentially weighted moving average (EWMA) control chart is developed for the same purpose as the Cumulative Sum (CUSUM) control chart. The multivariate extension of EWMA control chart (MEWMA) along with guidelines for designing multivariate process control chart was presented by Lowry et al. (1992). The ARL was used to compare the performance evaluation of the MEWMA control chart with Hotelling T^2 and MCUSUM charts. The ARL simulation results indicated that MEWMA performs better than MCUSUM when the shift happens before the application of the chart. When the shift occurs after the process is in control, the MCUSUM and MEWMA performs equally well. In order to assist in MEWMA control chart implementation, a program for estimating the in-control ARL for the MEWMA control chart was developed by Bodden and Rigdon (1999). In addition, Molnau et al. (2001) built an ARL calculation program for MEWMA control chart when the shift size, control limits and the smoothing parameter are known. Prahu and Runger (1997) provided suggestions for selecting parameters for MEWMA control chart based on the ARL and distribution of the run length acquired from Markov chain analysis. Their simulation study indicates that MEWMA smoothing parameter values between 0.1 and 0.5 are best in general.

Runger et al. (1999) proposed the U-transformation scheme to convert the original variables to lower dimensional subspace to improve the performance of MEWMA in detecting shifts in mean as the number of variables increase. The U-transformation is comparable to principle components excluding its ability to define the specific subspace for monitoring. The effect of non-normality on the MEWMA control chart performance in monitoring individual

observations was investigated by Stoumbos and Sullivan (2002). They verified that the Hotelling X^2 control chart is very sensitive to non-normality and the MEWMA is more robust than Hotelling X^2 control chart. Also, the extension of EWMA to subgroups of any size is possible. This extension can detect any shift size and direction effectively and robust to non-normality as well. Khoo (2003) suggested the exact covariance matrix calculation of MEWMA instead of asymptotic form to enhance the sensitivity of the chart in detecting small shifts. The asymptotic form covariance matrix was used in general when constructing the EWMA control chart, but this form takes time to initiate the control scheme during the start-up period and results in late detection of an out-of-control condition. Furthermore, Hawkins et al. (2007) proposed the full smoothing matrix of MEWMA control chart instead of using traditional diagonal matrix to improve the sensitivity of the chart at the start-up stage. The MEWMA control chart outperforms traditional EWMA in detecting the mean vector shifts.

2.1.4. Multivariate control using principal components analysis (PCA)

The principal components analysis (PCA) is a method for reducing the dimensionality of quality characteristics in multivariate processes to a new set of variables called principal components (Jolliffe, 2002). Each principal component is a linear combination of all original variables. The first principal component has the highest variance of all coefficient vectors of unit length. The second principal component has less variance and is orthogonal to the first principal component. The rest of principal components are listed in the same manner and the last principal component has the least variation. Schall and Chandra (1987) developed a method of monitoring manufacturing processes by using principal components with multivariate regression analysis. This process involves input and output variables when the number of input variables is larger than the number of output variables. The quality of the output variables depends on the quality of the input variables. Therefore the relationship between input and output variables is established by multiple regression analysis while the principal components

analysis is used to reduce the number of input and output variables. The control limits were calculated based on the Shewhart control chart for monitoring each principal component separately.

Runger (1996) applied the principal components analysis for designing a multivariate control chart that is sensitive to certain assignable cause and model-fixed assignable causes. The U^2 control chart is developed based on chi-square chart of U which is the vector of the first set of principal component with the most variance explained from original set of variables. The U^2 control chart is studied and discovered to be better, in detecting the mean vector shift, than X^2 control chart. Wikstrom et al. (1998) analyzed the electrolysis process using multivariate statistical process control. The principal components analysis was utilized to get principal components and residuals from the linear combination model. The principal components and residuals were charted on traditional multivariate control charts to create a simultaneous principal components monitoring and residual tracking control chart (SMART). The SMART chart is appeared to be more sensitive to mean vector shift than the traditional multivariate control charts.

2.1.5. Alternative methods

Sullivan and Woodall (2000) proposed the retrospective phase I multivariate control chart to detect mean vector shift or covariance matrix shift for individual observations data. The chart is based on the likelihood ratio test of the mean vector and covariance matrix shift. In retrospective stage, the parameters are estimated and this estimation is poor when multiple outliers exist. The control chart is more effective in detecting shift than Hotelling T^2 control chart. Feltz and Shiau (2001) used an empirical Bayesian approach to build the multivariate control chart. In this chart, the multivariate normal distribution with mean is derived from the posterior distribution of the Bayesian theory and the process parameters are estimated to apply

to multivariate processes structure. Vargus (2003) presented the Hotelling T^2 control chart for retrospective phase I based on minimum volume ellipsoid (MVE) estimators of process mean vector and covariance matrix. The method is proved to be more robust than the standard technique to estimate mean vector and covariance matrix for Hotelling T^2 control chart. Aparisi et al. (2004) invented the supplementary run rules to accompany X^2 control chart to enhance the performance in detecting small or moderate mean vector shifts. The X^2 control chart with run rules has better performance than X^2 control chart without run rules with respect to ARL performance for various shift sizes. The performance of X^2 control chart with run rules are further improved by Koutras et al. (2006) by incorporating two run rules into the modification of X^2 control chart.

2.2. Multivariate Autocorrelated Process Monitoring and Control

2.2.1. Vector autoregressive (VAR) and support vector regression (SVR) Model

The vector autoregressive (VAR) model for multivariate autocorrelated data is analogous to univariate autocorrelated data. The data can be modeled as a time series and the residuals of the model can be monitored on the traditional multivariate control chart. Kalgonda and Kulkarni (2004) monitored the mean vector of autocorrelated process by fitting the first-order vector autoregressive model VAR(1) and proposed the Z-chart based on the single step finite intersection test. The proposed control chart has the ability to identify the variable causing the out-of-control signal. In addition, Noorossana and Vaghefi (2006) investigated the performance of MCUSUM chart when data exhibits autocorrelation using ARL as an assessment for various mean shifts. They showed that the method of multivariate time series model improves the performance of MCUSUM in detecting the mean vector shifts.

The VAR control chart was proposed by Jarrett and Pan (2007) for monitoring Hotelling T^2 statistics. The VAR chart is the integration of univariate residuals control chart

with multivariate Hotelling T^2 control chart. The effect of parameter shifts which consist of mean shift, covariance shift and autocorrelation structure of the model shift are also examined. The mean shift effect depends on the shift size and the estimation of VAR parameter matrices. The covariance matrix shift and autocorrelated structure of the model shift also gives an out-of-control signal on the VAR control chart. Issam and Mohamed (2008) applied support vector regression (SVR) method to get residuals and monitored on MCUSUM control chart. SVR is an extension of support vector machine (SVM) algorithm with linear regression technique (Scholkopf and Smola, 1998). The SVR-based control chart is discovered to be sensitive to small shifts in the mean vector.

2.2.2. Multivariate autocorrelated using principal components analysis

Ku et al. (1995) proposed a dynamic principal components analysis (DPCA) model for monitoring multivariate processes when variables are correlated in time. The number of time lag in the variables and principal components need to be identified before applying the DPCA method. The data matrix is formed based on the time lag information and the principal components are determined for establishing the dynamic linear relationship. Then, the T^2 statistics with control limits are calculated for process monitoring. Mastrangelo et al. (1996) reviewed several multivariate process monitoring methods and demonstrated the effectiveness of principal component analysis when the data has serial correlation over time. They applied the principal components analysis to the process with different time series models of seven variables. Once the optimal number of principal components is determined, the values of the original variables are substituted back into the linear combination equations to obtain the z-scores. The z-scores are plotted on the chart for monitoring the process shifts. Runger (1996) also explored the principal component analysis method for autocorrelated and cross-correlated process. The uncorrelated residuals vectors based on principal components are computed and are monitored on traditional multivariate control chart.

Wikstrom et al. (1998) applied multivariate time series analysis and principal components analysis to electrolysis process. The original eight variables in the electrolysis process are represented by two uncorrelated principal components. These two components are investigated using autoregressive integrated moving average (ARIMA) model and the model coefficients are estimated by using the partial least square regression. The univariate control chart for monitoring are chosen based on the category of multivariate time series model. Tsung (1999) developed the adaptive principal component monitoring (APCM) for the automatic-controlled process. The principal components extracted from the process input-output matrix are combined with the knowledge of control actions to design statistical process control. The input-output combination and the residuals are monitored simultaneously on different charts for process change detection. Later, Kano et al. (2001) applied principal component analysis to on-line process monitoring by moving the time-window. This approach is referred to as moving principal component analysis (MPCA). The directions of principal components are used to monitor a change in the correlation structure of process variables. Both dynamic principal components developed by Tsung (1999) and Kano et al. (2001) are more effective for monitoring dynamic processes than the static principal components.

2.2.3. Multivariate autocorrelated using partial least square regression

Partial least square regression was introduced by Horst (1961) to find the highest possible correlation between two sets of variables. Then the second highest pair of correlation is retrieved and it is orthogonal to the first pair. This procedure is continued until only a pair from one set of variables left. Kresta et al. (1991) and Macgregor et al. (1994) presented the basic partial least square methodology in monitoring continuous processes. Kourti et al. (1996) and Martin et al. (1999) showed the practical application of partial least square technique for continuous industrial process monitoring. The process fault detection and diagnosis are also discussed for controlling product quality. The performance of multivariate process monitoring

based on partial least square was demonstrated by Simoglou et al. (2000). The partial least square performance is quite comparable to canonical variate analysis in their ability to predict quality variables. Wang et al. (2003) proposed the recursive partial least square algorithm to monitor the non-stationary and time varying processes. This application of recursive partial least square algorithm reduces the false alarm better than traditional multivariate statistical process control framework. Kourti (2005) gave the outline application of partial least square in monitoring different multivariate process situations such as continuous process and missing process data.

2.2.4. State-space model and neural networks model

Triantafyllopoulos (2006) proposed the Bayesian state space model to form a one-step forecast distribution and obtained Bayes' factor of the process. The Bayes' factor which is the distance between the one-step forecast mean and target mean vector are monitored using the modified univariate EWMA control chart. In addition to being robust to non-normality, this approach can control mean vector and covariance matrix at the same time. Arkat et al. (2007) designed residuals control chart based on the artificial neural network-based model for AR(1) process. The residuals from the model were plotted on MCUSUM for process monitoring. The proposed scheme performance outperforms standard MCUSUM control chart and is more sensitive to small shifts than the time series-based control chart. Another state space model based control chart is developed by Vargas et al. (2009) for the case of strong cross correlation among variables. The state space model is used to estimate covariance matrix and the Hotelling T^2 control chart with control limits is built on this estimation of covariance matrix. This method uses fewer number of parameters estimations and is able to get independent residuals by eliminating the cross-correlation effect. Hwang and Wang (2010) proposed a neural network based identifier (NNI) control chart with additional capability of identifying the variable responsible for the out-of control signal. The NNI control chart does

not require model fitting or parameter estimation and the performance is superior to Hotelling T^2 control chart, EWMA chart and Z-chart for detecting small to moderate shifts.

2.2.5. Other methods

Chan and Li (1994) proposed the M-chart which is an extension of multivariate Shewhart control charts for monitoring linear trends of the time series process using dimension reduction approach. The dimension reduction approach is developed by using projection pursuit technique which is similar to principal component analysis except the original data structure information is still maintained after the projection of the data to lower dimension. Loredó et al. (2002) extended the regression adjustment method of Hawkins (1993) for autocorrelated data. The residuals from the multiple regression model are monitored on the univariate traditional Shewhart, CUSUM and EWMA. The residual-based control chart performance is compared with original observation-based control chart by using ARL when the variables are autocorrelated. The residual-based control chart is more sensitive to mean shift than the original observation-based control chart under the assumption of correct model.

Badcock et al. (2004) proposed two methods based on transforming original time series variables into a linear combination of new set of uncorrelated components. This method focuses on developing the components with specific correlation properties, unlike the principal components analysis that obtains the components according to their maximum variance. The first method from Badcock et al. (2004) modified the signal-noise decomposition method in image analysis to get different autocorrelation structure of components to describe the behavior of multivariate process. The second method applied the concept of principal components analysis to find the temporally uncorrelated components for all time lags. Jiang (2004) proposed multivariate autocorrelated process control chart by using a generalized likelihood ratio test (GLRT). The paper also compared the performance of T^2 control chart, observational

GLRT control chart and residuals GLRT control chart in monitoring multivariate vector from univariate autocorrelated process transformation.

CHAPTER 3. PAPER 1 PRINCIPAL COMPONENTS REGRESSION CONTROL FOR MULTIVARIATE AUTOCORRELATED CASCADE PROCESSES¹

3.1. Introduction

Most of the research work in statistical process control is widely concentrated on monitoring the output variables of the process constantly on the control chart. If an out-of-control signal is detected, the root cause of the signal is identified and resolved. Moreover, in some cases, the quality of process is described by a relationship between the response variable and predictor variables. If there is any change or interruption in predictor variables, the response variable is directly affected and results in out-of-control signals on the control chart. This type of process is sometimes referred to as a cascade process (Hawkins, 1993). Such a process often involves several related input variables and output variables resulting in multicollinearity. Further, in some industries, these variables are often time series data, and autocorrelation exists in both input and output variables. Besides, as a result of increased sampling frequency capabilities in monitoring processes, process data are typically autocorrelated. The traditional multivariate control chart assumes that process observations are statistically independent and, therefore, does not capture the relationship between input variables and the output variable.

Furthermore, the investigation of the root cause of the out-of-control signal for the traditional multivariate control chart is complicated. As a result, a traditional multivariate control chart is not a very effective approach for monitoring high dimensional, multicollinear and autocorrelated systems. What is really needed is the development of a comprehensive approach for monitoring multivariate processes with both multicollinearity and

¹ Content in this chapter was previously published in International Journal of Quality Engineering and Technology, 1(3), 301-316 and was written by Canan Bilen-Green, Anakaorn Khan and Om Prakash Yadav. Anakaorn Khan was responsible in writing the draft of the paper and analyzing data.

autocorrelation, which can simultaneously capture the relationship between input variables and the output variable, so the root causes of out-of-control signals can also be identified. Some researchers have proposed the multivariate control chart for monitoring process output without taking the input variables into account (Jarrett and Pan, 2007; Jiang, 2004; Kalgonda and Kulkarni, 2004). The complexity of the multivariate autocorrelated problem is a challenging issue in statistical process control due to difficulties in identification of not only location but also root cause of out-of-control signals (Bersimis et al., 2007; Lowry and Montgomery, 1995; Woodall and Montgomery, 1999).

Approaches for monitoring multivariate autocorrelated processes when quality characteristics are directly affected by input variables can be divided into two categories. The first approach, initially developed by DiPaola (1945), is based on the regression control chart concept. The regression control chart is an integration of linear regression and control charting to monitor a quality characteristic of interest. Mandel (1969) combined the regression analysis and control chart to monitor man hours in the post office department. The regression line is established for the relationship between the man hours used and the volume of mail handled, and also served as the centreline with a standard error of estimate for upper and lower limits. This regression control chart is used as performance measurement and control tool in detecting assignable causes for process improvement. Hawkins (1991) used the idea of regression adjustment in the multivariate case by maximizing the relationship of related variables in order to increase the sensitivity of process monitoring. He proposed monitoring the process with Shewhart and CUSUM control charts based on the vector scaled residuals from the regression of each variable on all others. Hawkins (1993) extended the regression adjustment method for monitoring cascade processes. Loredó et al. (2002) developed an extension of regression adjustment method for multivariate quality control for autoregressive and correlated processes, especially for cascade processes with autocorrelated variables.

The second approach to monitoring autocorrelated processes is based on principal component analysis (PCA) introduced by Jackson (1980). Jolliffe (1972, 1973) illustrated numerous considerations and compared several principal components selection rules for real and artificial data. Bisgaard and Kulahci (2005) demonstrated the application of PCA for multivariate process monitoring. Schall and Chandra (1987) recommended using PCA for multivariate processes, when the number of input variables is less than the number of output variables, for handling the multicollinearity effect and simultaneously reducing the number of variables. A review of methodologies for multivariate autocorrelated process monitoring and the effectiveness of principal components can be found in Mastrangelo et al. (1996). Runger (1996) presented a model based on PCA for multivariate statistical process control where autocorrelation exists.

Multicollinearity in a process affects the least-square estimation of the regression coefficients. As a result, direct linear regression analysis is not valid and generates more false alarms on the control chart. Furthermore, the use of the regression control chart for processes in high dimensional systems is not practical. The traditional multivariate control chart performance deteriorates by the autocorrelation effect (Noorossana and Vaghefi, 2006). PCA has been used to reduce the size of variables, autocorrelation effect and multicollinearity problem for multivariate autocorrelated processes. However, a multiple regression model based on principal components has not been extended to monitoring the cascade process with input variables that significantly affect the quality of the output variables and exhibit both autocorrelation and multicollinearity effects. Additionally, the principal components selection rule and principal components interpretation are seldom discussed in quality control applications.

In this paper, a multiple regression model based on principal components is developed for monitoring multivariate autocorrelated processes to not only monitor process

output affected by many input variables but to also identify root causes of out-of control signals. We use the Principle Components Analysis to alleviate the multicollinearity problem among input variables and decrease the dimension of the variables. A multiple regression model with estimated parameters based on a selected number of principal components, determined through an integrated selection rule, is developed. The autoregressive error correction model is used to address the issue of dependent residuals from the multiple regression model. Further, interpretation of out-of-control signals based on the principal components in process monitoring is discussed.

The remainder of the paper is organized as follows: section 3.2 presents the proposed methodology in detail, followed by an actual industry implementation in section 3.3. Finally, a discussion and future work are given in section 3.4.

3.2. Principal Components Autoregressive Multiple Regression

The proposed methodology for multivariate autocorrelated process control uses a multiple regression control chart in conjunction with PCA. PCA is used to decrease the dimension of the variables in the process and reduce the multicollinearity effect (Jolliffe, 2002). The principal component selection rule is used to identify significant independent principal components with the linear combination of all variables. The response variable is regressed on all of the significant independent principal components by using the multiple regressions with autoregressive error correction model. Then, white noise series generated by the autoregressive error correction model is used to construct a control chart for future process monitoring. Components of the methodology are discussed in detail in the following subsections.

3.2.1. Multiple regression and autoregressive error correction model

Regression analysis is a statistical technique for examining and developing the relationship between response (output) and predictor (input) variables. When there is more than one predictor variable in the process, the multiple regression model is employed to capture the input and output relationship. Suppose a process of interest involves k independent variables, then the multiple regression model is given as follows:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki} + \varepsilon_i, (i = 1, 2, \dots, n) \quad (3.1)$$

where y_i is the i th observation of the response variable, β_k is the regression coefficient and

The key $\varepsilon_i \sim i.i.d.N(0, \sigma^2)$ assumption of the multiple regression model is that any one value of y is statistically independent of any other value of y . Therefore, checking the validity of the independence assumption is an essential step before further analysis of the data. The multiple regression model in (3.1) can be written in matrix notation as:

$$Y = XB + E \quad (3.2)$$

where $Y = [y_1 \ y_2 \ \dots \ y_n]'$, $E = [\varepsilon_1 \ \varepsilon_2 \ \dots \ \varepsilon_n]'$, $B = [\beta_0 \ \beta_1 \ \beta_2 \ \dots \ \beta_k]'$ and

$$X = \begin{bmatrix} x_{11} & x_{21} & \dots & x_{k1} \\ x_{12} & x_{22} & \dots & x_{k2} \\ \vdots & \vdots & & \vdots \\ x_{1n} & x_{2n} & \dots & x_{kn} \end{bmatrix}_{n(k+1)}$$

The regression coefficient estimates are given by $\hat{\beta} = (X'X)^{-1} X'Y$ variance-covariance matrix of the estimated regression coefficients can be represented as $Var(\hat{\beta}) = \sigma^2 (X'X)^{-1}$. The main assumption of ordinary regression analysis is that the errors should have mean of zero with constant variance and should be independent of each other. However, with time series process data, ordinary regression residuals are usually correlated over time. It is not desirable

to use ordinary regression analysis for time series process data. Therefore, we build a multiple regression model with an autoregressive series for the random error, accounting for the autocorrelation of the errors. Instead of the usual regression model, we use the following autoregressive error model:

$$y_t = XB + v_t$$

$$v_t = -\varphi_1 v_{t-1} - \varphi_2 v_{t-2} - \dots - \varphi_m v_{t-m} + \varepsilon_t \quad (3.3)$$

$$\varepsilon_i \sim i.i.d.N(0, \sigma^2)$$

where y_t is the response variable, v_t is the error term at time period t , and φ_m is autocorrelation parameter $\varphi_m < 1$. When the predictors in the process of interest have different units, standardization is necessary.

Consider equation (3.1), the independent variables are transformed as $x_{ji}^s = (x_{ji} - \bar{x})/s_j$ where x_{ji} is the standardized version of the i th observation for the j th variable x_j and $s_j = \sqrt{\sum_{i=1}^n (x_{ji} - \bar{x}_j)^2 / (n-1)}$. Thus, equation (3.1) can be rewritten in standardized version as:

$$y_i = \beta_0^s + \beta_1^s \left[\frac{x_{1i} - \bar{x}_1}{s_1} \right] + \beta_2^s \left[\frac{x_{2i} - \bar{x}_2}{s_2} \right] + \dots + \beta_k^s \left[\frac{x_{ki} - \bar{x}_k}{s_k} \right] + \varepsilon_i \quad (3.4)$$

If we separate the first column from the X matrix, then:

$$Y = \beta_0^s I + X^s B^s + E \quad (3.5)$$

where I is a vector of ones. It is noted that $X^s X^s$ is proportional to the correlation matrix for the predictor variables.

3.2.2. Principal components and selection rule

PCA is a method for alleviating the multicollinearity effect by reducing the dimensionality of the variables in the process (Jolliffe, 2002). In PCA, the original process

variables are transformed into a new set of orthogonal or uncorrelated variables called principal components. The PCA transformation ranks the new orthogonal variables in order of their importance. A number of principal components are selected to explain the most of the variability in the process, while others are eliminated to reduce the dimensionality of the variables. Let A be $k \times k$ matrix which consists of the normalized eigenvector such that $A'A = I$. Therefore, equation (3.5) can be written as:

$$Y = \beta_0^s I + X^s AA'B^s + E = \beta_0^s I + Z\gamma + E \quad (3.6)$$

where $Z = X^s A$ is an $n \times k$ matrix of the principal components and $\gamma = A'B^s$ is $k \times 1$ vector of new coefficients. The principal components are orthogonal to each other since:

$$Z'Z = (X^s A)'(X^s A) = A'X^{s'}X^s A = \text{diag}(\lambda_1 \quad \lambda_2 \quad \dots \quad \lambda_k) \quad (3.7)$$

where $\lambda_1 \quad \lambda_2 \quad \dots \quad \lambda_k$ are the eigenvalues of the matrix $X^{s'}X^s$, furthermore $\sum_{j=1}^m \lambda_j = k$. Then,

j th principal component can be computed as:

$$z_j = a_{1j}x_1^s + a_{2j}x_2^s + \dots + a_{kj}x_k^s \quad (3.8)$$

where $[a_{1j}, a_{2j}, \dots, a_{kj}]$ is the eigenvector associated with the eigenvalue λ_j , and z_j accounts for λ_j of the total variance. The least squares estimator for regression coefficient of the vector λ is given by $\hat{\gamma} = (Z'Z)^{-1}Z'Y$ and the variance for the estimated coefficients for $\hat{\gamma}$ can be computed as:

$$\text{Var}(\hat{\gamma}) = \hat{\sigma}^2 (Z'Z)^{-1} = \hat{\sigma}^2 \text{diag}(\lambda_1^{-1}, \lambda_2^{-2}, \dots, \lambda_k^{-k}) \quad (3.9)$$

The principal components are ranked according to the size of their eigenvalues. The highest eigenvalue is selected to be the first principal component, next highest eigenvalue is selected to be the second principal component, and so on. In PCA, it is critical to find the optimal number of principal components to be included in the multiple regression model without major loss of information. There are many rules for selecting principal components.

Each rule has different criteria which lead to a different number of principal components retained. Therefore, it is difficult for practitioners to determine the optimal number of principal components. For this reason, we propose a principal component selection rule derived from the integration of two standard principal component selection rules, found in the literature, to enhance principal component selection and to increase confidence in choosing the number of principal components. Both rules are based on variation in the principal components and work well in practice (Jolliffe, 2002). The first rule is based on the cumulative percentage of total variation of selected principal components, t_m , less than a predetermined cut-off point, t^* . However, this rule does not address the issue of determining the best cut-off point, t^* . The second rule is based on the size of variances of principal components, λ_j , and principal components with variance less than 0.7 are discarded as they contain less information than the original variables. Unlike the first rule, the second rule takes into account the sampling variation.

The selection rule we propose integrates these two rules while modifying the cut-off point used in the first rule. We use the range of cumulative percentage of total variation of 70% to 90%, instead of the exact cumulative percentage of total variation, to keep the potential number of principal components as much as possible without loss of major information of the process data. Principal components selected will simultaneously satisfy both rules. The proposed selection rule is given as:

$$t_m = \frac{100}{k} \sum_{j=1}^m \lambda_j, 70\% \leq t_m \leq 90\%, \text{ and } \lambda_j \geq 0.7 \quad (3.10)$$

where k is the original number of principal components, m is the number of principal components to be retained and t_m is cumulative percentage of total variation of selected principal components. For example, if the first three to six principal components variation is combined and it is in the range of 70% to 90%, then three to six principal components can be

retained. Furthermore, if the first four principal components have variances or eigenvalues larger than 0.7, then the final number to be retained are the first four principal components.

3.2.3. Interpretation of principal components

The correlation of the variables with the principal component is necessary for interpreting the importance of the k th variable to the j th principal component. The correlation coefficient between principal component z_j and variable x_k can be calculated as:

$$\rho_{z_j, x_k} = e_{jk} \sqrt{\lambda_j} / \sigma_{kk} \quad (3.11)$$

where e_{jk} is the eigenvector, λ_j is eigenvalue and σ_{kk} is standard deviation. However, this correlation coefficient does not measure the importance of an individual x_k to a response variable (y) in the presence of other independent variables. It is recommended in the PCA literature that only the eigenvector e_{jk} in absolute value should be used to interpret the components (Jolliffe, 2002). The correlation of the variables with principal component are simplified in the format of -1 , $+1$, (-1) and $(+1)$. The variable with maximum absolute correlation coefficient value for each individual principal component is identified, and this value is used as a comparable point for other variables. The correlation coefficient of other variables in absolute value greater than half of the absolute maximum correlation coefficient variables are indicated as -1 and $+1$, depending on their signs. The correlation coefficient of other variables in absolute value between a quarter and a half of the absolute maximum correlation coefficient variables are indicated as (-1) and $(+1)$.

The following steps describe the implementation procedure of the proposed approach.

Step 1: Verify that the process is stationary with multicollinearity and autocorrelated effects.

Step 2: Let be $x_{1i}, x_{2i}, \dots, x_{ni}$ standardized n correlated input variables. Use principal component analysis to transform these input variables into n uncorrelated variables z_1, z_2, \dots, z_n . These variables are called the principal components and can be denoted as PC_1, PC_2, \dots, PC_k where $k = n$. Apply the integrated principal component selection rule to reduce k principal components to m principal components. The new set of principal component is PC_1, PC_2, \dots, PC_m , where $m \leq k$.

Step 3: Let y_t be the standardized autocorrelated output variables. Regress on

PC_1, PC_2, \dots, PC_m , with the autoregressive error correction model to get the prediction y'_t of

which can be written as y'_t . The regression model is given by:

$$y'_t = \beta_0 + \beta_1 PC_1 + \beta_2 PC_2 + \dots + \beta_k PC_k + v_t$$

$$v_t = -\phi_1 v_{t-1} - \phi_2 v_{t-2} - \dots - \phi_p v_{t-p} + \varepsilon_t$$

Step 4: Generate and plot the white noise residuals $\tilde{y}_i = y_i - y'_i$ Shewhart control chart. The mean of the residuals is calculated with the upper and lower control limits of the control chart as follows:

$$\bar{\tilde{y}} \pm 3\sigma$$

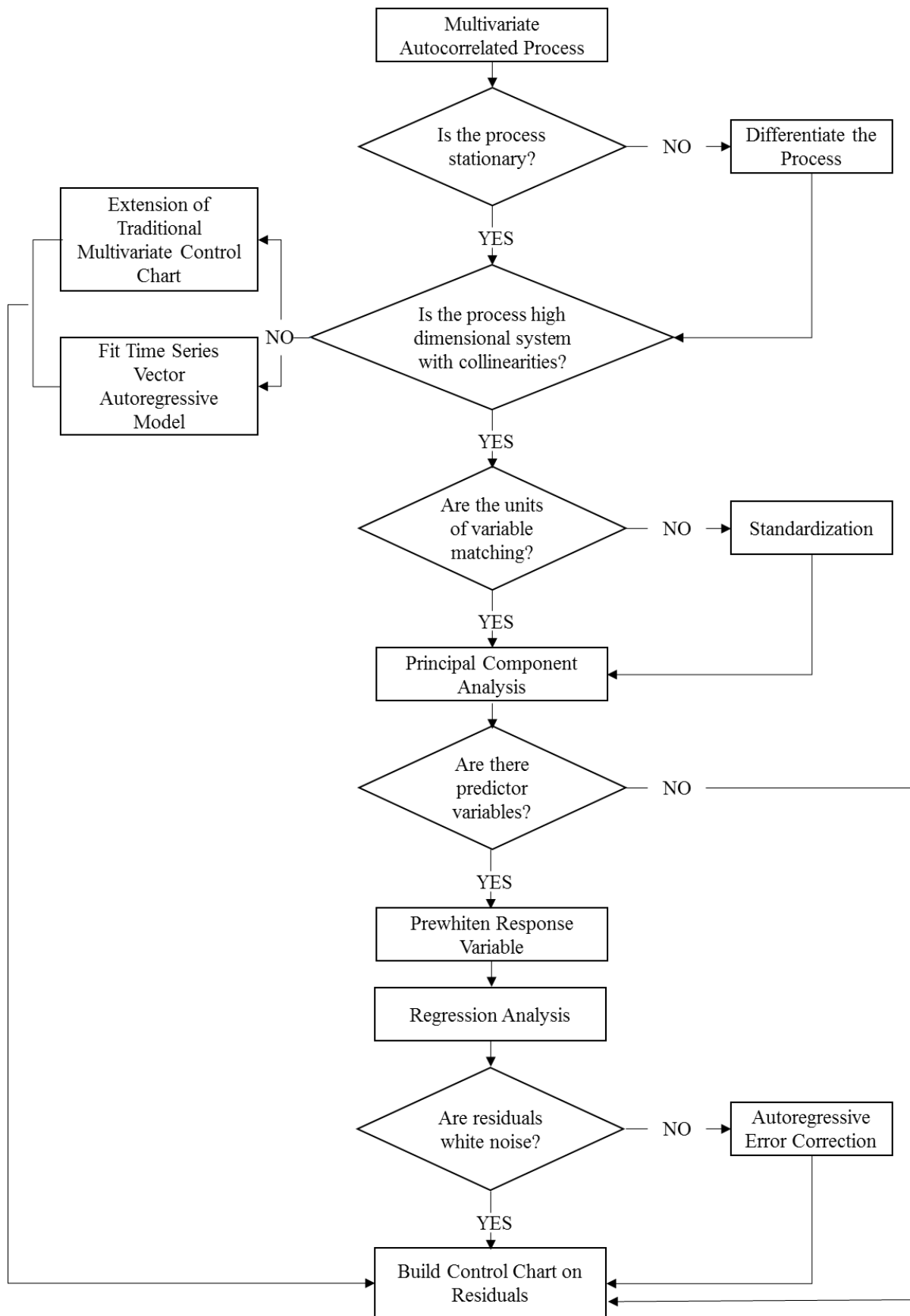


Figure 3.1. Flow chart for analyzing multivariate autocorrelated processes

A comprehensive guideline for multivariate autocorrelated processes analysis, monitoring and control is illustrated in Figure 3.1. In statistical process control, the process is assumed to be stationary, which means that statistical characteristics remain same as time changes. If the process is nonstationary, differentiation of the process is needed before further analysis is carried out. Makridakis et al. (1983) provide more details on how to handle non-stationary processes by differentiation. If the process is not a high dimensional system with multicollinearities, vector autoregressive control charts and modified traditional control charts are the options for monitoring the process. The method of monitoring the residuals from the vector autoregressive model is developed based on the technique that the residuals from fitting a time series model are independent and can be used to construct control charts (Jarrett and Pan, 2007; Jiang, 2004). The modified traditional multivariate control chart is developed from an extension of the multivariate Shewhart control chart for the autocorrelated and multicollinear process (Chan and Li, 1994; Charnes, 1995). Jiang (2004) proposed an extension of the Hotelling T^2 control chart for multivariate autocorrelated processes based on the generalized likelihood ratio test. In cases where the process of interest is a high dimensional system with multicollinearities, the PCA is necessary because variables with different units need to be standardized into identical units before further analysis is carried out (Jolliffe, 2002).

3.3. Application in Sugar Beets Pulp Drying Process

In this section, the application of the proposed approach is demonstrated by considering a real life example of the sugar beet pulp drying process discussed in Bilen et al. (2007, 2009). The general form of the sugar beet pulp drying process is shown in Figure 3.2. Sugar beet processing is the production of sugar (sucrose) from sugar beets. The major byproducts of sugar beet processing are pulp and molasses. Most of the molasses produced is processed further to remove the remaining sucrose. The pulp is dried, and sold as livestock

feed. If the moisture content of dry pulp is too high or low, it directly affects the quality of livestock feed and the transportation cost, as it weighs more with higher moisture content. Thus, it is important to monitor the moisture content of the dry pulp (y) that is affected by 12 measurable input variables as shown in Figure 3.3. The implementation procedure is divided into four steps as follows:

- 1) Check the process for stationarity multicollinearity and autocorrelation.
- 2) If the variables have different units, standardize the variables and use PCA to get independent predictor variables.
- 3) Construct a regression model to get a set of white noise residuals.
- 4) Build control charts on the white noise residuals for investigating the out-of-control signals and monitoring future process.

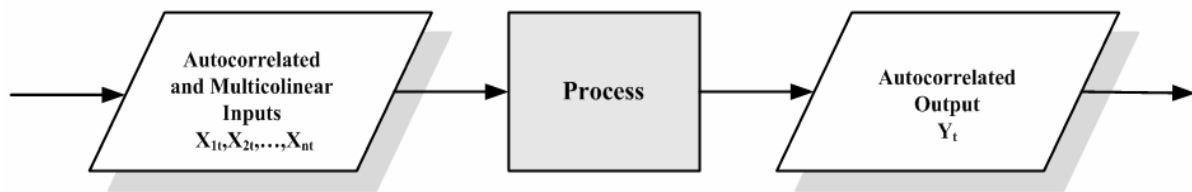


Figure 3.2. Pulp drying process in general

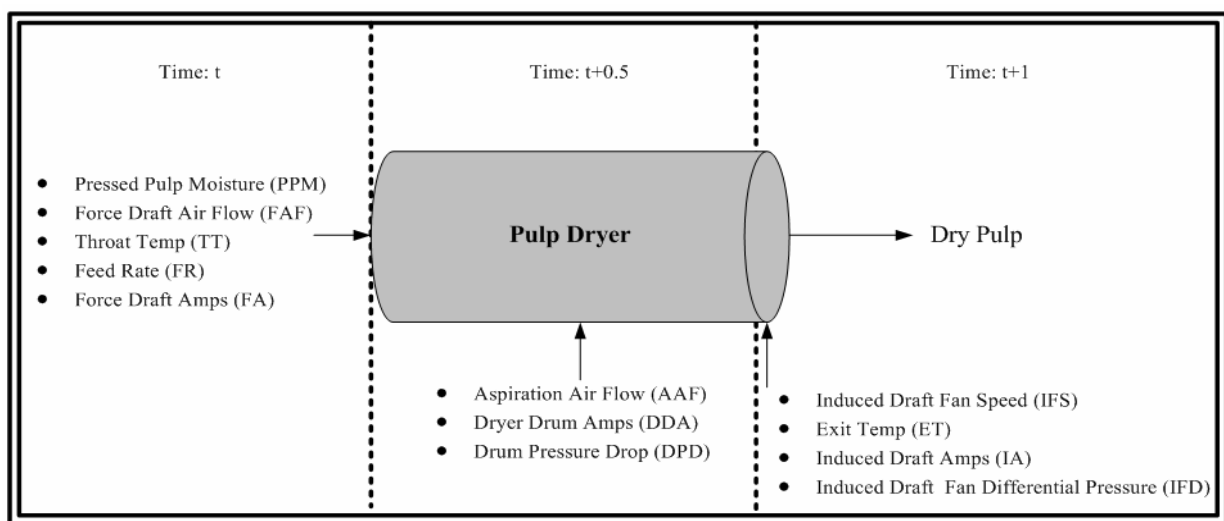


Figure 3.3. Pulp drying process according to time

A detailed discussion of dry pulp quality monitoring at each step of the analysis is given next.

Step 1: The sugar beet pulp drying process is shown in Figure 3. Historical data for a period of two months of production, 24 hours a day, are used in the analysis. Twelve predictor variables are adjusted according to time at t , $t + 0.5$ and $t + 1$. The observed pulp process data are somewhat messy and have a few missing points. Therefore, in order to construct the stationary control chart, 228 consistent observations, showing no obvious outliers, collected in the middle of the process are selected from a total 1,349 observations. Time series plots of all variables indicate a stationary process. The ACF and PACF plots indicate that all process variables are autocorrelated. The multicollinearity among input variables is identified from prior process knowledge.

Step 2: Due to different units of predictor variables, variables are standardized, as discussed in subsection 3.2.1. The SAS statistical analysis software is used to conduct PCA of the standardized variables. The eigenvalues for each principal component and correlation between principal components and standardized variables (correlation matrix) are provided in Table 3.1. Based on the integrated principal component selection rule, only the first four principal components are selected for further analysis. These principal components are used to explain the process without much loss of information contained in the 12 variables. Obviously, this reduction in the dimension of input variables reduces the complexity of the problem.

It is informative to determine the correlations of the predictor variables with the principal components in order to interpret the importance of the each variable to each principal component. The correlation coefficient between principal

component z_i , $i = 1,2,3,4$ and variable x_k , where $k = \text{PPM, FAF, TT, FR, FA, AAF, DDA, DPD, IFS, ET, IA}$ and IFD can be computed using equation (3.11). The first four eigenvectors associated with eigenvalues are presented in Table 3.2. Notable observations are that variable AAF receives the greatest weight in principal component 1; FAF, TT and FR are almost equally important to principal component 2; IFD and DPD contribute almost the same to principal component 3; and the most important variable to principal component 4 is DDA.

Step 3: Since the response variable (dry pulp moisture) is autocorrelated, it violates the independence assumption. Therefore, it is important to remove autocorrelation before the model is built by regressing four significant principal components. The method often used to deal with autocorrelation is to fit an ARIMA model for the input series so that it is sufficient to reduce the residuals to white noise (Chatfield, 2000). However, there are no ARIMA models that reasonably fit well for the response data since the residuals produced from the ARIMA model to build regression model will give an R-square value of almost zero. Therefore, we use the autoregressive error model. Again with the help of SAS, the following model coefficients are estimated by the maximum likelihood method:

$$\begin{aligned}
 y_t &= 9.6302 - 0.3199PC_1 + 0.1652PC_2 - 0.8376PC_3 + 0.4932PC_4 + v_t \\
 v_t &= 1.1188v_{t-1} - 0.5804v_{t-2} + 0.3945v_{t-3} + \varepsilon_t \\
 \varepsilon_t &\sim i.i.d.N(0, \sigma^2)
 \end{aligned}
 \tag{3.12}$$

The Durbin-Watson test for autocorrelation in residuals confirms that the residuals are white noise. And the intercept, components 1 to 4 and the coefficients for AR1, AR2 and AR3 are all significant, indicating that the variables associated with these coefficients are important and meaningful to the model. The R-square is 0.6137, which is not as high as desired. However, various factors, such as instrumentation,

gauging accuracy, process interruption by operators, and data recording errors affect data collection and the production process. Therefore, even a model with a moderate R-square is informative about the process and the relative importance of numerous variables to the process.

Step 4: From the analysis results, the mean of the process is -0.0016 and the standard deviation is 0.6933 . So the control limits for the Shewhart chart are computed as:

$$-0.0016 \pm 3(0.6933) = (-2.0815, 2.0783) \quad (3.13)$$

Table 3.1. Eigenvalues of the correlation matrix

PC	Eigenvalue	Proportion	Cumulative
1	4.6290	0.3858	0.3858
2	2.6333	0.2194	0.6052
3	1.6352	0.1363	0.7415
4	0.9960	0.0830	0.8245
5	0.6784	0.0565	0.8810
6	0.4533	0.0378	0.9188
7	0.3803	0.0317	0.9505
8	0.2762	0.0230	0.9735
9	0.1783	0.0149	0.9883
10	0.0694	0.0058	0.9941
11	0.0511	0.0043	0.9984
12	0.0195	0.0016	1.0000

The Shewhart Individuals control chart of dry pulp residuals is generated in Figure 3.4. The correlation matrix provides some information regarding the weight of the variables in the principal component. From Table 3.2, a simplified version of the correlation matrix is created

to interpret significant variables as shown in Table 3.3. For principal component 1 (PC1), eight variables, PPM, FA, IFS, ET, IA, IFD, AAF and DPD, are considered to be important based on their weights with maximum weight variables. Three variables, FAF, TT and FR, are significant in principal component 2 (PC2), even though their importance is minor compared to principal component 1. Principal component 3 (PC3) shows five variables, FA, IA, IFD, DPD, and DDA, as important, and only DDA is significant in principal component 4 (PC4). Since four principal components are regressed on the dry pulp, the weight of the variables changes according to equation (3.12). Table 3.4 shows the adjusted weight of each variable in each principal component and the overall weight of each variable in the model. When detecting an out-of-control signal, the sequence for checking process variables according to their priority should be: IFD, FA, DPD, IA, IFS, FR, TT, DDA, FAF, ET, PPM and AAF.

Table 3.2. Correlation between principal components and standardized variables

x	PC1	PC2	PC3	PC4
PPM	-.333218	0.146091	0.194914	0.118675
FAF	-.166576	-.539317	0.000217	-.024384
TT	0.202653	0.505909	0.056680	-.115743
FR	0.111124	0.512892	0.147625	-.057863
FA	0.293420	-.248280	0.414488	-.282509
IFS	0.380466	-.148433	0.119992	0.213348
ET	-.341491	0.231609	0.072179	-.090648
IA	-.225141	-.119583	0.395806	0.254624
IFD	0.338596	0.000655	0.466796	-.275830
AAF	-.439290	0.088861	0.107443	-.094088
DDA	0.067763	0.085509	0.353278	0.772074
DPD	0.312140	0.044331	-.484442	0.298605

Table 3.3. Simplified correlation between principal components and standardized variables

z	PC1	PC2	PC3	PC4
PPM	-	(+)	(+)	
FAF	(-)	-		
TT	(+)	+		
FR	(+)	+	(+)	
FA	+	(-)	+	(-)
IFS	+	(-)		(+)
ET	-	(+)		
IA	-		+	(+)
IFD	+		+	(-)
AAF	-			
DDA			+	+
DPD	+		-	(+)

There are several out-of-control signals in the Shewhart control chart given in Figure 3.4. It is difficult to identify a single source for each of the out-of-control signals since most variables are correlated with each other and the principal component is a linear combination of all other variables. A physical interpretation of the results is informative; for example, the aspiration air flow (AAF) is the air that is affected by the pressure difference between temperature at the beginning and at the end of the drum. If there is high pressure in the drum, then the pulp will have less moisture content, and vice versa. The throat temp (TT), the temperature at the beginning of the drum, behaves the same way as the exit temperature (ET), the temperature at the end point of the drum. If the exit temperature is higher, then the moisture in the pulp will decrease.

Table 3.4. Adjusted weight of each variable (X) standardized variables for each principal component and overall weight of each variable (X) to response variables (Y)

	-0.3199*PC1	0.1652*PC2	-0.3876*PC3	0.4932*PC4	y
PPM	0.10660	0.02413	-0.16326	0.05853	0.02600
FAF	0.05329	-0.08910	-0.00018	-0.01203	-0.04802
TT	-0.06483	0.08358	-0.04748	-0.05708	-0.08581
FR	-0.03555	0.08473	-0.12365	-0.02854	-0.10301
FA	-0.09387	-0.04102	-0.34718	-0.13933	-0.62139
IFS	-0.12171	-0.02452	-0.10051	0.10522	-0.14151
ET	0.10924	0.03826	-0.06046	-0.04471	0.04234
IA	0.07202	-0.01976	-0.33153	0.12558	-0.15368
IFD	-0.10832	0.00011	-0.39099	-0.13604	-0.63524
AAF	0.14053	0.01468	-0.08999	-0.04640	0.01881
DDA	-0.02168	0.01413	-0.29591	0.38079	0.07733
DPD	-0.09985	0.00732	0.40577	0.14727	0.46051

Furthermore, the feed rate (FR) and force draft air flow (FAF) will affect the duration of the pulp inside the drum; if the feed rate and force draft air flow are increased, the pulp will have less time inside the drum which directly leads to high moisture content in the pulp. The variables IFD, FA, DPD, IA, IFS, FR, TT, DDA, FAF, ET, PPM and AAF should be investigated as a sequence according to their importance once the out-of-control signal is detected. It is useful to construct an additional residual control chart of suspicious predictor variables to be able to identify the root cause and physical interpretation of the out-of-control signal. These signals needs to be investigated further based on the prior process knowledge.

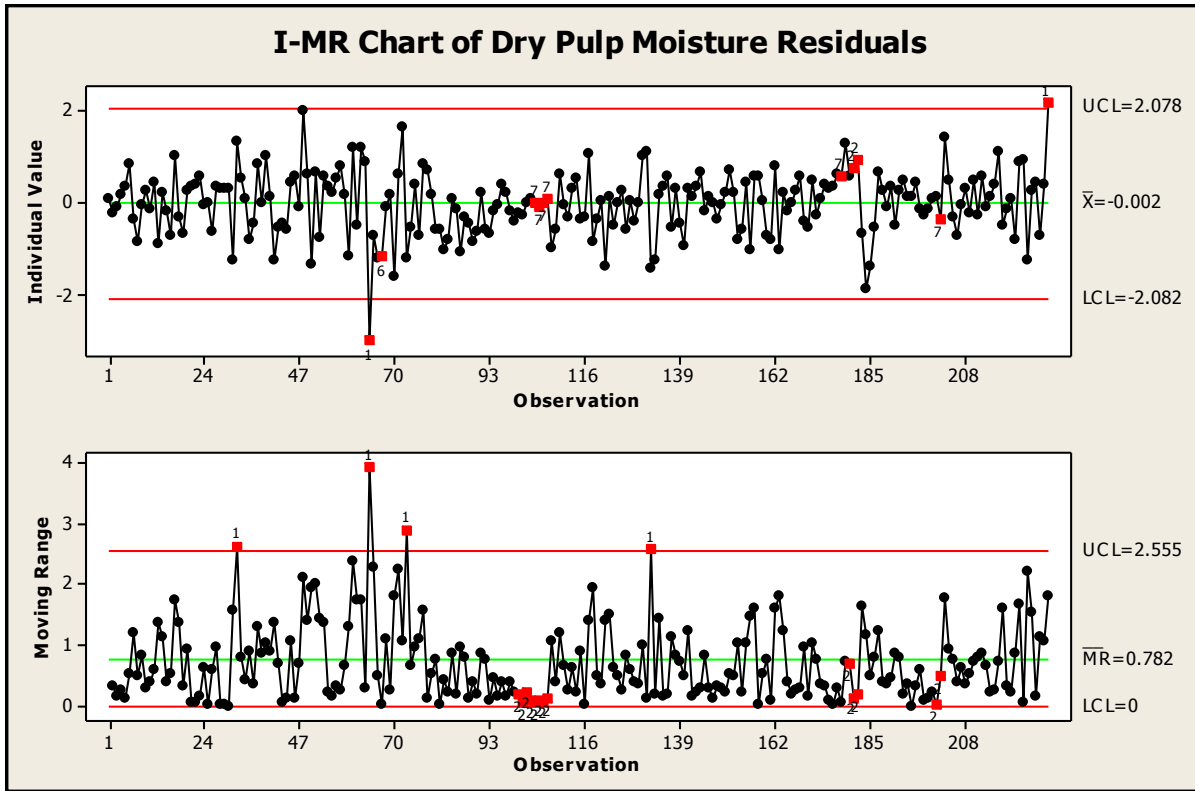


Figure 3.4. Shewhart control chart for dry pulp moisture residuals

3.4. Discussion

This paper developed a comprehensive framework for analyzing multivariate autocorrelated cascade processes. A guideline for handling autocorrelated and multicollinearity effects with a high dimensional system is also presented. The PCA reduces the number of input variables and removes the multicollinearity effect. The autoregressive error correction model along with multiple regression is applied to determine the relationship between output variables and the PCA of the original input variables. Independent residuals are generated based on the model. These white noise residuals can then be monitored using the Shewhart control chart. The application of the proposed framework has been demonstrated using a real life industry example. The cumulative sum and exponential weighted moving average can be used instead of the Shewhart control chart if the detection of small process shifts is desired. The proposed methodology can be expanded to monitor two or more response variables in multivariate cascade processes. Further, the model

developed to describe the relationship between input and output variables can also be used for process optimization as well as in-line, real-time control of the process.

CHAPTER 4. PAPER 2 DUAL MONITORING SCHEME FOR MULTIVARIATE AUTOCORRELATED CASCADE PROCESSES WITH EWMA AND MEWMA CONTROL CHARTS²

4.1. Introduction

Implementation of statistical process control (SPC) methods for monitoring multivariate autocorrelated cascade processes has been receiving increased attention in recent years. Traditionally, SPC is used to chart and monitor important process output variables. However quality in multivariate cascade processes can be described by the relationship between process input and output variables (Hawkins, 1993). For such processes, the latent relationship between process output and input variables can be used to predict the process output from input process variables. In fact, monitoring the process output can be more effective if the relationship between product and process variables can be used and incorporated (Mandel, 1969). By accounting for process input values, one can determine the expected value of process output and monitor the process through the white noise.

One such process, from the sugar beet industry as described by Bilen et al. (2010), is the pulp drying process (Figure 4.1). The major byproducts of sugar beet processing are pulp and molasses where the pulp is dried and used as livestock feed. Quality of the livestock feed depends on the moisture content of the dry pulp. The moisture content of the dry pulp (Y) is affected by measurable process characteristics or input variables, X_s , as shown in Figure 4.2.

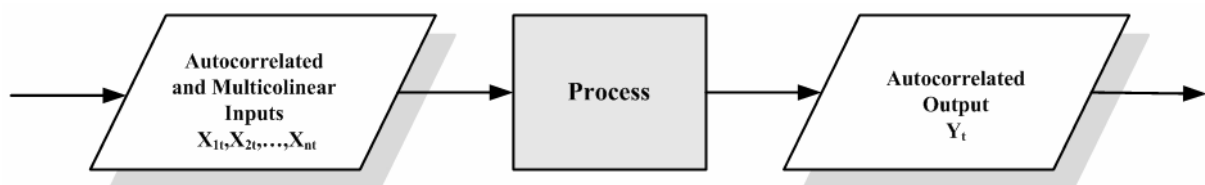


Figure 4.1. Pulp drying process in general

² Content in this chapter has been submitted for publication to Quality Engineering on April 28, 2014

To enhance the monitoring system, information on the process input variables can be used to successively determine and forecast the expected value of the process output so that the control limits for the output can be determined. Typically if predictor (process input) variables change, the response (the output variable) is directly affected. Basically, process output variable, Y , can be monitored by comparing its actual value with the expected range based on its relationship with the X s, i.e., using regression. Regression technique can be used as a tool to obtain those predicted value and the predefined range (Mandel, 1969). Under the controlled condition, the process input variables are at in-control state where joint means and covariances are at in-control values.

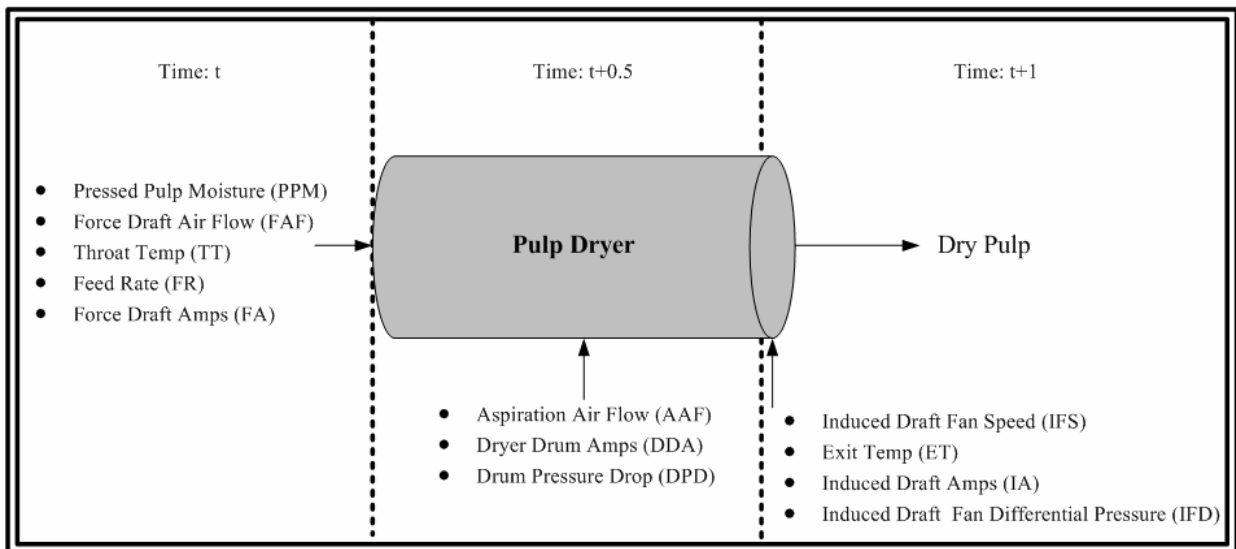


Figure 4.2. Pulp drying process according to time

For example, in the pulp drying process when the process is in-control state with no change in the relationship between process inputs and output, the moisture content of the dry pulp can be predicted based on input variable values. The actual moisture content can then be compared and charted against an expected/predicted value within a predefined range. The moisture content of the dry pulp Y is autocorrelated and, in this paper, we are interested in developing a method to control this autocorrelated process output, Y . For overall control of the multivariate cascade process, a monitoring scheme for the autocorrelated process output

which takes into account process input variables is necessary. In addition to the autocorrelation in the process output, for chemical processes such as the sugar beet pulp drying process, typically there is multicollinearity among key process input variables, X_s . For such cascade processes the multivariate regression control chart cannot be used to monitor the process output.

While multivariate control charts for monitoring process output variable without any consideration of input variables were proposed by several researchers (Jarrett and Pan, 2007; Jiang, 2004; Kalgonda and Kulkarni, 2004), there are also numerous examples and theoretical basis for monitoring the multicorrelated and autocorrelated processes based on Principal Components Analysis (Bisgaard and Kulahci, 2005; Jackson, 1980; Jolliffe, 1972 and 1973; Mastrangelo et al., 1996; Runger, 1996; Schall and Chandra, 1987). Principal Components Analysis effectively reduces the dimension of process input variables while retaining most of the process information. Detecting the out-of-control signal and finding the root cause of the process failure is complicated in multivariate autocorrelated statistical process control (Bersimis et al., 2007; Lowry and Montgomery, 1995; Woodall and Montgomery, 1999). Several types of multivariate control charts for monitoring process output variable without taking input variables into consideration have been proposed by a number of researchers (Jarrett and Pan, 2007; Jiang, 2004; Kalgonda and Kulkarni, 2004). The detection of the out-of-control signal can be even more difficult with the Principal Components Regression approach where information on the current process input values is used merely as regressor for Principal Components Regression. In fact these key process inputs/characteristics can also provide valuable information on the existing process performance. Further, multicollinearity in the process input variables X_s can affect the least squares estimation of parameters leading to incorrect control limits for the process output, Y . Hence, the multicollinearity of the input variables should be accounted for when monitoring multivariate cascade processes.

Bilen et al. (2010) presented and used the Principal Components Analysis to reduce the size of process input variables; eliminate the multicollinearity problem among input variables; and construct the multiple Principal Component Regression model for autocorrelated output variable. In this paper we extend the work of Bilen et al. (2010) by (i) improving the output monitoring scheme sensitivity using exponentially weighted moving average (EWMA) control scheme; (ii) using the Generalized Least Squares (*GLS*) approach on the Principal Components Regression in conjunction with the covariance structure of autoregressive error correction model on the residuals to obtain the best linear unbiased estimates of the autoregressive and the regression parameters; and (iii) applying the multivariate statistical process control scheme, namely MEWMA, to critical multicorrelated process inputs. The EWMA and MEWMA control charts are shown to be robust to non-normality, autocorrelation and effective in detecting small process shifts (Apley and Lee, 2008; Borrer et al., 2003; Carson and Yeh, 2008; Human et al., 2011; Koehler et al., 2001; Midi and Shabbak, 2011; Nuebauer, 1997; Lu, 1999; Scranton et al., 1996; Stoubos and Sullivan, 2002; Testik et al., 2003).

The proposed dual, simultaneous control of process inputs and output leads to better and thorough control of the process and faster identification of the out-of-control signal. This approach allows the process characteristics or input variables to be monitored and maintained at the in-control state which sequentially assists in maintaining the output characteristic at the desired level. In addition, the expected output can be regularly predicted and updated based on the observed realization of the process inputs varying within acceptable ranges. Then at each sampling interval, the expected output and its control limits can be updated based on the currently observed process input values. At the same time, the observed realization of process inputs are monitored and charted in order to monitor the current process characteristics. If the process inputs are not in-control state, one should not expect the output to be in-control state. Higher variations of the inputs lead to higher variation of the output. Rest of the paper is

organized as follows. Section 4.2. outlines the methodologies of principal components and principal components regression. Parameter estimation for the generalized least square regression and autoregressive error correction model is presented in Section 4.3. Section 4.4. illustrates the overall application of the dual control scheme and numerical results from the dual EWMA and MEWMA control chart scheme using the sugar beet pulp drying process data are illustrated in Section 4.4. The performance of the dual EWMA-MEWMA control scheme based on extensive numerical simulations is presented and discussed in Section 4.5. Finally, summary and discussion are given in Section 4.6.

4.2. Principal Component Regression Based on Generalized Least Squares under

Autoregressive Errors

The proposed methodology for multivariate autocorrelated cascade process control uses a multiple principal components regression technique to reduce the dimension of the variables in the process and eliminate the multicollinearity effect among the predictors (Jolliffe, 2002). The principal components selection rule is used to identify significant principal components which are linear combinations of process input variables. The response is regressed on selected significant independent principal components using the multiple Generalized Least Squares with autoregressive error correction model. Subsequently, white noise residuals are extracted and used to construct a control chart for process monitoring. Components of the proposed methodology are discussed in detail in the following subsections.

4.2.1. Principal component regression

Principal Component Regression (PCR) is a regression analysis technique based on Principal Components Analysis (PCA). In PCR, the response variable is linearly regressed on a select number of principal components (pc). Principal components are generally obtained by transforming the original input variables into uncorrelated input variables. The selected subset contains the pcs that explain most variability ($\geq 80\%$) in the process inputs. Remaining pcs are

discarded with minimal loss of information while reducing the dimension of the *pcs*. As *pcs* are uncorrelated there is no multicollinearity effect and hence multiple PCR based on Generalized Least Squares (*GLS*) can be used. In addition to reducing the dimension of regressor variables, PCR also provides accurate prediction of the response variable through appropriate selection of the *pcs* used in the regression model.

Let a sample of n collected single observation vectors of the process inputs be denoted by $x_1, \dots, x_t, \dots, x_n$ where $x_t = (x_{t1}, \dots, x_{tk})^T$ is an $k \times 1$ column vector of a single observation of the process inputs sampled at time t . In the dry pulp process, there are 12 process inputs, so $k = 12$. Let $\mathbf{X} = [\mathbf{x}_1 \dots \mathbf{x}_k]$ denotes an $n \times k$ matrix of the sampled process inputs where $\mathbf{x}_i = (x_{i1}, \dots, x_{in})^T$ represents the $n \times 1$ column vector the i^{th} input collected from the sample of size n . Let their associated n values of the process output be represented by an $n \times 1$ column vector $\mathbf{Y} = (Y_1, \dots, Y_n)^T$. Note that in this paper, both \mathbf{Y} and \mathbf{X} s are standardized to accommodate for difference in variation as well as units. The conventional multiple regression equation can be formulated and written in matrix form as:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e} \quad (4.1)$$

where $\boldsymbol{\beta}$ is $k \times 1$ vector regression parameters, \mathbf{e} is $n \times 1$ column vector of the error terms. Under independence assumption of the error terms, the estimated regression coefficients using ordinary least squares (*OLS*) estimator are given by $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}$. The variance-covariance matrix of the estimated regression coefficient is $Var(\hat{\boldsymbol{\beta}}) = \sigma_e^2 (\mathbf{X}^T \mathbf{X})^{-1}$.

To perform PCA and PCR, first an $n \times k$ matrix of *pcs* $\mathbf{Z} = [\mathbf{z}_1 \dots \mathbf{z}_m \mathbf{z}_{m+1} \dots \mathbf{z}_k]$ must be obtained. Note that \mathbf{z}_i is an $n \times 1$ vector of the i^{th} *pcs*. For unknown covariance matrix of \mathbf{X} , the principal component matrix is given by $\mathbf{Z} = \mathbf{X}\mathbf{A}$, where \mathbf{D} and \mathbf{A} are the eigenvalues and

eigenvectors of the sample variance-covariance matrix $\mathbf{S} = \mathbf{X}\mathbf{X}^T/(n-1)$, respectively. Then, \mathbf{Y} is regressed on the principal components and can be represented as:

$$\mathbf{Y} = \mathbf{Z}\boldsymbol{\gamma} + \mathbf{e} \quad (4.2)$$

where $\boldsymbol{\gamma}$ is $k \times 1$ vector of PCR coefficients which can be estimated using *OLS* as $\hat{\boldsymbol{\gamma}}_{OLS} = (\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T\mathbf{Y}$. Covariances of the estimated coefficients of $\hat{\boldsymbol{\gamma}}$ can be computed as $\text{Var}(\hat{\boldsymbol{\gamma}}_{OLS}) = \sigma_e^2(\mathbf{Z}^T\mathbf{Z})^{-1}$. For known values of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, the predicted response with respect to a single observation vector at time t , $x_t = (x_{t1}, \dots, x_{tk})^T$ can be obtained as $x_t^T\boldsymbol{\beta} = \boldsymbol{\beta}^T x_t$ or $z_t^T\boldsymbol{\gamma} = \boldsymbol{\gamma}^T z_t$ where $z_t = \mathbf{A}^T x_t$, $z_t = (z_{t1}, \dots, z_{tm}, z_{tm+1}, \dots, z_{tk})^T$. Also, the relationship between the known regression coefficients can be represented as $\boldsymbol{\gamma} = \mathbf{A}^T\boldsymbol{\beta}$ or $\boldsymbol{\beta} = \mathbf{A}\boldsymbol{\gamma}$. The regression coefficients vector $\boldsymbol{\gamma}$ can be theoretically transformed back to the original regression coefficient vector, $\boldsymbol{\beta}$. Based on observations from the pulp drying process, Bilen et al. (2010) showed that only four of 12 *pcs* should be statistically used to explain and represent the 12 process inputs, *Xs*. This is a great reduction in dimension of regressor variables. Hence, only those selected significant *pcs* should retained in Equation based on integration selection rule explained in the next subsection. Multiple PCR is then applied using selected *pcs*.

4.2.2. Selecting significant principal components

PCA is a method of transforming correlated set of original variables into uncorrelated new set of variables called *pcs* (Jolliffe, 2002). The eigenvalues of the *pcs* are ordered from largest to smallest. The first *pc* is selected based on the largest eigenvalue, the second *pc* is selected based on the second largest eigenvalue and so on. Selection of the optimal number of *pcs* to include in the multiple regression model is crucial for reducing the number of the variables while still being able to explain the majority of process input data information. The number of *pcs* is always less than or equal to the number of original variables. Given k

correlated variables of interest, PCA transforms k correlated variables into k uncorrelated pcs . The general selection rule by Jolliffe (2002) can be used to select the most significant number of $m pcs$, where $m \ll k$.

Bilen et al. (2010) developed an integrated selection rule to select $m pcs$ whose variance is at least 0.7 and the cumulative percentage of total variation of the principal components is greater than or equal to a threshold (t_m) value:

$$t_m = \frac{100}{k} \sum_{j=1}^m \lambda_j \text{ and } 70\% \leq t_m \text{ and } \lambda_j \geq 0.7 \quad (4.3)$$

where k and m are the original and the selected number of pcs , respectively. The variance the j^{th} principal component is given by λ_j . Selected number of pcs must satisfy both conditions: selected pcs must have cumulative percentage at least 70% of total variation and magnitude of pcs variance must be at least 0.7. Bilen et al. (2010) showed that this integrated selection rule is effective in reducing the number of pcs while retaining maximum information on the variances of the process for pulp drying process data. In this paper we use this same rule to (i) select the $n \times m$ matrix of pcs from the matrix \mathbf{Z} , and (ii) obtain the estimates of the PCR parameters γ in equation 4.2 based on selected m principal components. Without loss of generality, the principal component regressors are expressed as $mx1$ column vector of $z_t = (z_{t1}, \dots, z_{tm})^T$ and $\mathbf{Z} = [\mathbf{z}_1 \dots \mathbf{z}_m]$. For the pulp dryer process data, $m=4$.

4.2.3. Multiple generalized least squares regression on significant principal components under autoregressive error correction model

Under the autoregressive error correction model, Multiple PCR is carried based on the *GLS* method. Assuming that an autoregressive (AR) error correction model provides the best fit to the actual data, the regression model at time t can be expressed as:

$$y_t = x_t^T \boldsymbol{\beta} + \varepsilon_t \quad (4.4)$$

$$\varepsilon_t = \varphi_1 \varepsilon_{t-1} - \varphi_2 \varepsilon_{t-2} - \dots - \varphi_p \varepsilon_{t-p} + V_t$$

where y_t is the response variable; ε_t is the correlated residuals at time period t ; φ_p is autocorrelation coefficient of lag p and $|\varphi_p| < 1$; and white noise $v_t \sim iid N(0, \sigma^2)$. If *OLS* is applied to time series process output, the *OLS* residuals will remain correlated over time resulting in three significant shortcomings: 1) incorrect statistical tests of significance of the parameters and the confidence limit for predicted value \hat{y} , 2) inefficient estimates of regression coefficient, and 3) lack of improvement in predicting the future value since ordinary regression residuals are not independent. Under the AR model, in this paper we use multiple *GLS* to obtain an unbiased estimates of: regression parameter, γ ; autoregressive parameter, φ_i ; and the variance of the white noise, σ_e^2 . This multiple *GLS* is a better approach for estimating of these parameters. When *GLS* is employed in equation [1], the regression coefficient estimates are given by $\hat{\boldsymbol{\beta}}_{GLS} = (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1} \mathbf{Y}$ where \mathbf{V} is the covariance matrix of ε_t . The variance-covariance matrix of the estimated regression coefficient is represented by $Var(\hat{\boldsymbol{\beta}}_{GLS}) = (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1}$. For $m=4$, the regression model at time t can be expressed as:

$$y_t = z_t^T \boldsymbol{\beta} + \varepsilon_t \quad (4.5)$$

$$\varepsilon_t = \varphi_1 \varepsilon_{t-1} - \varphi_2 \varepsilon_{t-2} - \dots - \varphi_p \varepsilon_{t-p} + V_t$$

where $z_t = (z_{t1}, \dots, z_{t4})^T$. Replacing \mathbf{X} with the matrix of the selected *pcs*, \mathbf{Z} , the estimated regression coefficient of $\boldsymbol{\beta}$ can be calculated as:

$$\hat{\boldsymbol{\beta}}_{GLS} = (\mathbf{Z}^T \mathbf{V}^{-1} \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{V}^{-1} \mathbf{Y} \quad (4.6)$$

and the variance-covariance matrix of $\hat{\boldsymbol{\beta}}_{GLS}$ can be computed as $Var(\hat{\boldsymbol{\beta}}_{GLS}) = (\mathbf{Z}^T \mathbf{V}^{-1} \mathbf{Z})^{-1}$.

With unknown autoregressive error model parameters φ_q and σ_e^2 , the parameter estimates of

$\hat{\boldsymbol{\beta}}_{GLS}$ and $\hat{\mathbf{V}}$ must be obtained:

1. Obtain $\hat{\boldsymbol{\beta}}_{GLS}$ by assuming $\mathbf{V} = \sigma_e^2 \mathbf{I}_{n \times n}$ and compute residuals $\hat{\varepsilon}_t$.
2. Fit residuals $\hat{\varepsilon}_t$ with AR(p) model estimates $\hat{\varphi}_1, \dots, \hat{\varphi}_p$ and $\hat{\sigma}_e^2$ from this model:

$$\hat{\varepsilon}_t = \varphi_1 \hat{\varepsilon}_{t-1} - \varphi_2 \hat{\varepsilon}_{t-2} - \dots - \varphi_p \hat{\varepsilon}_{t-p} + \hat{v}_t$$

3. Estimate $\hat{\mathbf{V}}$ with respect to AR(p) model based on estimates $\hat{\varphi}_1, \dots, \hat{\varphi}_p, \hat{\sigma}_e^2$.
4. Re-compute $\hat{\boldsymbol{\beta}}_{GLS} = (\mathbf{Z}^T \hat{\mathbf{V}}^{-1} \mathbf{Z})^{-1} \mathbf{Z}^T \hat{\mathbf{V}}^{-1} \mathbf{Y}$ and re-compute residuals $\hat{\varepsilon}_t$.
5. Re-fit residuals $\hat{\varepsilon}_t$ from step 4 with AR(p) model, re-estimate the $\hat{\varphi}_1, \dots, \hat{\varphi}_p$ and $\hat{\sigma}_e^2$.
6. Repeat steps 2 to 5 until estimates for $\hat{\varphi}$ and $\hat{\sigma}_e^2$ converge.

A flow chart showing the sequence of steps, for determining the *GLS* regression coefficients, autocorrelation parameters and the white noise variance, is also provided in Figure 4.3.

Once *GLS* converges, the estimate of white noise residuals series \hat{v}_t is determined and used to monitor the process using the EWMA control chart. Without structural changes in the relationship between the process input and output variables, $\hat{v}_t \sim iid.N(0, \hat{\sigma}_e^2)$. Then the EWMA control chart applied on \hat{v}_t can be calculated as:

$$U_t = \lambda \hat{v}_t + (1 - \lambda) U_{t-1} \text{ for } t = 1, 2, \dots, n \quad (4.7)$$

with $U_0 = 0$ and $0 < \lambda \leq 1$ is a smoothing parameter of the EWMA control chart. The appropriate value of λ is provided by Lucas and Saccucci (1990). The variance of control statistic is computed as:

$$\sigma_U^2 = \left(\frac{\lambda}{2 - \lambda} \right) \left[1 - (1 - \lambda)^{2i} \right] \hat{\sigma}_e^2 \quad (4.8)$$

The upper and lower control limits, UCL and LCL, of the EWMA are given as:

$$\begin{aligned} \text{UCL}_U &= +L\sigma_U \\ \text{LCL}_U &= -L\sigma_U \end{aligned} \quad (4.9)$$

where (λ, L) are chosen with respect to a desired false alarm rate. In this paper, the parameters (λ, L) are selected with respect to false alarm rate of 500.

4.2.4. Multivariate exponential weighted moving average (MEWMA) control charts on process inputs

In this paper we propose using the MEWMA control chart to monitor key process inputs. We assume that under normal operating conditions, the process input characteristics X_s are normally non-constant but subject to an in-control state. Under the in-control state, process adjustment or feedback control is not required or implemented. However, the process output characteristic, may fluctuate within controlled range of variation as a result of random but controllable nature of the process inputs X_s . Given a known relationship between process output and inputs, the EWMA statistic monitors shifts in the structure of the relationship between process output and inputs. Since the shifts in process inputs can directly affect the output, in this paper we also use the MEWMA to help monitor process inputs. By adding multivariate SPC scheme onto input X_s , both the process and product characteristics can be monitored. Monitoring process characteristics will help prevent any significant changes in the process output. Similar to EWMA, the MEWMA control chart has been shown to be robust to non-normality (Lowry et al., 1992). So in this paper we apply MEWMA to the critical multicorrelated input X_s in order to help identify whether process inputs are in-control. This concept of dual monitoring of both process inputs X_s and output Y simultaneously is more effective and extends the work of Bilen et al. (2010).

Similar to the U^2 chart that monitors the subset of variables where the mean shifts are likely to occur in a subspace, in this paper we focus on key process input characteristics X_s . Key process variables are those with maximum weights in the first principal components. Based on our process knowledge, these variables are actually the critical inputs to the pulp drying process. These variables are monitored using the MEWMA chart. While the MEWMA

chart can also be used to monitor principal components as suggested by Mastrangelo et al. (1996), to avoid the unnecessary difficulty in identifying the source of an out-of-control signal based on principal components, in this paper the MEWMA chart is applied directly to identified key process inputs.

Let \mathbf{X}_i denote a single observation column vector of $\mathbf{X} = (X_1, \dots, X_m)$ sampled at time i . Recall that all the X s are standardized with zero means and unit variances. Without loss of generality, the mean of \mathbf{X} is zero vector and the covariance matrix is Σ . The \mathbf{X}_i is assumed to be independent over time. If $r_1 = r_2 = \dots = r_m = r$, then the MEWMA is given as follows:

$$\mathbf{U}_i = r\mathbf{X}_i + (1-r)\mathbf{U}_{i-1} \quad (4.10)$$

where $\mathbf{U}_0 = 0$ and $0 < r \leq 1$. The MEWMA control chart will show the out-of-control signal, when:

$$T_i^2 = \mathbf{U}_i^T \Sigma_{\mathbf{u}_i}^{-1} \mathbf{U}_i > h_4 \quad (4.11)$$

where $h_4 > 0$ for $m = 4$ is chosen in the aim of achieving predetermined in control ARL (Montgomery, 2012). Lowry et al. (1992) presented the h_4 value for different number of variables of interest and various values of in-control ARL. The covariance matrix can be calculated as:

$$\Sigma_{\mathbf{u}_i} = \left(\frac{r}{2-r} \right) \left[1 - (1-r)^{2i} \right] \Sigma \quad (4.12)$$

Prabhu and Runger (1997) recommended using $r = 0.2$ for in-control ARL of 500 with $k = 4$

In this paper we also choose $r = 0.2$ and h_4 for an in-control ARL of 500.

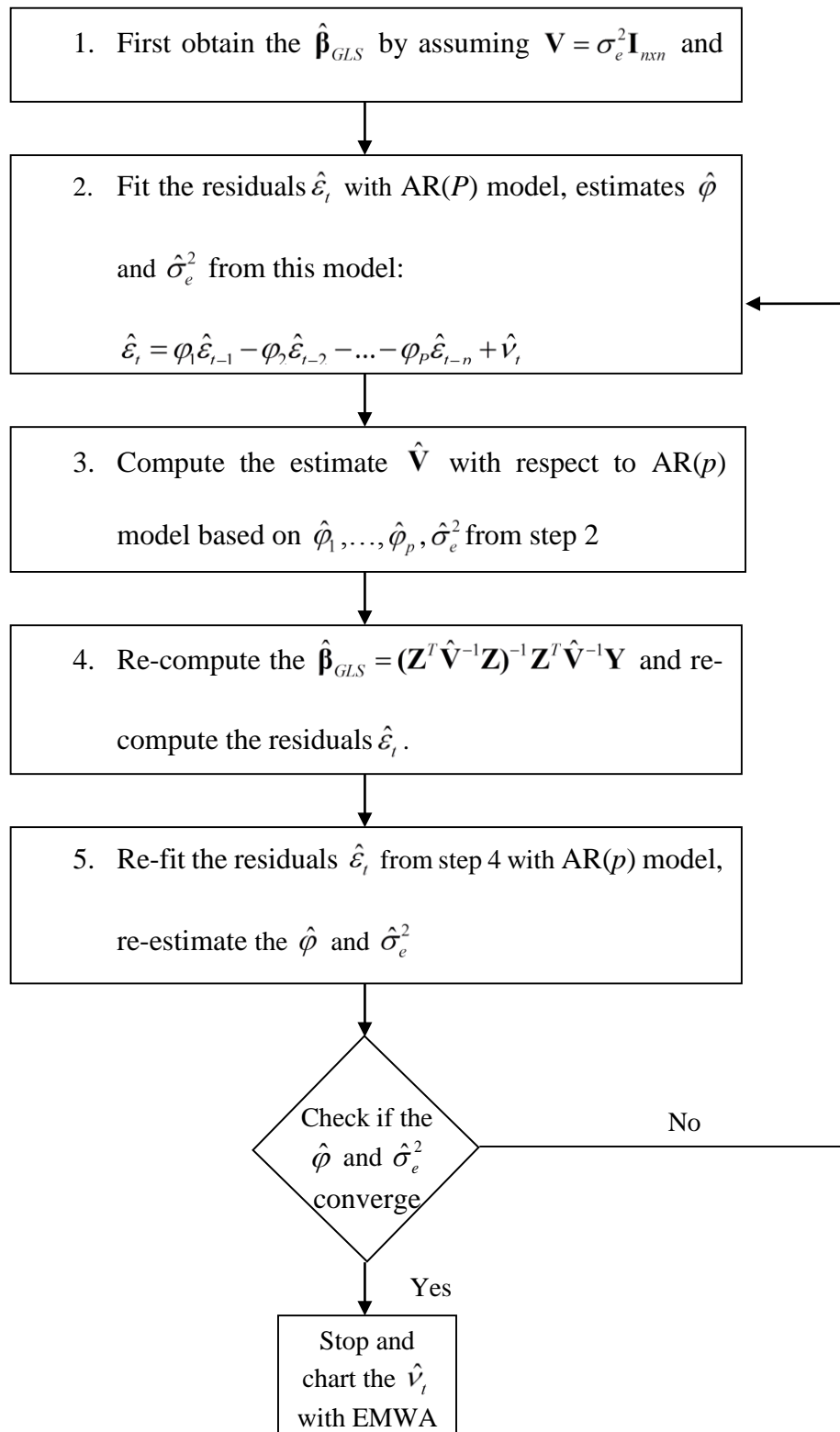


Figure 4.3. Autoregressive error correction parameters estimation procedure

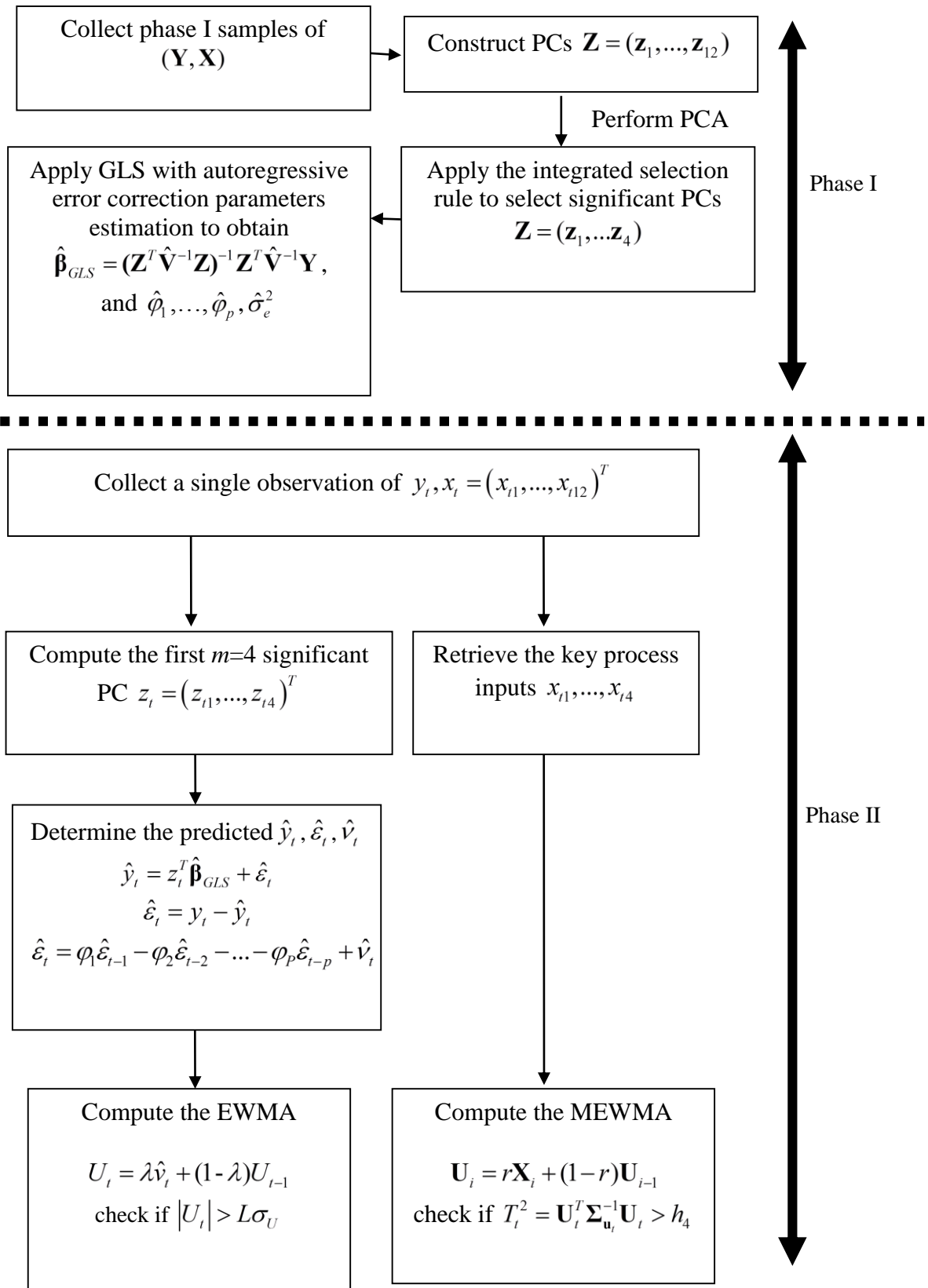


Figure 4.4. The general configuration of dual EWMA and MEWMA control chart scheme

4.3. Dual EWMA and MEWMA Control Chart Diagram

Dual EWMA and MEWMA control chart scheme is developed to monitor both process input variables, X_s , and product or output variable, Y , simultaneously. The predicted Y is based on PCR from first selected pcs . The residual-based EWMA is used to monitor structural changes in the relationships between process inputs and output. With no change in the structure the residuals, i.e. the difference between observed Y and the predicted Y , under GLS will be independent and normally distributed with mean zero and constant variance. When there is a change in the PCR model, the white noise residuals will no longer center around zero. So the EWMA statistic will signal a change in the relationship between process inputs and output as well as changes among the process input variables. When mean or covariance of X changes to an out-of-control state, the structure of PCR will also change. This will result in changes in the mean of the residuals and white noises since the pc regressor is determined based on the in-control state of covariance of X . Therefore, the EWMA can capture not only changes in structural relationship but also changes in the key process input parameters. In contrast, the MEWMA is used to directly monitor the shift in the mean of key process variables, X_s . The MEWMA can signal changes in the covariance of X as well. Therefore our dual monitoring scheme can capture changes in (i) the mean, (ii) the covariance, and (iii) the structure for Y and X_s . Hence this dual control is more effective in identifying changes in both process inputs and output. The construction and use of our dual schemes is divided into two phases: the dual control chart construction (Phase I) and implementation (Phase II).

During Phase I, a sample of single observations $y_i, x_i = (x_{i1}, \dots, x_{i12})^T \quad i=1, \dots, n$ are collected from an in-control state. Once the matrices of (Y, X) are constructed, PCA is conducted and the matrix of significant pcs , Z , is obtained. Next the multiple PCR based on GLS is performed to obtain the parameter estimates of regression coefficients $\hat{\beta}_{GLS}$ and

parameters of the autoregressive error model, $\hat{\phi}_1, \dots, \hat{\phi}_p, \hat{\sigma}_e^2$. These estimates are used for prediction and calculation of the residuals and white noise during Phase II.

During the implementation stage (Phase II), a single observation $y_t, x_t = (x_{t1}, \dots, x_{t12})^T$ at time t is sampled. The first m significant pcs , $z_t = (z_{t1}, \dots, z_{t4})^T$, are calculated and used to predict the process output with respect to those currently values of x_t through z_t using the $\hat{\beta}_{GLS}, \hat{\phi}_1, \dots, \hat{\phi}_p, \hat{\sigma}_e^2$ estimated from phase I. Next the predicted $\hat{y}_t, \hat{\varepsilon}_t, \hat{v}_t$ are determined so that the residual-based EWMA statistic can be charted. At the same time, key process inputs, $x_t = (x_{t1}, \dots, x_{t4})^T$, are monitored with the MEWMA chart. When the EWMA or MEWMA statistic gives an out-of-control signal, the process is stopped to search for assignable causes. For out-of-control signals on the MEWMA chart, key process inputs are investigated. A conventional decomposition can be applied to identify potential process input variables responsible for the out-of-control signal (Montgomery, 2012). For signals on the EWMA chart, the controlling condition of the process output is deemed to be out-of-control and an assignable cause is searched. Figure 4.4 gives an overall configuration of the implementation of the dual EWMA and MEWMA control chart scheme.

4.4. Implementation of the Dual Control Scheme

We use historical data, for a period of two months of production, from the pulp drying process shown in Figure 4.2 to illustrate the steps of the dual control scheme. The observed process pulp process data are somewhat messy and have a few missing points. Therefore, in order to construct the stationary control chart, 228 consistent (and continuous) observations, showing no obvious outliers, were identified and used from a total 1,349 observations during phase I construction. Time series plots of process variables indicate a stationary process. Figure 4.5 shows the time series plot of process output, dry pulp moisture content. The ACF and PACF plots indicate that all process variables are autocorrelated. multicollinearity among

input variables is identified and also affirmed by prior knowledge of the process. PCA is used to alleviate the multicollinearity problem among process variables yet maximally maintaining the relationship structure between process inputs and output, and at the same time reducing the dimension of the charting variables. Then the model-based white noise are extracted from the PCR under *GLS* and charted for monitoring the product/output.

Principal components of the multivariate process input variables of the sugar beet processing is selected using the Bilen et al. (2010) integrated selection rule. The autoregressive model is used to represent the autocorrelated moisture content of the dry pulp. The PCR is performed by regressing the autocorrelated moisture content on the significant principal components. This is carried as the autoregressive error correction model is applied to the time dependent residuals from the multiple PCR model so that the stationary white noise can be obtained and charted on the univariate Shewhart type control chart.

The results of the eigenvectors and their associated eigenvalues are presented in Table 4.1. The linear combination coefficients of the original process input variables are shown in Table 4.2. Note that according to the selection rule in equation (4.3), key *pcs* are chosen for constructing the relationship between the moisture content, *Y*, and *Xs* through the principal component, *Z*.

Through principal components analysis and knowledge of the process, four key process input variables are identified and monitored: aspiration air flow (AAF), induced draft fan speed (IFS), exit temperature (ET) and induced draft fan differential pressure (IFD). These four variables received maximum weights in the first principal components out of the four significant principal components. To increase detection power, EWMA control chart is applied on the white noise. On the process inputs side, MEWMA is used to monitor identified key multivariate process input variables. These key process inputs are not only the main contributing factor of the PCR but are also vital characteristics to the product output, moisture

content of the dry pulp. For the construction of the EWMA control chart, the parameter parameters (λ, L) are selected to be $(0.2, 2.962)$ for false alarm rate, ARL_0 , of 500. Similarly, MEWMA control chart parameters (r, h_4) are selected to be $(0.2, 16)$ for false alarm rate, ARL_0 , of 500. These parameters were empirically justified using simulation. So the statistics of the dual control scheme are

$$EWMA_t = 0.2\hat{v}_t + 0.8EWMA_{t-1} \quad (4.13)$$

$$CL_{EWMA_t} = \pm 2.962\sigma_{EWMA_t}$$

$$\sigma_{EWMA_t}^2 = \left(\frac{1}{9}\right) [1 - 0.8^{2t}] 0.6933^2$$

The MEWMA statistic for $\mathbf{X} = (X_1, \dots, X_4)$ is

$$MEWMA_t = 0.2\mathbf{X}_t + 0.8MEWMA_{t-1} \quad (4.14)$$

$$T_t^2 = MEWMA_t^T \Sigma_{MEWMA_t}^{-1} MEWMA_t$$

$$\Sigma_{EWMA_t} = \left(\frac{1}{9}\right) [1 - 0.8^{2t}] \hat{\Sigma}, h_4 = 16.$$

At each time interval, process inputs and output are automatically recorded by the pulp drying process sensors. The EWMA-MEWMA statistics are then computed and updated at each sampling interval. The computed statistics are then charted and compared against their control limits. The application of this dual scheme results in higher performance of the overall process control.

By adding multivariate SPC scheme onto process input variables, Xs, both the process and product characteristics are monitored. The extracted residuals conditioned on the current process input values provide reference information on changes in the structural relationship between the product and the process. Monitoring of the process characteristics help identify any developing change in the process output, moisture content of dry pulp, well in advance.

Therefore multivariate control charting, of key multicorrelated process input variables, X_s , helps identify whether the process inputs are in-control. This approach yields better understanding and more effective control for multivariate cascade processes. The performance of the proposed dual EWMA-MEWMA control scheme has also been studied using numerical simulation. The results of the numerical study are discussed in the next section.

Table 4.1. Eigenvalues of the correlation matrix

PC	Eigenvalue	Proportion	Cumulative
1	4.6290	0.3858	0.3858
2	2.6333	0.2194	0.6052
3	1.6352	0.1363	0.7415
4	0.9960	0.0830	0.8245
5	0.6784	0.0565	0.8810
6	0.4533	0.0378	0.9188
7	0.3803	0.0317	0.9505
8	0.2762	0.0230	0.9735
9	0.1783	0.0149	0.9883
10	0.0694	0.0058	0.9941
11	0.0511	0.0043	0.9984
12	0.0195	0.0016	1.0000

Table 4.2. Correlation between principal components and standardized variables

x	PC1	PC2	PC3	PC4
PPM	-.333218	0.146091	0.194914	0.118675
FAF	-.166576	-.539317	0.000217	-.024384
TT	0.202653	0.505909	0.056680	-.115743
FR	0.111124	0.512892	0.147625	-.057863
FA	0.293420	-.248280	0.414488	-.282509
IFS	0.380466	-.148433	0.119992	0.213348
ET	-.341491	0.231609	0.072179	-.090648
IA	-.225141	-.119583	0.395806	0.254624
IFD	0.338596	0.000655	0.466796	-.275830
AAF	-.439290	0.088861	0.107443	-.094088
DDA	0.067763	0.085509	0.353278	0.772074
DPD	0.312140	0.044331	-.484442	0.298605

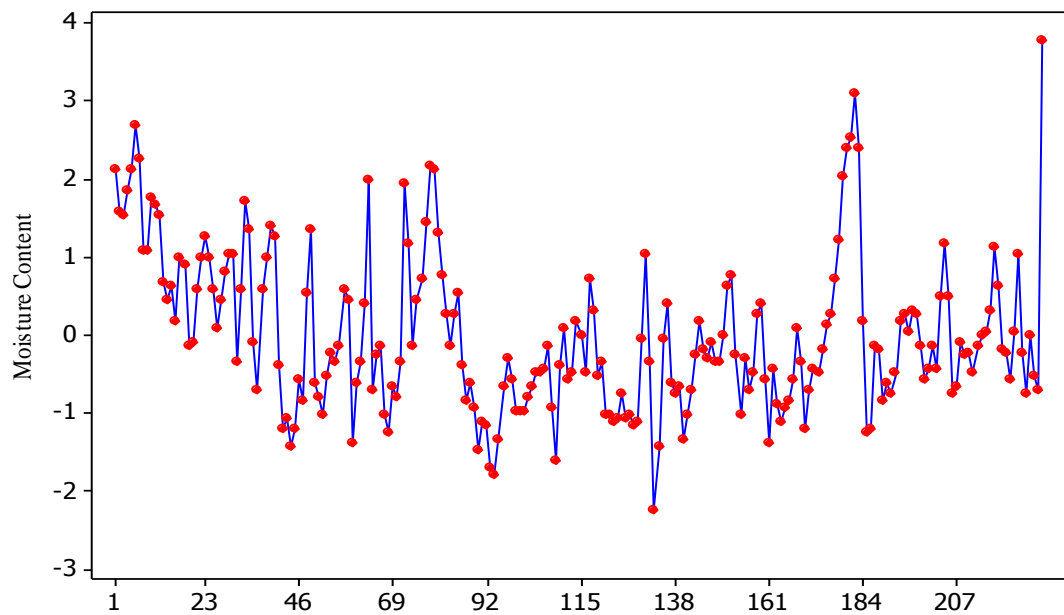


Figure 4.5. Time series plot of moisture content (output) for the pulp drying process

4.5. Performance of the Proposed Dual Scheme

We evaluated the performance of the proposed dual control scheme using simulated data similar to the observed data from the pulp drying process. The simulation study conducted to investigate the performance of the dual EWMA and MEWMA control scheme was based on 100,000 replications. The data were generated based on the same structure of actual pulp drying process. Performance of the dual control scheme was examined for changes in the mean, covariance, and structural relationship (Table 4.3). We evaluated performance in terms of out-of-control average run length (ARL_1), the number of points that must be plotted before an out-of-control signal is detected, for each chart as well as the joint ARL_1 , the minimum average run length for the dual monitoring scheme. In general, ARL_1 should be small when there are changes in the process corresponding to higher power of detection. The minimum ARL_1 and the probability that the out-of-control condition will be detected first by the MEWMA control chart provide information on which chart will yield an out-of-control signal. The EWMA chart is sensitive to structural and variance changes in all process inputs whereas the MEWMA is more sensitive to changes in the mean and possibly variance changes in X_1, X_2, X_3, X_4 .

For Scenario 1, the mean vector of the key process input variables being monitored by the MEWMA were simultaneously shifted by $\delta\boldsymbol{\mu} = \delta\mu_{x_1}, \dots, \delta\mu_{x_4}$ with $\delta = 0.25, 0.50, 1.00, 1.50, 2.00$ and 3.00 (Table 4.3). This scenario is used to assess the performance of the MEWMA chart directly. As expected unlike the EWMA chart, the MEWMA detects changes in the means $\delta\boldsymbol{\mu} = \delta\mu_{x_1}, \dots, \delta\mu_{x_4}$. The minimum ARL_1 represents the ARL based on joint control which will always be smaller than the individual ARL. Under in-control state, the joint ARL would represent the more frequent false alarm. The last column shows the proportion that the signal is almost certainly given by the MEWMA.

Table 4.3. Scenarios for studying the dual control scheme performances

	Process Input Variables			Structural Relationship
	$\mathbf{X} = (X_1, \dots, X_4)$		$\mathbf{X} = (X_5, \dots, X_{12})$	$\boldsymbol{\beta} = (\beta_1, \dots, \beta_4)$
	Changes in $\boldsymbol{\mu}$	Changes in $\boldsymbol{\Sigma}$	Changes in $\boldsymbol{\mu}$	Changes in β 's
Scenario 1	√			
Scenario 2			√	
Scenario 3	√		√	
Scenario 4		√		
Scenario 5				√
Scenario 6		√		√
Scenario 7	√	√		√

For scenario 2, while the mean vector of the key process input variables remain same, other variables shift by $\delta\boldsymbol{\mu} = \delta\mu_{x_5}, \dots, \delta\mu_{x_{12}}$ where $\delta = 0.25, 0.50, 1.00, 1.50, 2.00,$ and 3.00 . Changes in the means of Xs will cause the predicted moisture content to drift out of in-control state. So one would expect the signal to be given in the EWMA chart rather than the MEWMA. However, since the prediction is based on the same structural relationship, one would expect the predicted moisture content to be adjusted or accounted for by the current values of all Xs. With no changes in structural relationship or $\boldsymbol{\beta} = (\beta_1, \dots, \beta_4)$, changes in Xs leads to changes in the actual moisture content as well as the predicted value. The white noise extracted from the difference between the actual and the predicted moisture content still represents the pure white noise with the same mean at zero. The EWMA statistic is still a function of the pure white noise. Therefore we would expect both charts not to signal any structural changes or key process input variable changes. Table 4.5 summarizes the performance of the dual control scheme under scenario 2. As expected, none of the charts signal changes either in structural or

in key process input variables. The ARLs are simply the ARL_0 . Note that the joint ARL_1 becomes joint ARL_0 showing more frequent false alarm as a consequence of using joint control charts. Since none of the charts detect a change, the proportion or probability that a signal is detected by MEWMA or EWMA is 50%. The last column affirms that a false alarm is equally likely to be detected by either chart.

Table 4.4. Performances of the dual scheme under scenario 1 (standard deviation of the ARL_1 are in parenthesis)

δ	ARL_1 of MEWMA	ARL_1 of EWMA	Minimum ARL_1	Probability of First Detecting by MEWMA
0.25	19.0603 (15.4784)	494.1302 (427.7206)	18.3549 (15.0822)	0.9565 (0.2040)
0.50	3.5773 (1.7748)	489.1419 (422.1123)	3.5517 (1.7632)	0.9888 (0.1052)
1.00	1.0285 (0.1667)	491.4255 (426.1608)	1.0285 (0.1666)	1.0000 (0.0032)
1.50	1.0000	493.1678 (426.8112)	1.0000	1.0000
2.00	1.0000	493.1123 (426.5532)	1.0000	1.0000
3.00	1.0000	491.0701 (426.1637)	1.0000	1.0000

For scenario 3, the mean of all Xs has shifted by $\delta\boldsymbol{\mu} = \delta\mu_{x_1}, \dots, \delta\mu_{x_{12}}$ where $\delta = 0.25, 0.50, 1.00, 1.50, 2.00,$ and 3.00 . One would expect the signal to be given by both the MEWMA chart rather than the EWMA. Changes in the means of Xs will likely cause the actual moisture

content to drift out of the in-control state. As long as there is no change in structural relationship, the computed white noise still represents the pure white noise and no signal is given from the EWMA chart. In contrast, key process input variable changes will be noticed by the MEWMA. as shown in Table 4.6. In fact the results for scenario 1 and 3 are theoretically same. The last column shows that the MEWMA will detect the out-of-control signal.

Table 4.5. Performances of the dual scheme under scenario 2 (standard deviation of the ARL_1 are in parenthesis)

δ	ARL_1 of MEWMA	ARL_1 of EWMA	Minimum ARL_1	Probability of First Detecting by MEWMA
0.25	500.6007 (494.8401)	490.1605 (425.0081)	260.3702 (242.5750)	0.5128 (0.4998)
0.50	501.5634 (495.6951)	492.4464 (426.1649)	260.3858 (241.9645)	0.5131 (0.4998)
1.00	499.3547 (494.1531)	492.0592 (427.2602)	259.5897 (241.5918)	0.5152 (0.4998)
1.50	498.3966 (494.2622)	490.7534 (241.4834)	259.5816 (241.4834)	0.5167 (0.4997)
2.00	499.6636 (495.6035)	492.4494 (426.5958)	259.0994 (241.4450)	0.5146 (0.4998)
3.00	496.7262 (493.2899)	493.0368 (426.6435)	259.5923 (242.2439)	0.5186 (0.4997)

Table 4.6. Performances of the dual scheme under scenario 3 (standard deviation of the ARL₁ are in parenthesis)

δ	ARL ₁ of MEWMA	ARL ₁ of EWMA	Minimum ARL ₁	Probability of First Detecting by MEWMA
0.25	19.0120 (15.3732)	490.2401 (423.7091)	18.2816 (14.9554)	0.9552 (0.2068)
0.50	3.5803 (1.7748)	493.5123 (426.5843)	3.5559 (1.7622)	0.9896 (0.1015)
1.00	1.0296 (0.1699)	492.8393 (428.1624)	1.0296 (0.1699)	1.0000
1.50	1.0000	487.1425 (422.6859)	1.0000	1.0000
2.00	1.0000	491.8194 (425.1505)	1.0000	1.0000
3.00	1.0000	491.5078 (425.3678)	1.0000	1.0000

In scenario 4, the variances of the key process input variables have been simultaneously changed to $\delta\sigma_{x_1}^2, \dots, \delta\sigma_{x_4}^2$ where $\delta = 0.25, 0.50, 1.00, 1.50, 2.00$ and 3.00 . This scenario represents changes in the covariance structure of key process inputs. Changes in the covariance will change the principal component structure. Since the moisture content is related to latent variables through principal components, actual moisture content will deviate from the in-control state. With covariance changed, the computed principal components based on the estimated eigenvector from phase I will no longer represent the actual principal component. The predicted moisture content based on the principal component coefficient parameters from

phase I will differ from the actual one. Hence, the computed white noises will have different mean from the in-control state value of zero. The signal should be detected by the EWMA chart. However, changes in the covariance matrix will also impact the performance of the MEWMA. Table 4.7 shows the performance of the dual control scheme for scenario 4. As expected, the EWMA chart detects changes very fast compared to the MEWMA chart which is also affected by changes in both mean and covariance. The EWMA chart is significantly likely to detect small changes. As the magnitude of variance change gets larger, the MEWMA is more sensitive to the covariance changes. This can be seen from the ARL_1 and the probability that a signal is first detected by the MEWMA tends to increase as δ increases.

Table 4.7. Performances of the dual scheme under scenario 4 (standard deviation of the ARL_1 are in parenthesis)

δ	ARL_1 of MEWMA	ARL_1 of EWMA	Minimum ARL_1	Probability of First Detecting by MEWMA
0.25	38.6669 (34.7900)	2.9396 (3.7926)	2.8099 (3.4471)	0.0312 (0.1737)
0.50	15.8032 (12.8817)	2.7015 (3.3700)	2.4329 (2.6837)	0.0744 (0.2625)
1.00	7.3309 (5.3682)	2.3583 (2.7606)	1.9546 (1.7755)	0.1523 (0.3594)
1.50	4.9617 (3.4567)	2.1045 (2.3039)	1.6862 (1.3195)	0.2212 (0.4151)
2.00	3.8669 (2.6337)	2.0356 (2.1801)	1.5549 (1.1058)	0.2966 (0.4568)
3.00	2.7847 (1.8436)	4.4892 (6.2645)	1.7303 (1.2590)	0.5637 (0.4959)

Table 4.8. Performances of the dual scheme under scenario 5 (standard deviation of the ARL_1 are in parenthesis)

γ	ARL_1 of MEWMA	ARL_1 of EWMA	Minimum ARL_1	Probability of First Detecting by MEWMA
0.50	501.2508 (492.7092)	85.4207 (96.1684)	72.5400 (80.9204)	0.1377 (0.3446)
0.75	501.3622 (497.8904)	297.4333 (310.1911)	185.2244 (190.9700)	0.3634 (0.4810)
0.90	500.2131 (493.7417)	474.7670 (426.0455)	250.6095 (246.0354)	0.4954 (0.5000)
1.10	501.3524 (496.5488)	472.6846 (424.8884)	249.8301 (244.9613)	0.4929 (0.5000)
1.25	500.9720 (498.4079)	297.1395 (310.4359)	184.4976 (190.4891)	0.3638 (0.4811)
1.5	500.6268 (492.5258)	85.6666 (96.6659)	72.8204 (81.1120)	0.1350 (0.3417)

Scenario 5 represents the change in the structural relationship between the moisture content and the latent variables. The regression of parameter $\beta_1, \beta_2, \beta_3,$ and β_4 were changed to $\gamma\beta_1, \gamma\beta_2, \gamma\beta_3, \gamma\beta_4$ where $\gamma = 0.5, 0.75, 0.9, 1.1, 1.25,$ and 1.5 . We would expect only the EWMA chart to respond to these changes. The results in Table 4.8 agree with our analysis. As γ gets smaller or larger, the changes in $\beta_1, \beta_2, \beta_3,$ and β_4 are prominent enough to cause the predicted moisture content based on the in-control β_1, \dots, β_4 to be significantly different from the actual moisture content driven by the out-of-control β s. EWMA and MEWMA are equally

likely to detect an out of control signal for $\gamma = 0.9- 1.1$. However, MEWMA is less likely to detect small or large shifts compared to EWMA.

Scenario 6 represents the case where both covariance of key process inputs $\delta\sigma_{x_1}^2, \dots, \delta\sigma_{x_4}^2$ and the structural relationship $(\gamma\beta_1, \gamma\beta_2, \gamma\beta_3, \gamma\beta_4)$ are simultaneously changed. This concurrent change in covariance and regression parameters β should cause a significant change in the structural relationship. Hence the EWMA chart should be sensitive and give an out-of-control signal in the structural change faster than the MEWMA chart. This has been confirmed with the results shown in Table 4.9.

Table 4.9. Performances of the dual scheme under scenario 6 (standard deviation of the ARL_1 are in parenthesis)

δ	γ	ARL of MEWMA	ARL of EWMA	Minimum ARL_1	Probability of First Detecting by MEWMA
0.25	0.50	38.5370 (34.9076)	5.2215 (7.4383)	4.6157 (6.1379)	0.0755 (0.2642)
0.50	0.75	15.7900 (12.8968)	3.4805 (4.6861)	2.9416 (3.4250)	0.1076 (0.3099)
1.00	0.90	7.3189 (5.3478)	2.5568 (3.1088)	2.0564 (1.9231)	0.1661 (0.3722)
1.50	1.10	4.9581 (3.4553)	1.9833 (2.0776)	1.6305 (1.2439)	0.2081 (0.4059)
2.00	1.25	3.8681 (2.6340)	1.7792 (1.6907)	1.4626 (0.9669)	0.2677 (0.4428)
3.00	1.5	2.7939 (1.8494)	2.2944 (2.6555)	1.4724 (0.9551)	0.4539 (0.4539)

For the last scenario, there are changes in the means, variances of X_1, X_2, X_3, X_4 as well as changes in regression parameters, β s. This change should be detected by both charts as such a change will affect structure of the model. While both MEWMA and EWMA are able to detect the change, MEWMA is more sensitive to detecting changes as shown in Table 4.10. As illustrated by our performance study, the dual EWMA-MEWMA scheme provides double safeguard against changes from in-control to out-of-control conditions for the entire multivariate cascade process. The proposed dual scheme of monitoring both the product and the process characteristics is more effective than the conventional approach of only monitoring the product.

Table 4.10. Performances of the dual scheme under scenario 7 (standard deviation of the ARL_1 are in parenthesis)

δ	γ	ARL_1 of MEWMA	ARL_1 of EWMA	Minimum ARL_1	Probability of First Detecting by MEWMA
0.25	0.50	10.9540 (8.6405)	5.2383 (7.4395)	3.4526 (3.8515)	0.2259 (0.4182)
0.50	0.75	3.2609 (2.1214)	3.4542 (4.6413)	1.7695 (1.2758)	0.4251 (0.4944)
1.00	0.90	1.1680 (0.4465)	2.5285 (3.0409)	1.0666 (0.2784)	0.9132 (0.2815)
1.50	1.10	1.0043 (0.0681)	1.9577 (1.9844)	1.0015 (0.0397)	0.9973 (0.0518)
2.00	1.25	1.0001 (0.0084)	1.7618 (1.6416)	1.0000 (0.0055)	1.0000 (0.0063)
3.00	1.5	1.0000 0	2.2501 (2.5005)	1.0000 0	1.0000 0

4.6. Summary and Discussion

In this paper we developed a dual statistical process monitoring scheme for multivariate autocorrelated cascade processes where both process output and inputs are jointly controlled using a combination of residual-based EWMA and MEWMA control charts. The proposed methodology acquires information on the process inputs, evaluates whether the inputs are in-control state by control charting the correlated process inputs with the MEWMA control chart. In the meantime, process input information is used to predict and compare the output characteristic to the actual output. The difference between the actual and the predicted output indicates whether the operating behavior of the process is expected, accepted or in-control state. The difference is corrected for time-series autoregressive model and charted on the EWMA control chart derived from *GLS* PCR and AR parameters. This dual control scheme has improved performance in detecting out-of-control conditions either from the inputs of the process or from the process dynamic itself.

The dual EWMA-MEWMA control chart scheme provides the capability of monitoring the inputs, assessing and then controlling the output. The dual control scheme has advantages over conventional control chart schemes that rely only on the process output, e.g., residual-based control chart or principal component regression chart. The EWMA statistic is designed to signal either structural changes or changes in the dispersion and covariance of the process input variables. Whereas the MEWMA is designed to detect changes in the mean of the process inputs. Under seven different scenarios, the dual EWMA-MEWMA monitoring scheme has shown superior performance in detecting changes in the process. A comprehensive performance analysis used simulated data drawn from the characteristics of the sugar beet pulp drying process.

An advantage of the proposed dual scheme is its ability to monitor and maintain in-control process characteristics or input variables and hence maintain the output characteristic

at the desired level. In addition, the expected output can be regularly predicted and updated based on the observed realization of the process input values varying within suitable ranges. High variation in the inputs leads to high variation in the output. Monitoring and controlling the process inputs are necessary for effective control of the process. Therefore the simultaneous monitoring schemes of both process inputs and output presented in this paper leads to better control of both process and product characteristics.

In this paper we used PCA to alleviate the multicollinearity problem in the process input variables. The *GLS* PCR based on selected significant components were applied to establish the relationship model between process input variables and product characteristics. The PCR technique is a well-accepted tool for establishing the relationship between input and output variables for multivariate cascade processes. However, the conventional PCR is based on the assumption that the response is independently distributed. In this paper, we devised the PCR based on *GLS* that captures the autocorrelated cascade nature of the multivariate process. To implement this concept, two concurrent estimation stages must be performed; one for the regression and one for the autoregressive error. This approach has been employed extensively in time series model estimation. Our application of the *GLS*-based PCR for multivariate autocorrelated cascade processes is an intuitive and efficient methodology for monitoring the whole process.

The dual control scheme combined with exponential weight moving average control chart is robust and sensitive to small to medium shifts in the process. The dual monitoring scheme presented in this paper can be extended to monitor multivariate cascade processes where the relationship between the process inputs and output are nonlinear. Similarly, the sensitivity of the dual monitoring scheme to detect small process shifts can be enhanced by using the MCUSUM control chart. However, the residual-based control chart can always be employed once a process-product relationship is established. Overall, the dual EWMA-

MEWMA control chart is an effective scheme for monitoring and controlling multivariate cascade process.

CHAPTER 5. CONCLUSIONS AND FUTURE RESEARCH

5.1. Conclusion

This dissertation developed a dual statistical process monitoring scheme for the multivariate autocorrelated cascade processes where both process output and inputs are jointly controlled using a combination of the residual-based EWMA and MEWMA control chart. In general one can enhance monitoring of both output and inputs by incorporating the information on the process input variables to successively forecast the expected value of the output so that the control limits for the output can be determined and updated. Typically if predictor (the process variables) are altered or changed, the response (the output variable) will be directly affected. Under the controlled condition, the process input variables are at the in-control state where the joint mean and covariance are both at in-control values. When the process is in-control state with no change in the relationship between Y and X_s , process output can be predicted corresponding to the actual current input variable values. The actual process output can then compared and charted against predicted value within a predefined range.

Therefore this dissertation devised a comprehensive methodology to control the overall process by acquiring information on the process inputs and evaluating whether process inputs are in-control state by control charting the correlated process inputs with the MEWMA control chart. In the meantime, the process input information is used to project the output characteristic so that the actual output can be monitored. The difference between actual and predicted output provides information for assessing evidence whether the operating behavior of the process is the same or being different from the expected, accepted or in-control state. This difference is corrected for time series autoregressive model and charted on the EWMA control chart derived from generalized least square, principal component regression and autoregressive parameters. This dual control scheme has improved performance in detecting the out-of-control condition either from the inputs of the process or from the process dynamic itself.

The dual EWMA-MEWMA control chart scheme provides the capability of monitoring the inputs, assessing and then controlling the output. This dual control provides the advantages over conventional control chart schemes that focus only on the output side, e.g., residual-based control chart or principal component regression chart. The EWMA statistic is designated to signal any changes in either the structural change or the changes in the dispersions and covariance of the process input variables whereas the MEWMA is responsible for detecting mainly the change in the mean of the process inputs. Under seven different scenarios, the dual EWMA-MEWMA control scheme has shown superior performance in detecting changes and shifts in the process. The performance analysis has been conducted based on simulated data drawn from the sugar beet dry pulping process. The simulation study affirms that the desired performance of the dual control scheme is achieved as expected.

This research concept leads to several advantages in that first the process characteristics or input variables can be monitored and maintained at the in-control state which will sequentially assist in maintaining the output characteristic at the desired level. In addition, the expected output can be regularly predicted and updated based on the observed realization of the process input values varying within suitable ranges. Higher variations of the inputs lead to higher variation of the output. Monitoring and controlling the inputs are necessary for effective quality control. Therefore the simultaneous monitoring scheme of both inputs and output presented in this paper is intuitive and leads to better control of both the process and the product characteristics.

5.2. Discussion and Future Research Direction

This research proposed using the generalized least squares and principal component regression technique to measure the relationship between process inputs and outputs based on selected significant principal components. This research used the principal component analysis to alleviate the multicollinearity problem among process input variables. This approached

established the relationship model between the product characteristic and the process variables. The principal component analysis technique is a well-accepted tool for establishing the relationship between process input and output variables of a multivariate cascade process. Hence the main mechanism of this proposed methodology can be validated. However, the conventional principal component analysis will rely on the assumption that the response is independently distributed. In this research, we devised the principal component analysis based on generalized least squares that can capture the relationship structure of autocorrelated cascade process. To implement this concept, two concurrent estimation stages must be performed; one for the regression and one for the autoregressive error. This has been employed extensively in time series model estimation. Our application of the generalized least squares based principal component regression for multivariate autocorrelated cascade process is a well constructive and efficient methodology for monitoring the whole process. However another technique such as multivariate partial least squares might also be used to derive the relationship by accounting for the multicollinearity of the input variables. In our future work, we will devise our methodology around the partial least squares and compare with the current generalized least squares based principal component regression to see difference which one is more superior to the multivariate cascade process.

This research increases the detection power of the dual control chart by using exponential weight moving averages that are robust and sensitive to small to medium shifts. In our future research, different schemes such as MCUSUM might also be studied and compared with the MEWMA. Lastly if there are nonlinear relationships between process inputs and output, then different types of regression should be adopted. However the residual-based control chart can always be employed once the relationship is established. Therefore the dual EWMA-MEWMA control chart is an overall effective scheme to monitor and control the multivariate cascade process. Therefore one could adopt and compare our results with another

sensitive control chart such as CUSUM and MCUSUM. So our future work will study with the method to enhance the effectiveness of the dual control scheme.

REFERENCES

- Aparisi, F. (1996) 'Hotelling's T^2 control chart with adaptive sample sizes', *International Journal of Production Research*, Vol.34, Issue.10, pp.2853–2862.
- Aparisi, F., Champ, C.W. and Garcia, Diaz J.C. (2004) 'A performance Hotelling's T^2 control chart with supplementary run rules', *Quality Engineering*, Vol.16, Issue.3, pp.359–368.
- Apley, D. W. and Lee, H. C. (2008) 'Robustness Comparison of Exponential weighted moving-average charts on autocorrelated data and on residuals' *Journal of Quality Technology*, Vol. 40, No. 4, pp.428–447.
- Arkat, J., Niaki, S.T.A. and Abbasi, B. (2007) 'Artificial neural networks in applying MCUSUM residuals charts for AR(1) processes', *Applied Mathematics and Computation*, Vol.189, Issue.2, pp.1889-1901.
- Badcock, J., Bailey, T.C., Jonathan, P. and Krzanowski, W.J. (2004) 'Two projection methods for use in the analysis of multivariate process data with an illustration in petrochemical production', *Technometrics*, Vol.46, Issue.4, pp.392–403.
- Bersimis, S., Psarakis S., and Panaretos, J. (2007) 'Multivariate statistical process control chart: An overview', *Quality and Reliability Engineering International*, Vol.23, Issue.5, pp.517-543.
- Bilen, C., Chen, X., Khan, A. and Yadav, O.P. (2007) 'Multiple regression control chart integrated with principal component analysis', *Proceedings of the Industrial Engineering Research Conference*, 19–23 May, Nashville, Tennessee.
- Bilen, C., Khan, A. and Yadav, O.P. (2009) 'Multivariate autocorrelated process monitoring with multiple regression and principal components', *Proceedings of the Industrial Engineering Research Conference*, 30 May–3 June, Miami, Florida.
- Bilen, C., Khan, A., and Yadav, O. P. (2010) 'Principal components regression control for multivariate autocorrelated cascade processes', *International Journal of Quality Engineering and Technology*, Vol. 1, No. 3, pp.301-315.
- Bisgaard, S. and Kulahci, M. (2005) 'Quality quandaries: the application of principal component analysis for process monitoring', *Quality Engineering*, Vol. 18, No. 1, pp.95–103.
- Bodden, K.M. and Rigdon, S.E. (1999) 'A program for approximating the in-control ARL for the MEWMA chart', *Journal of Quality Technology*, Vol.31, No.1, pp.120–123.
- Borror, C. M., Montgomery D. C. and Runger, G. C. and (2003) 'Robustness of the EWMA control chart to non-normality' *Journal of Quality Technology*, Vol. 31, No. 3, pp.309–316.

- Carson, P. K. and Yeh, A. B. (2008) 'Exponential weighted moving average (EWMA) control charts for monitoring analytical process' *Industrial and Engineering Chemistry Research*, Vol. 47, No. 2, pp.405–411.
- Chan, L.K. and Li, G.Y. (1994) 'A multivariate control chart for detecting linear trends', *Communications in Statistics-Simulation and Computation*, Vol. 23, No. 4, pp.997–1012.
- Charnes, J.M. (1995) 'Tests for special causes with multivariate autocorrelated data', *Computers and Operational Research*, Vol. 22, No. 4, pp.443–453.
- Chatfield, C. (2000) *The Analysis of Time Series: An Introduction*, 5th ed., Chapman and Hall/CRC, Florida.
- DiPaola, P.P. (1945) 'Use of correlation in quality control', *Industrial Quality Control*, Vol. 2, No. 1, pp.10–14.
- Feltz, C.J. and Shiau, J.J.H. (2001) 'Statistical process monitoring using an empirical Bayes multivariate process control chart', *Quality and Reliability Engineering International*, Vol.17, Issue.2, pp.119–124.
- Hawkins, D.M. (1991) 'Multivariate quality control based on regression-adjusted variables', *Technometrics*, Vol. 33, Issue. 1, pp.61-75.
- Hawkins, D.M. (1993) 'Regression adjustment for variables in multivariate quality control', *Journal of Quality Technology*, Vol. 25, No. 3, pp.170–182.
- Hawkins, D.M., Choi, S.C. and Lee, S. (2007) 'A general multivariate exponential weighted moving-average control chart', *Journal of Quality Technology*, Vol.39, No.2, pp.118-125.
- Healy, J.D. (1987) 'A note on multivariate CUSUM procedures', *Technometrics*, Vol.29, Issue.4, pp. 409-412.
- Holgersson, H.E.T. (2004) 'Testing for multivariate autocorrelation', *Journal of Applied Statistics*, Vol.31, No.4, pp.379-395.
- Horst, P. (1961) 'Relations among m sets of measures', *Psychometrika*, Vol. 26, No. 2, pp. 129-149.
- Hotelling, H. (1947) 'Multivariate quality control-illustrated by the air testing of sample bombsights'. In: Eisenhart, C., Hastay, M.W., and Wallis, W.A. (eds), *Techniques of Statistical Analysis* (pp. 111-184) New York: Mcgraw Hill.
- Human, S. W., Kritzing, P. and Chakraborti, S. (2011) 'Robustness of EWMA control chart for individual observations' *Journal of Applied Statistics*, Vol. 38, No. 10, pp.2071–2087.

- Hwang, H.B. and Wang, H. (2010) 'Shift detection and source identification in multivariate autocorrelated processes', *International Journal of Production Research*, Vol.48, Issue.3, pp.835–859.
- Issam, B.K. and Mohamed, L. (2008) 'Support vector regression based residual MCUSUM control chart for autocorrelated process', *Applied Mathematics and Computation*, Vol.201, Issue.1-2, pp.565-574.
- Jackson, J.E. (1980) 'Principal components and factor analysis: Part I – principal components', *Journal of Quality Technology*, Vol. 12, No. 4, pp.201–213.
- Jackson, J.E. (1985) 'Multivariate quality control', *Communications in Statistics-Theory and Methods*, Vol.14, Issue.11, pp.2657-2688.
- Jarrett, J.E. and Pan X. (2007) 'The quality control chart for monitoring multivariate autocorrelated processes', *Computational Statistics and Data Analysis*, Vol. 51, No. 8, pp.3862–3870.
- Jiang, W. (2004) 'Multivariate control charts for monitoring autocorrelated processes', *Journal of Quality Technology*, Vol.36, No.4, pp.367–379.
- Jolliffe, I.T. (1972) 'Discarding variables in a principal component analysis. I: artificial data', *Applied Statistics*, Vol. 21, No. 2, pp.160–173.
- Jolliffe, I.T. (1973) 'Discarding variables in a principal component analysis. II: real data', *Applied Statistics*, Vol. 22, No. 1, pp.21–31.
- Jolliffe, I.T. (2002) *Principal Component Analysis*, 2nd ed., Springer, New York.
- Kalgonda, A.A. and Kulkarni, S.R. (2004) 'Multivariate quality control chart for autocorrelated processes', *Journal of Applied Statistics*, Vol. 31, No. 3, pp.203-215.
- Kano, M., Hasebe, S., Hashimoto, I. and Ohno, H. (2001) 'A new multivariate statistical monitoring method using principal component analysis', *Computers and Chemical Engineering*, Vol.25, Issue.7-8, pp.1103–1113.
- Khoo, M. (2003) 'Increasing the sensitivity of multivariate EWMA control chart', *Quality Engineering*, Vol.16, Issue.1, pp.75-85.
- Kim, K., Mahmoud, M.A. and Woodall, W.H. (2003) 'On the monitoring of linear profiles', *Journal of Quality Technology*, Vol.35, No.3, pp.317–328.
- Koehler, A. B., Mark, N. B. and O'Connell R. T. (2001) 'EWMA control charts for autoregressive processes' *Journal of the Operation Research Society*, Vol. 52, No. 6, pp.699–707.

- Kourti, T. (2005) 'Application of latent variable methods to process control and multivariate statistical process control in industry', *International Journal of Adaptive Control and Signal Processing*, Vol. 19, No. 4, pp. 213-246.
- Kourti, T. and MacGregor, J.F. (1996) 'Multivariate SPC methods for process and product monitoring', *Journal of Quality Technology*, Vol. 28, No. 4, pp. 409-428.
- Kourti, T., Lee, J. and MacGregor, J.F. (1996) 'Experience with industrial application of projection methods for multivariate statistical process control', *Computers and Chemical Engineering*, Vol. 20, Supplement 1, pp. S745-S750.
- Koutras, M.V., Bersimis, S. and Antzoulakos, D.L. (2006) 'Improving the performance of the chi-square control chart via runs rules', *Methodology and Computing in Applied Probability*, Vol.8, Issue.3, pp.409-426.
- Kresta, J.V., MacGregor, J.F. and Marlin, T.E. (1991) 'Multivariate statistical monitoring of process operating performance', *The Canadian Journal of Chemical Engineering*, Vol. 69, No. 1, pp. 35-47.
- Ku, W., Storer, R.H. and Georgakis, C. (1995) 'Disturbance detection and isolation by dynamic principal component analysis', *Chemometrics and Intelligent Laboratory Systems*, Vol.30, Issue.1, pp.179-196.
- Loredo, E.N., Jearkraporn, D. and Borrer, C.M. (2002) 'Model-based control chart for autoregressive and correlated data', *Quality and Reliability Engineering International*, Vol. 18, No. 6, pp.489-496.
- Lowry, C.A. and Montgomery, D.C. (1995) 'A review of multivariate control charts', *IIE Transactions*, Vol.27, Issue.6, pp.800-810.
- Lowry, C.A., Woodall, W. H., Champ, C. W. and Rigdon, S. E. (1992) 'A multivariate exponential weighted moving average control chart', *Technometrics*, Vol. 34, No. , pp.46-53.
- Lu, C. W. (1999) 'EWMA control charts for monitoring the mean of autocorrelated processes' *Journal of Quality Technology*, Vol. 31, No. 2, pp.166-188.
- MacGregor, J.F. and Kourti, T. (1995) 'Statistical process control of multivariate processes', *Control Engineering Practice*, Vol. 3, No. 3, pp. 403-414.
- Macgregor, J.F., Jaeckle, C., Kiparissides, C. and Koutoudi, M. (1994) 'Process monitoring and daiganosis by multiblock PLS methods', *American Institute Chemical Engineers Journal*, Vol. 40, No.5, pp. 826-838.
- Makridakis, S., Wheelwright, S.C. and McGee, V.E. (1983) *Forecasting Methods and Applications*, 2nd ed., Wiley, New York.
- Mandel, B.J. (1969) 'The regression control chart', *Journal of Quality Technology*, Vol. 1, No. 1, pp.1-9.

- Martin, E.B., Morris, A.J. and Kiparissides, C. (1999) 'Manufacturing performance enhancement through multivariate statistical process control', *Annual Reviews in Control*, Vol. 23, pp. 35-44.
- Mason, R.L. and Young, J.C. (2007) 'Detecting dependent observations in multivariate statistical process control', *Quality Progress*, Vol.40, No.9, pp.56-58.
- Mason, R.L., Chou, Y-M., and Young, J.C. (2001) 'Applying Hotelling's T^2 statistic to batch processes', *Journal of Quality Technology*, Vol.33, No.4, pp.466-479.
- Mastrangelo, C.M., Runger, G.C. and Montgomery, D.C. (1996) 'Statistical process monitoring with principle components', *Quality and Reliability Engineering International*, Vol. 12, No. 3, pp.203-210.
- Midi, H. and Shabbak, A. (2011) 'Robust multivariate control charts to detect small shifts in mean' *Mathematical Problems in Engineering*, Vol. 2011, Hindawi Publishing Corporation, pp.1-19.
- Molnau, W.E., Runger, G.C., Montgomery, D.C., Skinner, K.R., Lored, E.N. and Prabhu, S.S. (2001) 'A program for ARL calculation for multivariate EWMA charts', *Journal of Quality Technology* Vol.33, No.4, pp.515-521.
- Montgomery, D. C. (2012) *Introduction to Statistical Quality Control*, 7th Edition, John Wiley and Sons.
- Nedumaran, G. and Pignatiello, J.J. (2000) 'On constructing T^2 control charts for retrospective examination', *Communications in Statistics—Simulation and Computation*, Vol.29, Issue.2, pp.621-632.
- Neubauer, A. S. (1997) 'The EWMA control chart: properties and comparison with other quality-control procedures by computer simulation' *Clinical Chemistry*, Vol. 43, No. 4, pp.594-601.
- Nomikos, P. and MacGregor, J.F. (1995) 'Multivariate SPC Charts for Monitoring batch processes', *Technometrics*, Vol.37, No.1, pp.41-59.
- Noorossana, R. and Vaghefi S.J.M. (2006) 'Effect of autocorrelation on performance of the MCUSUM control chart', *Quality and Reliability Engineering International*, Vol. 22, No. 2, pp.191-197.
- Prabhu, S. S. and Runger, G.C. (1997) 'Designing a multivariate EWMA control chart' *Journal of Quality Technology*, Vol. 29, No. 1, pp.8-15.
- Runger, G.C. (1996) 'Multivariate statistical process control for autocorrelated processes', *International Journal of Production Research*, Vol.34, Issue.6, pp.715-1724.
- Runger, G.C. (1996) 'Projections and the U^2 multivariate control chart', *Journal of Quality Technology*, Vol.28, No.3, pp.313-319.

- Runger, G.C. and Testik, M.C. (2004) 'Multivariate extensions to cumulative sum control charts', *Quality and Reliability Engineering International*, Vol. 20, Issue. 6, pp. 587-606.
- Runger, G.C., Keats, J.B., Montgomery, D.C. and Scranton, R.D. (1999) 'Improving the performance of a multivariate EWMA control chart', *Quality and Reliability Engineering International*, Vol.15, Issue.3, pp.161–166.
- Schall, S. and Chandra, J. (1987) 'Multivariate quality control using principal components', *International Journal of Production Research*, Vol.25, Issue.4, pp.571-588.
- Scholkopf, B. and Smola, A (1998) 'A tutorial on support vector regression', NeuroCOLT Technical Report TR 1998-030, Royal Holloway College.
- Scranton, R., Runger, G. C., Keats, J. B. and Montgomery, D. C. (1996) 'Efficient shift detection using multivariate exponential-weighted moving average control charts and principal components', *Quality and Reliability Engineering International*, Vol. 12, No. 3, pp.165–171.
- Simoglou, A., Martin, E.B. and Morris, A.J. (2000) 'Multivariate statistical process control of an industrial fluidized-bed reactor', *Control Engineering Practice*, Vol. 8, No. 8, pp. 893-909.
- Stoumbos, Z. and Sullivan, J.H. (2002) 'Robustness to non-normality of the multivariate EWMA control chart', *Journal of Quality Technology*, Vo.34, No.3, pp.260-276.
- Sullivan, J.H. and Woodall, W.H. (1996) 'A comparison of multivariate control charts for individual observations', *Journal of Quality Technology*, Vol.28, No.4, pp.398–408.
- Sullivan, J.H. and Woodall, W.H. (2000) 'Change-point detection of mean vector or covariance matrix shifts using multivariate individual observations', *IIE Transactions*, Vol.32, Issue.6, pp.537–549. *Technometrics*, Vol. 33, No. 1, pp.61–75.
- Testik, M. C., Runger, G. C. and Borrer, C. M. (2003) 'Robustness properties of multivariate EWMA control charts' *Quality and Reliability Engineering International*, Vol. 19, No. 1, pp.31–38.
- Triantafyllopoulos, K. (2006) 'Multivariate control charts based on Bayesian state space models', *Quality and Reliability Engineering International*, Vol.22, Issue.6, pp.693-707.
- Tsung, F. (1999) 'Improving automatic-controlled process quality using adaptive principal component monitoring', *Quality and Reliability Engineering International*, Vol.15, Issue.2, pp.135–142.
- Vargas, J.A. (2003) 'Robust estimation in multivariate control charts for individual observations', *Journal of Quality Technology*, Vol.35, No.4, pp.367–376.

- Vargas, M., Alfaro, J.L. and Mondejar, J. (2009) 'On the run length of a state-space control chart for multivariate autocorrelated data', *Communications in Statistics-Simulation and Computation*, Vol.38, Issue.9, pp.1823-1833
- Wang, X., Kruger, U. and Lennox, B. (2003) 'Recursive partial least squares algorithms for monitoring complex industrial processes', *Control Engineering Practice*, Vol. 11, No. 6, pp. 613-632.
- Wikstrom. C., Albano, C., Eriksson, L., Friden, H., Johansson, E., Nordahl, A., Rannar, S., Sandberg, M., Kettaneh-Wold, N. and Wold, S. (1998) 'Multivariate process and quality monitoring applied to an electrolysis process: Part I. Process supervision with multivariate control charts', *Chemometrics and Intelligent Laboratory Systems*, Vol.42, Issue.1-2, pp.233-240.
- Woodall, W.H. and Montgomery, D.C. (1999) 'Research issues and ideas in statistical process control', *Journal of Quality Technology*, Vol. 31, No. 4, pp.376-386.
- Woodall, W.H. and Ncube, M.M. (1985) 'Multivariate CUSUM quality control procedures', *Technometrics*, Vol.27, Issue.3, pp. 285-292.
- Woodall, W.H., Spitzner, D.J., Montgomery, D.C. and Gupta, S. (2004) 'Using control charts to monitor process and product quality profiles', *Journal of Quality Technology*, Vol. 36, No.3, pp.309-320.

APPENDIX A. DRY PULP PROCESS DATA

Data	PPM	FAF	TT	FR	FA	IFS	ET	IA	IFD	AAF	DDA	DPD	DPM
1	73.95	0.046	715.33	21.008	16.82	620.41	105.25	111.25	5.004	1.04	23.523	0.332	11.75
2	73.6	0.047	710.99	21.82	16.835	620.41	105.15	112.71	5.019	1.038	23.014	0.331	11.15
3	73.05	0.056	687.97	21.776	16.849	620.41	105.24	112.57	5.034	1.036	24.862	0.33	11.1
4	72.5	0.076	652.03	21.329	16.864	620.41	105.33	112.15	5.05	1.034	25.376	0.329	11.45
5	71.9	0.06	691.46	20.721	16.878	620.41	105.42	111.72	5.065	1.033	23.535	0.328	11.75
6	71.3	0.058	676.46	22.176	16.892	620.41	105.52	111.3	5.081	1.031	23.643	0.327	12.4
7	71.45	0.069	664.19	21.726	16.907	620.41	105.61	110.87	5.096	1.029	22.77	0.326	11.9
8	71.6	0.08	670.21	21.935	16.921	620.42	105.7	110.45	5.111	1.027	24.377	0.325	10.6
9	72.4	0.054	707.22	21.818	16.935	620.42	105.8	110.03	5.127	1.026	25.502	0.324	10.6
10	73.2	0.051	699.77	21.759	16.95	620.42	105.89	109.6	5.142	1.024	24.346	0.323	11.35
11	72.5	0.08	664.38	22.155	16.964	620.42	105.98	109.75	5.157	1.022	24.842	0.322	11.25
12	71.8	0.057	712.99	22.56	16.978	620.42	106.12	110.28	5.173	1.02	24.224	0.321	11.1
13	72.05	0.072	690.63	21.21	16.993	620.42	105.73	110.57	5.188	1.018	23.135	0.32	10.15
14	72.3	0.042	771.04	23.667	17.007	620.42	106.5	110.43	5.204	1.017	24.441	0.319	9.9
15	72.15	0.044	748.01	23.495	17.022	620.42	106.45	110.3	5.219	1.015	23.749	0.318	10.1
16	72	0.046	763.74	21.789	17.036	620.42	106.41	110.16	5.234	1.013	25.021	0.317	9.6
17	72.15	0.042	729.84	22.601	17.05	620.42	106.37	110.03	5.25	1.011	26.343	0.316	10.5
18	72.3	0.064	679.85	22.826	17.065	620.42	106.33	109.89	5.265	1.01	23.987	0.315	10.4
19	72.45	0.051	725.65	22.86	17.079	620.42	106.28	109.76	5.28	1.008	23.657	0.314	9.25
20	72.6	0.052	728.09	22.999	17.093	620.43	106.24	109.62	5.296	1.006	24.073	0.313	9.3
21	72.1	0.045	716.57	22.971	17.108	620.43	106.2	109.49	5.311	1.004	23.508	0.312	10.05
22	71.6	0.051	722.18	22.909	17.122	620.43	106.16	109.35	5.326	1.002	23.822	0.311	10.5
23	72.1	0.044	735.2	22.956	17.136	620.43	106.11	109.22	5.342	1.001	24.655	0.31	10.8
24	72.6	0.044	758.32	22.935	17.151	620.43	106.07	109.08	5.357	0.999	25.582	0.309	10.5
25	72.4	0.018	819.64	23.042	17.165	620.43	106.03	108.95	5.373	0.997	24.236	0.308	10.05
26	72.2	0.047	746.52	22.405	17.18	620.43	105.99	108.81	5.388	0.995	23.319	0.307	9.5
27	72.45	0.047	732.13	21.858	17.194	620.43	105.94	108.68	5.403	0.994	23.572	0.306	9.9
28	72.7	0.051	730.33	22.526	17.208	620.43	105.9	108.54	5.419	0.992	22.742	0.305	10.3
29	72.45	0.057	713.32	22.789	17.223	620.43	105.86	108.4	5.434	0.99	24.51	0.304	10.55
30	72.2	0.054	729.92	22.386	17.237	620.43	105.81	108.27	5.449	0.988	24.695	0.303	10.55
31	72.1	0.045	748.6	22.049	17.251	620.43	105.77	108.13	5.465	0.987	23.31	0.302	9
32	72	0.059	708.11	22.817	17.266	620.44	105.73	108	5.48	0.985	24.011	0.301	10.05
33	72	0.057	711.57	22.603	17.28	620.44	105.69	107.8	5.496	0.983	24.361	0.3	11.3

Data	PPM	FAF	TT	FR	FA	IFS	ET	IA	IFD	AAF	DDA	DPD	DPM
34	72	0.058	713.52	23.007	17.294	620.44	105.64	107.36	5.511	0.981	22.931	0.299	10.9
35	71.35	0.042	761.13	22.49	17.309	620.44	105.6	106.93	5.526	0.979	21.702	0.298	9.3
36	70.7	0.037	765.3	22.436	17.323	620.44	105.56	106.5	5.542	0.978	21.654	0.297	8.6
37	70.55	0.067	689.41	22.781	17.338	620.44	106.5	106.06	5.557	0.976	22.849	0.296	10.05
38	70.4	0.059	712.94	22.576	17.352	620.44	105.73	105.63	5.572	0.974	24.228	0.295	10.5
39	70.15	0.059	706.76	22.514	17.366	620.44	107.35	105.2	5.588	0.972	23.499	0.294	10.95
40	69.9	0.063	703.41	21.507	17.381	620.44	106.69	104.94	5.603	0.971	21.958	0.293	10.8
41	70.2	0.074	719.98	22.445	17.395	620.44	105.53	105.13	5.619	0.969	21.934	0.292	8.95
42	70.5	0.027	801.69	23.053	17.409	620.44	108.23	105.33	5.634	0.967	22.917	0.291	8.05
43	70.15	0.071	705.41	22.661	17.424	620.44	107.89	105.52	5.649	0.965	24.284	0.29	8.2
44	69.8	0.056	743.11	23.047	17.438	620.45	105.77	105.72	5.665	0.964	25.049	0.289	7.8
45	69.75	0.051	761.48	23.248	17.452	620.45	106.34	105.92	5.68	0.962	23.798	0.288	8.05
46	69.7	0.024	804.24	23.014	17.467	620.45	105.57	106.11	5.695	0.96	24.093	0.287	8.75
47	69.85	0.047	768.99	22.848	17.481	620.45	104.79	106.31	5.711	0.958	23.162	0.286	8.45
48	70	0.08	685.38	22.995	17.496	620.45	104.59	106.5	5.726	0.956	22.071	0.285	10
49	71.2	0.06	714.41	22.729	17.51	620.45	104.51	106.7	5.742	0.955	23.636	0.284	10.9
50	72.4	0.051	757.38	23.296	17.524	620.45	104.42	106.9	5.757	0.953	23.708	0.283	8.7
51	71.45	0.061	730.97	23.197	17.539	620.45	104.34	107.09	5.772	0.951	23.347	0.282	8.5
52	70.5	0.027	803.46	23.467	17.553	620.45	104.26	107.29	5.788	0.949	23.355	0.281	8.25
53	70.25	0.034	782.71	23.609	17.567	620.45	104.17	107.48	5.803	0.948	24.167	0.28	8.8
54	70	0.048	765.9	22.837	17.582	620.45	104.09	107.68	5.818	0.946	23.551	0.279	9.15
55	70.7	0.042	766.58	22.638	17.596	620.45	104.01	107.87	5.834	0.944	22.714	0.278	9
56	71.4	0.043	743.33	23.397	17.61	620.46	103.92	108.07	5.849	0.942	24.181	0.277	9.25
57	72.1	0.055	713.38	23.382	17.625	620.46	103.84	108.27	5.865	0.941	25.394	0.276	10.05
58	72.8	0.067	700.31	23.343	17.639	620.46	103.76	108.46	5.88	0.939	23.883	0.275	9.9
59	72.4	0.046	793.52	23.843	17.654	620.46	103.66	108.66	5.895	0.937	22.001	0.274	7.85
60	72	0.008	838.79	23.361	17.668	620.46	103.41	108.85	5.911	0.935	22.833	0.273	8.7
61	71.05	0.026	777.09	22.549	17.682	620.46	103.16	109.05	5.926	0.933	24.128	0.272	9
62	70.1	0.072	697.51	20.543	17.697	620.46	102.91	109.25	5.941	0.932	22.773	0.274	9.85
63	70.5	0.104	600.64	16.917	17.711	620.46	90.878	113	5.463	0.93	20.921	0.28	11.6
64	70.9	0.207	484.78	13.869	18.46	620.46	104.68	109.61	5.007	0.928	23.099	0.287	8.6
65	70.65	0.115	602.14	19.83	17.268	620.46	104.41	109.26	5.012	0.926	23.029	0.294	9.1
66	70.4	0.099	605.58	19.412	17.201	620.46	104.15	108.9	5.018	0.925	22.343	0.301	9.25
67	69.85	0.067	654.31	20.741	17.444	620.46	103.89	108.5	5.023	0.923	23.78	0.307	8.25
68	69.3	0.113	583.24	17.979	17.686	620.47	103.61	108	5.029	0.921	22.781	0.314	8

Data	PPM	FAF	TT	FR	FA	IFS	ET	IA	IFD	AAF	DDA	DPD	DPM
69	69.3	0.093	658.45	18.092	17.575	620.47	103.28	107.49	5.035	0.919	22.176	0.321	8.65
70	69.3	0.059	695.78	21.923	16.878	620.47	102.95	106.98	5.04	0.917	21.903	0.327	8.5
71	70.9	0.058	696.32	21.737	16.884	620.47	102.73	106.47	5.046	0.916	22.224	0.334	9
72	72.5	0.055	691.69	22.54	16.889	620.47	102.84	105.96	5.052	0.914	22.592	0.341	11.55
73	72.15	0.045	731.97	22.142	16.895	620.47	102.95	105.45	5.057	0.912	22.567	0.347	10.7
74	71.8	0.054	720.54	21.908	16.9	620.47	103.06	104.95	5.063	0.91	23.992	0.354	9.25
75	71.65	0.055	727.41	22.261	16.906	620.47	103.17	104.44	5.069	0.909	24.166	0.361	9.9
76	71.5	0.055	692.42	21.964	16.911	620.47	103.27	104.21	5.074	0.907	23.508	0.367	10.2
77	72	0.054	716.06	21.721	16.917	620.47	103.38	104.26	5.08	0.905	23.73	0.374	11
78	72.5	0.042	743.77	22.139	16.922	620.47	103.49	104.31	5.086	0.903	23.269	0.381	11.8
79	72	0.047	735.82	21.988	16.928	620.47	103.6	104.36	5.091	0.902	21.35	0.387	11.75
80	71.5	0.05	729.38	21.634	16.934	620.48	103.71	104.41	5.097	0.9	20.411	0.394	10.85
81	72.05	0.05	728.4	22.193	16.939	620.48	103.82	104.47	5.103	0.898	22.356	0.401	10.25
82	72.6	0.043	743.43	21.815	16.945	620.48	103.93	104.52	5.108	0.896	23.857	0.407	9.7
83	71.7	0.048	713.78	22.01	16.95	620.48	104.05	104.57	5.114	0.894	24.155	0.414	9.25
84	70.8	0.039	754.15	21.534	16.956	620.48	104.16	104.62	5.12	0.893	24.388	0.421	9.7
85	70.8	0.042	742.39	21.705	16.961	620.48	104.27	104.67	5.125	0.891	23.686	0.427	10
86	70.8	0.048	720.49	21.912	16.967	620.48	104.38	104.72	5.131	0.889	23.45	0.434	8.95
87	70.75	0.049	755.34	22.074	16.972	620.48	104.49	104.78	5.137	0.887	23.276	0.441	8.45
88	70.7	0.047	733.14	22.353	16.978	620.48	104.6	104.83	5.142	0.886	24.278	0.447	8.7
89	71.3	0.049	736.62	21.711	16.983	620.48	103.82	104.88	5.148	0.884	24.448	0.45	8.35
90	71.9	0.051	705.64	21.617	16.989	620.48	104.09	104.93	5.154	0.882	22.166	0.45	7.75
91	71.15	0.059	700.41	22.05	16.995	620.49	103.77	104.98	5.159	0.88	21.092	0.45	8.15
92	70.4	0.051	727.5	22.063	17	620.49	103.82	105.03	5.165	0.879	21.556	0.45	8.1
93	70.15	0.049	718.58	21.931	17.006	620.49	103.88	105.09	5.171	0.877	22.069	0.45	7.5
94	69.9	0.046	753.95	22.276	17.011	620.49	103.93	105.14	5.176	0.875	22.831	0.45	7.4
95	69.85	0.046	756.94	22.546	17.017	620.49	103.99	105.19	5.182	0.873	23.778	0.449	7.9
96	69.8	0.052	727.17	22.263	17.022	620.49	104.04	105.24	5.187	0.871	23.053	0.449	8.65
97	70.05	0.052	702.15	21.792	17.028	620.49	104.09	105.29	5.193	0.87	22.354	0.449	9.05
98	70.3	0.05	736.84	21.897	17.033	620.49	104.15	105.34	5.199	0.868	22.571	0.449	8.75
99	70.35	0.047	753.67	21.987	17.039	620.49	104.2	105.4	5.204	0.866	22.645	0.449	8.3
100	70.4	0.062	707.28	22.168	17.045	620.49	104.26	105.45	5.21	0.864	22.181	0.448	8.3
101	70.7	0.044	746.16	23.474	17.05	620.49	104.31	105.5	5.216	0.863	23.273	0.448	8.3
102	71	0.051	752.26	23.153	17.056	620.49	104.36	105.55	5.221	0.861	23.419	0.448	8.5
103	70.55	0.042	767.26	23.388	17.061	620.5	104.42	105.6	5.227	0.859	22.683	0.448	8.65

Data	PPM	FAF	TT	FR	FA	IFS	ET	IA	IFD	AAF	DDA	DPD	DPM
104	70.1	0.036	767.86	22.774	17.067	620.5	104.47	105.65	5.233	0.857	22.712	0.448	8.85
105	70.3	0.032	769.85	22.435	17.072	620.5	104.53	105.71	5.238	0.856	22.425	0.448	8.85
106	70.5	0.043	754.43	22.848	17.078	620.5	104.58	105.76	5.244	0.854	22.914	0.447	8.9
107	70.9	0.034	780.73	22.745	17.083	620.5	104.63	105.81	5.25	0.852	24.568	0.447	9.25
108	71.3	0.032	783.94	23.17	17.089	620.5	104.69	105.86	5.255	0.85	23.958	0.447	8.35
109	71	0.031	788.68	22.849	17.094	620.5	103.79	105.91	5.261	0.848	22.535	0.447	7.6
110	70.7	0.037	746.38	23.031	17.1	620.5	103.65	105.96	5.267	0.847	26.786	0.447	8.95
111	70.7	0.041	755.81	22.949	17.106	620.5	103.57	106.02	5.272	0.845	27.536	0.447	9.5
112	70.7	0.043	748.95	23.113	17.111	620.5	103.5	106.07	5.278	0.843	24.343	0.446	8.75
113	70.95	0.043	758.68	22.582	17.117	620.5	103.42	106.12	5.284	0.841	23.256	0.446	8.85
114	71.2	0.046	759.91	23.115	17.122	620.5	103.35	106.17	5.289	0.84	22.137	0.446	9.6
115	70.65	0.035	785.27	22.937	17.128	620.51	103.27	106.22	5.295	0.838	22.402	0.446	9.4
116	70.1	0.039	781.52	23.016	17.133	620.51	103.2	106.27	5.301	0.836	24.207	0.446	8.85
117	70.2	0.036	775.23	21.74	17.139	620.51	103.12	106.33	5.306	0.834	24.884	0.445	10.2
118	70.3	0.046	746.44	21.967	17.144	620.51	103.05	106.38	5.312	0.832	23.111	0.445	9.75
119	69.95	0.044	750	22.325	17.15	620.51	102.97	106.43	5.318	0.831	21.861	0.445	8.8
120	69.6	0.042	755.57	22.417	17.155	620.51	102.9	106.48	5.323	0.829	23.48	0.445	9
121	70	0.044	757.89	21.668	17.161	620.51	102.83	106.53	5.329	0.827	25.019	0.445	8.25
122	70.4	0.045	734.02	22.2	17.167	620.51	102.75	106.58	5.335	0.825	23.952	0.445	8.25
123	69.9	0.036	759.25	22.452	17.172	620.51	102.81	106.64	5.34	0.824	23.102	0.444	8.15
124	69.4	0.045	728.3	22.558	17.178	620.51	102.87	106.69	5.346	0.822	21.814	0.444	8.2
125	69.45	0.054	733.42	22.705	17.183	620.51	102.93	106.74	5.351	0.82	21.778	0.444	8.55
126	69.5	0.04	750.82	22.64	17.189	620.51	102.99	106.79	5.357	0.818	22.936	0.444	8.2
127	69.25	0.044	731.88	22.365	17.194	620.52	103.06	106.84	5.363	0.817	24.039	0.444	8.25
128	69	0.047	732.04	22.633	17.2	620.52	103.12	106.89	5.368	0.815	23.921	0.443	8.1
129	69.25	0.046	738.12	22.957	17.205	620.52	103.18	106.95	5.374	0.813	22.988	0.443	8.15
130	69.5	0.043	739.26	22.517	17.211	620.52	103.24	107	5.38	0.811	22.077	0.443	9.35
131	69.95	0.044	755.81	23.285	17.217	620.52	103.3	107.05	5.385	0.809	21.423	0.443	10.55
132	70.4	0.049	731.26	22.343	17.222	620.52	103.36	107.1	5.391	0.808	21.527	0.443	9
133	70.2	0.043	761.93	22.496	17.228	620.52	103.42	107.15	5.397	0.806	21.495	0.443	6.9
134	70	0.047	751.76	22.012	17.233	620.52	103.48	107.2	5.402	0.804	23.822	0.442	7.8
135	70	0.035	764.57	21.569	17.239	620.52	103.54	107.26	5.408	0.802	24.461	0.442	9.35
136	70	0.051	731.74	22.371	17.244	620.52	103.6	107.31	5.414	0.801	23.132	0.442	9.85
137	70	0.049	740.85	22.702	17.25	620.52	103.66	107.36	5.419	0.799	21.456	0.442	8.7
138	70	0.052	741.84	22.688	17.255	620.52	103.72	107.41	5.425	0.797	20.278	0.442	8.55

Data	PPM	FAF	TT	FR	FA	IFS	ET	IA	IFD	AAF	DDA	DPD	DPM
139	69.8	0.052	739.26	22.212	17.261	620.53	102.91	107.46	5.431	0.795	21.072	0.442	8.65
140	69.6	0.013	843.68	21.961	17.266	620.53	103.44	107.51	5.436	0.794	23.712	0.441	7.9
141	69.7	0.045	730.71	22.986	17.272	620.53	102.5	107.56	5.442	0.792	24.511	0.441	8.25
142	69.8	0.046	746.16	22.101	17.278	630.4	102.06	107.62	5.448	0.79	24.007	0.441	8.6
143	69.35	0.057	722.9	22.953	17.283	630.4	101.87	107.67	5.453	0.788	26.342	0.441	9.1
144	68.9	0.041	750.53	22.914	17.289	630.4	101.95	107.72	5.459	0.786	26.221	0.441	9.6
145	69.5	0.051	741.84	22.618	17.294	630.4	102.03	107.77	5.465	0.785	23.605	0.44	9.2
146	70.1	0.042	746.69	22.691	17.3	630.41	102.11	107.82	5.47	0.783	22.755	0.44	9.05
147	69.7	0.055	726.37	21.84	17.305	630.41	102.18	107.79	5.476	0.781	23.867	0.44	9.3
148	69.3	0.046	736.94	22.657	17.311	630.41	102.26	107.7	5.482	0.779	24.634	0.44	9
149	69.4	0.045	724.34	22.545	17.316	630.41	102.34	107.61	5.487	0.778	25.106	0.44	9
150	69.5	0.057	698.35	22.141	17.322	630.42	102.42	107.51	5.493	0.776	25.031	0.44	9.4
151	70	0.045	727.07	22.568	17.327	630.42	102.5	107.42	5.498	0.774	23.834	0.439	10.1
152	70.5	0.03	765.9	23.042	17.333	630.42	102.58	107.33	5.504	0.772	23.993	0.439	10.25
153	70.2	0.051	754.82	22.297	17.339	630.42	102.65	107.24	5.51	0.771	23.579	0.439	9.1
154	69.9	0.049	730.09	22.085	17.344	630.42	102.73	107.15	5.515	0.769	22.269	0.439	8.25
155	69.45	0.05	723.04	22.769	17.35	630.43	102.81	107.05	5.521	0.767	22.796	0.439	9.05
156	69	0.041	740.64	22.343	17.355	630.43	102.75	106.96	5.527	0.765	23.683	0.438	8.6
157	69.5	0.037	769.7	21.444	17.361	630.43	102.66	106.87	5.532	0.763	23.691	0.438	8.85
158	70	0.061	698.05	22.755	17.366	630.43	102.56	106.78	5.538	0.762	23.831	0.438	9.7
159	69.75	0.052	734.25	22.419	17.372	630.43	102.46	106.69	5.544	0.76	24.637	0.438	9.85
160	69.5	0.046	739.64	22.477	17.377	630.44	102.36	106.59	5.549	0.758	23.99	0.438	8.75
161	69.5	0.044	733.2	21.648	17.383	630.44	102.26	106.5	5.555	0.756	24.081	0.438	7.85
162	69.5	0.047	737.24	22.349	17.389	630.44	102.17	106.41	5.561	0.755	24.538	0.437	8.9
163	69.55	0.052	740.64	22.398	17.394	630.44	102.07	106.32	5.566	0.753	23.109	0.437	8.4
164	69.6	0.041	768.22	22.731	17.4	630.44	101.97	106.23	5.572	0.751	23.878	0.437	8.15
165	69.8	0.049	723.28	21.762	17.405	630.45	101.87	106.13	5.578	0.749	24.182	0.437	8.35
166	70	0.059	705.42	22.072	17.411	630.45	101.69	106.04	5.583	0.747	23.421	0.437	8.45
167	70.05	0.052	729.01	22.006	17.416	630.45	101.5	105.95	5.589	0.746	23.002	0.437	8.75
168	70.1	0.073	687.04	21.917	17.422	630.45	101.31	105.86	5.595	0.744	24.754	0.436	9.5
169	70.2	0.07	697.01	22.106	17.427	630.45	101.12	105.77	5.6	0.742	24.218	0.436	9
170	70.3	0.063	700.09	21.763	17.433	630.46	100.92	105.68	5.606	0.74	22.003	0.436	8.05
171	70.05	0.062	683.93	22.433	17.438	630.46	100.98	105.58	5.612	0.739	22.453	0.436	8.6
172	69.8	0.061	698.24	21.564	17.444	630.46	101.06	105.49	5.617	0.737	23.252	0.436	8.9
173	69.5	0.056	731.52	21.691	17.45	630.46	101.15	105.4	5.623	0.735	25.129	0.435	8.85

Data	PPM	FAF	TT	FR	FA	IFS	ET	IA	IFD	AAF	DDA	DPD	DPM
174	69.2	0.061	714	21.912	17.455	630.47	101.23	105.31	5.629	0.733	25.162	0.435	9.2
175	69.35	0.057	721.18	21.773	17.461	630.47	101.32	105.22	5.634	0.732	24	0.435	9.55
176	69.5	0.054	730.17	22.036	17.466	630.47	101.4	105.12	5.64	0.73	23.07	0.435	9.7
177	69.4	0.069	703.12	21.692	17.472	630.47	101.48	105.03	5.646	0.728	22.916	0.435	10.2
178	69.3	0.055	722.74	21.707	17.477	630.47	101.57	104.94	5.651	0.726	24.226	0.435	10.75
179	69.8	0.051	748.51	22.623	17.483	630.48	101.65	104.85	5.657	0.724	23.512	0.434	11.65
180	70.3	0.05	752.89	22.049	17.488	630.48	101.74	104.76	5.662	0.723	22.067	0.434	12.05
181	70.3	0.033	775.73	22.673	17.494	630.48	101.82	104.66	5.668	0.721	22.681	0.434	12.2
182	70.3	0.054	750.5	22.352	17.499	630.48	101.99	104.57	5.674	0.719	23.374	0.434	12.85
183	70.95	0.057	723.12	22.033	17.505	630.48	102.47	104.48	5.679	0.717	23.598	0.434	12.05
184	71.6	0.035	776.04	22.265	17.511	630.49	102.79	104.39	5.685	0.716	25.044	0.433	9.6
185	71.15	0.053	741.11	22.928	17.516	630.49	102.53	104.3	5.691	0.714	24.504	0.433	8
186	70.7	0.05	759.8	22.482	17.522	630.49	102.28	105.56	5.696	0.712	24.022	0.433	8.05
187	70.55	0.03	795.14	22.013	17.527	630.49	102.02	107.45	5.702	0.71	23.922	0.433	9.25
188	70.4	0.055	729.2	22.144	17.533	640.25	101.75	108.74	5.708	0.709	22.406	0.433	9.2
189	70.15	0.044	784.16	22.184	17.538	640.28	101.45	108.56	5.713	0.707	22.395	0.433	8.45
190	69.9	0.053	747.76	22.061	17.544	635.39	101.15	108.38	5.719	0.705	22.464	0.432	8.7
191	69.75	0.047	751.65	22.118	17.549	630.47	100.9	108.2	5.725	0.703	22.296	0.432	8.55
192	69.6	0.051	739.67	21.481	17.555	630.45	100.94	108.02	5.73	0.701	22.655	0.432	8.85
193	70.05	0.051	743.18	21.26	17.561	630.43	100.98	107.84	5.736	0.7	24.275	0.432	9.6
194	70.5	0.061	708.56	22.011	17.566	630.41	101.02	107.66	5.742	0.698	24.735	0.432	9.7
195	70.6	0.057	733.39	21.755	17.572	630.39	101.05	107.48	5.747	0.696	24.058	0.432	9.45
196	70.7	0.059	738.44	22.15	17.577	630.37	101.09	107.31	5.753	0.694	24.858	0.431	9.75
197	70.25	0.045	750.01	22.811	17.583	630.35	101.13	107.13	5.759	0.693	26.626	0.431	9.7
198	69.8	0.046	729.99	22.635	17.588	630.33	101.16	106.95	5.764	0.691	26.205	0.431	9.25
199	69.75	0.053	739.66	23.074	17.594	630.31	101.2	106.94	5.77	0.689	23.476	0.431	8.75
200	69.7	0.065	715.13	22.839	17.599	630.29	101.24	106.94	5.776	0.687	22.026	0.431	8.9
201	69.65	0.044	719.23	22.724	17.605	630.27	101.28	106.93	5.781	0.686	23.654	0.43	9.25
202	69.6	0.05	731.15	22.547	17.61	630.25	101.31	106.93	5.787	0.684	26.179	0.43	8.9
203	69.6	0.043	769.61	23.158	17.616	630.23	101.35	106.92	5.793	0.682	25.83	0.43	9.95
204	69.6	0.037	765.73	22.559	17.622	630.21	101.39	106.92	5.798	0.68	24.206	0.43	10.7
205	69.95	0.044	741.81	22.294	17.627	630.19	101.42	106.92	5.804	0.678	24.177	0.43	9.95
206	70.3	0.041	738.45	23.487	17.633	630.17	101.46	106.91	5.81	0.677	24.066	0.43	8.55
207	70.3	0.039	760.92	22.388	17.638	630.15	101.5	106.91	5.815	0.675	24.93	0.429	8.65
208	70.3	0.051	750.41	22.944	17.644	630.13	101.54	106.91	5.821	0.673	24.212	0.429	9.3

Data	PPM	FAF	TT	FR	FA	IFS	ET	IA	IFD	AAF	DDA	DPD	DPM
209	69.9	0.042	761.36	22.823	17.649	630.11	101.57	106.9	5.826	0.671	23.703	0.429	9.1
210	69.5	0.04	767.11	23.393	17.655	630.09	101.61	106.9	5.832	0.67	24.392	0.429	9.15
211	69.75	0.049	745.6	22.714	17.66	630.07	101.65	106.9	5.838	0.668	22.871	0.429	8.85
212	70	0.058	736.4	23.193	17.666	630.05	101.68	106.89	5.843	0.666	23.697	0.428	9.25
213	69.9	0.039	771.85	22.563	17.671	630.03	101.72	106.89	5.849	0.664	25.617	0.428	9.4
214	69.8	0.031	763.27	22.054	17.677	630.02	101.76	106.24	5.855	0.663	26.437	0.428	9.45
215	70	0.043	736.31	21.381	17.683	630	101.8	106.01	5.86	0.661	25.179	0.428	9.75
216	70.2	0.045	747.09	22.063	17.688	629.98	101.83	105.78	5.866	0.659	23.397	0.428	10.65
217	70.3	0.049	738.43	22.53	17.694	629.96	101.87	105.56	5.872	0.657	24.174	0.428	10.1
218	70.4	0.024	793.34	22.053	17.699	629.94	101.63	105.33	5.877	0.655	23.091	0.427	9.2
219	70.05	0.041	768.42	22.315	17.705	629.92	101.34	105.1	5.883	0.654	21.886	0.427	9.15
220	69.7	0.033	755.53	22.551	17.71	629.9	101.05	104.87	5.889	0.652	24.003	0.427	8.75
221	69.65	0.046	746.1	22.477	17.716	629.88	100.83	104.64	5.894	0.65	24.258	0.427	9.45
222	69.6	0.033	760.75	23.265	17.721	629.86	100.69	104.41	5.9	0.648	24.095	0.427	10.55
223	69.9	0.028	791.82	22.373	17.727	629.84	100.55	104.18	5.906	0.647	24.778	0.427	9.15
224	70.2	0.044	742.27	22.511	17.733	629.82	100.41	103.95	5.911	0.645	23.308	0.426	8.55
225	69.9	0.044	739	22.431	17.738	629.8	100.27	104.22	5.917	0.643	23.161	0.426	9.4
226	69.6	0.044	749.07	22.653	17.744	629.78	100.13	106.23	5.923	0.641	24.139	0.426	8.8
227	69.95	0.045	738.74	22.508	17.749	629.76	99.991	108.25	5.928	0.639	25.28	0.426	8.6
228	70.3	0.046	733.64	22.798	17.755	620.52	111.79	105.74	5.019	0.638	23.698	0.426	13.6

APPENDIX B. SIMULATION PROGRAM (MATLAB PROGRAM)

B.1. Main Simulation

```
clear all ,close all ,clc

% Parameter Estimated from Bilen e. a. 2010 %%
load standardized_data

X_matrix=[data(:,1:12)];
Y_data=data(:,14); var(Y_data);

% Define number of Xs
p=12;

% In control parameter
mean_X_0=mean(X_matrix);
mean_X_1=mean(X_matrix);

phi1_0 = 1.1188;
phi2_0 = -0.5804;
phi3_0 = 0.3945;
var_Z_0 = 0.6933^2;

phi1_1 = 1.1188;
phi2_1 = -0.5804;
phi3_1 = 0.3945;
var_Z_1 = 0.6933^2;

X_4_MEWMA=[X_matrix(:,6:7) X_matrix(:,9:10)];
Cov_X_4_MEWMA=cov(X_4_MEWMA);

R=chol(cov(X_matrix));
[Eigen_Vector_Matrix,latent,explained] = pcacov(cov(X_matrix));
Eigen_Value_Matrix=diag(latent(1:4));
pc=4;

P_Matrix=Eigen_Vector_Matrix(:,1:pc);

beta_0= [9.6094; -0.3165; -0.1608 ; 0.8057; 0.4770];
beta_1= beta_0;

H_Y=2.98; lambda_Y=0.2;
H_X=16.15; lambda_X=0.2;

simulation=100000; %subgroup_size=1;

% Define the number of simulation
Time_Series_NOISE=generate_AR3(50000,phi1_1,phi2_1,phi3_1,var_Z_1);

% Generate the time series noise

%Mean(Time_Series_NOISE),var(Time_Series_NOISE),plot(1:10000,Time_Series_
NOISE(1:10000))

% In Control Stage
mean_X_1 = mean(X_matrix) ;
beta_1= beta_0;
```

```

R_out_of_control=R;
P_Matrix_out_of_control=P_Matrix;

counter_X_Y=ARL_EWMA_Y_X(simulation,p, pc, R, R_out_of_control,
P_Matrix, P_Matrix_out_of_control, Eigen_Value_Matrix, Cov_X_4_MEWMA ,
mean_X_0, mean_X_1, beta_0, beta_1, H_Y, lambda_Y, H_X, lambda_X,
Time_Series_NOISE,phi1_0,phi2_0,phi3_0,var_Z_0,phi1_1,phi2_1,phi3_1,
var_Z_1);

data_case_0=cell(1,3);
data_case_0{1,1}=counter_X_Y ;
data_case_0{1,2}=mean(counter_X_Y);
data_case_0{1,3}=std(counter_X_Y);

save data_case_0 data_case_0
load data_case_0
disp(' The following is the mean of ARL of X, Y, minMRL, prop.of
detected by MEWA of X')
disp(data_case_0{1,2})
disp(' The following is the STD of ARL of X, Y, minMRL, prop.of
detected by MEWA of X')
disp(data_case_0{1,3})

% Out of Control Stage
% Changes in Mean of X only
% Case 1 mean_X6 7 9 10 is out of control
clear i counter_X_Y Eigen_Vector_Matrix1 latent1 explained1
delta=[ 0.25 0.5 1 1.5 2 3];
beta_1= beta_0;
data_case_1=cell(length(delta),3);
for i=1:length(delta)
mean_X_1 = mean(X_matrix) + delta(i).*[0 0 0 0 0 1 1 0 1 1 0 0];
R_out_of_control=R;
P_Matrix_out_of_control=P_Matrix;

counter_X_Y=ARL_EWMA_Y_X(simulation,p, pc, R, R_out_of_control,
P_Matrix, P_Matrix_out_of_control, Eigen_Value_Matrix, Cov_X_4_MEWMA ,
mean_X_0, mean_X_1, beta_0, beta_1, H_Y, lambda_Y, H_X, lambda_X,
Time_Series_NOISE,phi1_0,phi2_0,phi3_0,var_Z_0,phi1_1,phi2_1,phi3_1,
var_Z_1);

data_case_1{i,1}=counter_X_Y ;
data_case_1{i,2}=mean(counter_X_Y);
data_case_1{i,3}=std(counter_X_Y);
end

save data_case_1 data_case_1
load data_case_1
disp(' The following is the mean of ARL of X, Y, minMRL, prop.of
detected by MEWA of X')
data_case_1{:,2}
disp(' The following is the STD of ARL of X, Y, minMRL, prop.of detected
by MEWA of X')
data_case_1{:,3}
% Case 2 mean_X6 7 9 10 is in control but the others are out of control
clear i counter_X_Y Eigen_Vector_Matrix1 latent1 explained1
delta=[ 0.25 0.5 1 1.5 2 3];
beta_1= beta_0;
data_case_2=cell(length(delta),3);

```

```

for i=1:length(delta)
    mean_X_1 = mean(X_matrix) + delta(i).*[1 1 1 1 1 0 0 1 0 0 1 1];
    R_out_of_control=R;
    P_Matrix_out_of_control=P_Matrix;
    counter_X_Y=ARL_EWMA_Y_X(simulation,p, pc, R, R_out_of_control,
P_Matrix, P_Matrix_out_of_control, Eigen_Value_Matrix, Cov_X_4_MEWMA ,
mean_X_0, mean_X_1, beta_0, beta_1, H_Y, lambda_Y, H_X, lambda_X,
Time_Series_NOISE,phi1_0,phi2_0,phi3_0,var_Z_0,phi1_1,phi2_1,phi3_1,
var_Z_1);

    data_case_2{i,1}=counter_X_Y ;
    data_case_2{i,2}=mean(counter_X_Y);
    data_case_2{i,3}=std(counter_X_Y);
end

save data_case_2 data_case_2
load data_case_2
disp(' The following is the mean of ARL of X, Y, minMRL, prop.of
detected by MEWA of X')
data_case_2{:,2},
disp(' The following is the STD of ARL of X, Y, minMRL, prop.of detected
by MEWA of X')
data_case_2{:,3}

% Case 3 All mean_X are out of control
clear i counter_X_Y Eigen_Vector_Matrix1 latent1 explained1
delta=[0.25 0.5 1 1.5 2 3];
beta_1= beta_0;
data_case_3=cell(length(delta),3);
for i=1:length(delta)
    mean_X_1 = mean(X_matrix) + delta(i).*[1 1 1 1 1 1 1 1 1 1 1 1];
    R_out_of_control=R;
    P_Matrix_out_of_control=P_Matrix;
    counter_X_Y=ARL_EWMA_Y_X(simulation,p, pc, R, R_out_of_control,
P_Matrix, P_Matrix_out_of_control, Eigen_Value_Matrix, Cov_X_4_MEWMA ,
mean_X_0, mean_X_1, beta_0, beta_1, H_Y, lambda_Y, H_X, lambda_X,
Time_Series_NOISE,phi1_0,phi2_0,phi3_0,var_Z_0,phi1_1,phi2_1,phi3_1,
var_Z_1);

    data_case_3{i,1}=counter_X_Y ;
    data_case_3{i,2}=mean(counter_X_Y);
    data_case_3{i,3}=std(counter_X_Y);
end

save data_case_3 data_case_3
load data_case_3
disp(' The following is the mean of ARL of X, Y, minMRL, prop.of
detected by MEWA of X')
data_case_3{:,2},
disp(' The following is the STD of ARL of X, Y, minMRL, prop.of detected
by MEWA of X')
data_case_3{:,3}

% Case 4 Cov_X (X6,X7,X9,X10)is out of control
clear i counter_X_Y Eigen_Vector_Matrix1 latent1 explained1
delta=[0.25 0.5 1 1.5 2 3];
beta_1= beta_0;
data_case_4=cell(length(delta),3);
for i=1:length(delta)
    mean_X_1 = mean_X_0 ;

```

```

R_out_of_control=chol(cov(X_matrix)+delta(i).*diag([0 0 0 0 0 1 1 0 1
1 0 0] ));

[Eigen_Vector_Matrix1,latent1,explained1] = pcacov(
cov(X_matrix)+delta(i).*diag([0 0 0 0 0 1 1 0 1 1 0 0] ));
Eigen_Value_Matrix1=diag(latent1(1:4));
pc=4;
P_Matrix_out_of_control=Eigen_Vector_Matrix1(:,1:pc);

counter_X_Y=ARL_EWMA_Y_X(simulation,p, pc, R,
R_out_of_control, P_Matrix, P_Matrix_out_of_control, Eigen_Value_Matrix,
Cov_X_4_MEWMA , mean_X_0, mean_X_1, beta_0, beta_1, H_Y, lambda_Y, H_X,
lambda_X,
Time_Series_NOISE,phi1_0,phi2_0,phi3_0,var_Z_0,phi1_1,phi2_1,phi3_1,
var_Z_1);

data_case_4{i,1}=counter_X_Y ;
data_case_4{i,2}=mean(counter_X_Y);
data_case_4{i,3}=std(counter_X_Y);
end

save data_case_4 data_case_4
load data_case_4
disp(' The following is the mean of ARL of X, Y, minMRL, prop.of
detected by MEWA of X')
data_case_4{:,2},
disp(' The following is the STD of ARL of X, Y, minMRL, prop.of detected
by MEWA of X')
data_case_4{:,3}

% Case 5 Regression parameter BETA (B1,B2, B3, B4) are changed
clear i counter_X_Y Eigen_Vector_Matrix1 latent1 explained1

beta_change=[0.5 0.75 0.9 1.1 1.25 1.5];
data_case_5=cell(length(beta_change),3);

for i=1:length(beta_change)
mean_X_1 = mean_X_0 ;
R_out_of_control=R;
P_Matrix_out_of_control=P_Matrix;
beta_1= [9.6094; -0.3165*beta_change(i); -0.1608*beta_change(i);
0.8057*beta_change(i); 0.4770*beta_change(i)];

counter_X_Y=ARL_EWMA_Y_X(simulation,p, pc, R, R_out_of_control,
P_Matrix, P_Matrix_out_of_control, Eigen_Value_Matrix, Cov_X_4_MEWMA ,
mean_X_0, mean_X_1, beta_0, beta_1, H_Y, lambda_Y, H_X, lambda_X,
Time_Series_NOISE,phi1_0,phi2_0,phi3_0,var_Z_0,phi1_1,phi2_1,phi3_1,
var_Z_1);

data_case_5{i,1}=counter_X_Y ;
data_case_5{i,2}=mean(counter_X_Y);
data_case_5{i,3}=std(counter_X_Y);
end

save data_case_5 data_case_5
load data_case_5

```

```

disp(' The following is the mean of ARL of X, Y, minMRL, prop.of
detected by MEWA of X')
data_case_5{:,2},
disp(' The following is the STD of ARL of X, Y, minMRL, prop.of detected
by MEWA of X')
data_case_5{:,3}

% Case 6 Cov_X (X6,X7,X9,X10) is out of control AND Regression parameter
BETA (B1,B2, B3, B4) are changed
clear i counter_X_Y Eigen_Vector_Matrix1 latent1 explained1

delta=[ 0.25 0.5 1 1.5 2 3];
beta_change=[0.5 0.75 0.9 1.1 1.25 1.5];
data_case_6=cell(length(delta),3);

for i=1:length(delta)
mean_X_1 = mean_X_0 ;
beta_1= [9.6094; -0.3165*beta_change(i); -0.1608*beta_change(i);
0.8057*beta_change(i); 0.4770*beta_change(i)];
R_out_of_control=chol(cov(X_matrix)+delta(i).*diag([0 0 0 0 0 1 1 0 1
1 0 0] ));

[Eigen_Vector_Matrix1,latent1,explained1] = pcacov(
cov(X_matrix)+delta(i).*diag([0 0 0 0 0 1 1 0 1 1 0 0] ));
Eigen_Value_Matrix1=diag(latent1(1:4));
pc=4; % the significant number of the PC is 4
P_Matrix_out_of_control=Eigen_Vector_Matrix1(:,1:pc);

counter_X_Y=ARL_EWMA_Y_X(simulation,p, pc, R, R_out_of_control,
P_Matrix, P_Matrix_out_of_control, Eigen_Value_Matrix, Cov_X_4_MEWMA ,
mean_X_0, mean_X_1, beta_0, beta_1, H_Y, lambda_Y, H_X, lambda_X,
Time_Series_NOISE,phi1_0,phi2_0,phi3_0,var_Z_0,phi1_1,phi2_1,phi3_1,
var_Z_1);

data_case_6{i,1}=counter_X_Y ;
data_case_6{i,2}=mean(counter_X_Y);
data_case_6{i,3}=std(counter_X_Y);
end

save data_case_6 data_case_6
load data_case_6
disp(' The following is the mean of ARL of X, Y, minMRL, prop.of
detected by MEWA of X')
data_case_6{:,2},
disp(' The following is the STD of ARL of X, Y, minMRL, prop.of detected
by MEWA of X')
data_case_6{:,3}

% Case 7 Cov_X and Mean(X6,X7,X9,X10) are out of control and Regression
parameter BETA (B1,B2, B3, B4) are changed
clear i counter_X_Y Eigen_Vector_Matrix1 latent1 explained1
delta=[0.25 0.5 1 1.5 2 3];
beta_change=[0.5 0.75 0.9 1.1 1.25 1.5]; ;
data_case_7=cell(length(delta),3);

for i=1:length(delta)
mean_X_1 = mean(X_matrix) + delta(i).*[0 0 0 0 0 1 1 0 1 1 0 0];
beta_1= [9.6094; -0.3165*beta_change(i); -0.1608*beta_change(i);
0.8057*beta_change(i); 0.4770*beta_change(i)];

```

```

R_out_of_control=chol(cov(X_matrix)+delta(i).*diag([0 0 0 0 0 1 1 0 1
1 0 0] ));

[Eigen_Vector_Matrix1,latent1,explained1] = pcacov(
cov(X_matrix)+delta(i).*diag([0 0 0 0 0 1 1 0 1 1 0 0] ));
Eigen_Value_Matrix1=diag(latent1(1:4));
pc=4;
P_Matrix_out_of_control=Eigen_Vector_Matrix1(:,1:pc);

counter_X_Y=ARL_EWMA_Y_X(simulation,p, pc, R, R_out_of_control,
P_Matrix, P_Matrix_out_of_control, Eigen_Value_Matrix, Cov_X_4_MEWMA ,
mean_X_0, mean_X_1, beta_0, beta_1, H_Y, lambda_Y, H_X, lambda_X,
Time_Series_NOISE,phi1_0,phi2_0,phi3_0,var_Z_0,phi1_1,phi2_1,phi3_1,
var_Z_1);

data_case_7{i,1}=counter_X_Y ;
data_case_7{i,2}=mean(counter_X_Y);
data_case_7{i,3}=std(counter_X_Y);
end

save data_case_7 data_case_7
load data_case_7
disp(' The following is the mean of ARL of X, Y, minMRL, prop.of
detected by MEWA of X')
data_case_7{:,2},
disp(' The following is the STD of ARL of X, Y, minMRL, prop.of detected
by MEWA of X')
data_case_7{:,3}

```

B.2. Average Run Length (ARL) of EWMA and MEWMA

```

function counter_X_Y=ARL_EWMA_Y_X(simulation,p, pc, R, R_out_of_control,
P_Matrix, P_Matrix_out_of_control, Eigen_Value_Matrix, Cov_X_4_MEWMA ,
mean_X_0, mean_X_1, beta_0, beta_1, H_Y, lambda_Y, H_X, lambda_X,
Time_Series_NOISE,phi1_0,phi2_0,phi3_0,var_Z_0,phi1_1,phi2_1,phi3_1,
var_Z_1);

counter_X_Y=[];
for sim=1:simulation
%% Note that the first 3 observations will not be tested, just to get the
warming of the residual process
clear X_Generated T_Generated meanY_Generated Y_Generated
Subset_Time_Series_NOISE a0 a1 a2 a3
s = 0;
% s=sample_counter -->that will be used to count the no. of sample
a0=find(abs(Time_Series_NOISE)<0.01);
a1=length(a0);
a2=ceil(rand*a1/2);
% define the random seed to pick the subset of the time series noise
a3=a0(a2);
Subset_Time_Series_NOISE=Time_Series_NOISE(a3:50000,:);
% random chosen part of the time series noise

counter_for_X=[]; counter_for_Y=[];
out_of_control_id_X=0; out_of_control_id_Y=0;

while out_of_control_id_X==0 | out_of_control_id_Y==0;

```



```

s = s + 1 ;

%% Each single sample will be generated from
%% X --> T score--> Mean(Y)--> Y

X_Generated(s,:) = ( R_out_of_control'*normrnd(0,1,p,1)
)' + mean_X_1;
% Generate X from out-of-control
T_Generated(s,:) =
X_Generated(s,:)*P_Matrix_out_of_control;
meanY_Generated(s,:) = [1 T_Generated(s,:)]*beta_1;
% But regress using the in-control beta
Y_Generated(s,:) = meanY_Generated(s,:)+
Subset_Time_Series_NOISE(s);
X_4_MEWMA = [ X_Generated(s,6:7)
X_Generated(s,9:10)];

%% MEWMA for X
%% 1.Compute the MEWMA for statistics of the X6 X7 X9 X10
% since we start from s=4,... not from 1
if s==1,
MEWMA_Statistic_X(s,:)=lambda_X*X_4_MEWMA'+(1-lambda_X)*zeros(4,1);
else
MEWMA_Statistic_X(s,:)=lambda_X*X_4_MEWMA'+(1-
lambda_X)*MEWMA_Statistic_X(s-1,:);
end

%% 2.compute the control chart statistics
%% This is to get the plotted statistics as if the process is in control
if s<=6, T2_Statistic_X(s)=0;
else T2_Statistic_X(s)=( 1/
(lambda_X/(2-lambda_X)) * ( 1 - ((1-lambda_X)^(2*(s-6))) )
)*MEWMA_Statistic_X(s,:)*inv(Cov_X_4_MEWMA )*MEWMA_Statistic_X(s,:);
end

%% 3. Define the control limit for MEWMA
if s>=7, MEWMA_X_CL=H_X;
else MEWMA_X_CL=1000;
end

%% 4. Check if the MEWMA is in control
if abs(T2_Statistic_X(s))>MEWMA_X_CL,
out_of_control_id_X = 1 ;
counter_for_X=[counter_for_X max(0,s-
6)];
%keep the ARL for MEWMA
end

%% EWMA for White Noise generated from the Residual

%% 1. Get the residual as if the process is in control
%if s<=3, residual(s)=0; wn(s)=0;

residual(s)=Y_Generated(s,:)- [ 1
X_Generated(s,:)*P_Matrix]*beta_0;
if s<=3, wn(s)=0;
else wn(s)=residual(s)-phi1_0*residual(s-1)-
phi2_0*residual(s-2)-phi3_0*residual(s-3);
end

```

```

%% Compute the EWMA
    if s<=1, EWMA_Y(s)=lambda_Y*wn(s)+(1-lambda_Y)*0;
    else    EWMA_Y(s)=lambda_Y*wn(s)+(1-
lambda_Y)*EWMA_Y(s-1);
    end

%% Define the control limit for EWMA
    if s>=4, EWMA_Y_CL(s)=H_Y* sqrt( var_Z_0 *
(lambda_Y/(2-lambda_Y))* ( 1 - ((1-lambda_Y)^(2*(s-3))) ) );

% we need at least 3 sample to get the estimates of Y1, Y2, Y3 and
% another 3 to get estimates of wn(1) wn(2) wn(3)
    else    EWMA_Y_CL(s)=1000;
    end

%%Check if the EWMA is in control
    if abs(EWMA_Y(s))>EWMA_Y_CL(s),
        out_of_control_id_Y=1 ;
        counter_for_Y=[counter_for_Y max(0,s-3)]; %
keep the ARL for MEWMA
    end

    if mod(s,100)==0, s; end

end

%of while out_of_control_id = 0;

    if mod(sim,1000)==0, sim, end
    ARL_X=min(counter_for_X);
    ARL_Y=min(counter_for_Y);
    counter_X_Y(sim,:) = [ARL_X ARL_Y min(ARL_X,ARL_Y)
min(ARL_X,ARL_Y)==ARL_X];
end

%for sim=1:simulation

```

B.3. Theoretical Covariance of Y

```

function
[Estimated_var_Y,Theory_COV_Y]=Compute_Theory_COV_Y_AR3(t,phi1,phi2,phi3,va
r_Z)

% Find Theoretical Covariance for GLS regression

rho=zeros(t,1);
for i=1:t
    if i==1,    rho(i) = (phi1 + phi2*phi3) / ( 1-phi2 - (phi1+phi3)*phi3)
;
    elseif i==2, rho(i) = (phi1+phi3)*rho(i-1) +phi2;
    elseif i==3, rho(i) = phi1*rho(i-1) + phi2*rho(i-2) +phi3;
    else    rho(i) = phi1*rho(i-1) + phi2*rho(i-2) +phi3*rho(i-3);
    end
end
var_Z
Estimated_var_Y = var_Z/(1-phi1*rho(1)-phi2*rho(2)-phi3*rho(3))

```

```

Theory_COV_Y=eye(t);
  for i=1:t,
    for j=1:t
      if i==j, Theory_COV_Y(i,j)=Estimated_var_Y;
      else    Theory_COV_Y(i,j)=Estimated_var_Y*rho(abs(j-i)); end
    end
  end

end

%rho(1:3)

```

B.4. Generate AR(3)

```

function y=generate_AR3(t,phi1,phi2,phi3,var_Z)

%head start of 100
m=t+100;
y=zeros(m,1);
for i=1:m
  if i==1,    y(i)= normrnd(0,sqrt(var_Z));
  elseif i==2 y(i)= normrnd(0,sqrt(var_Z));
  elseif i==3 y(i)= normrnd(0,sqrt(var_Z));
  else      y(i)=phi1*y(i-1)+phi2*y(i-2)+phi3*y(i-3) +
normrnd(0,sqrt(var_Z));
  end
end

y=y(m-t+1:m);

```

APPENDIX C. MULTIPLE REGRESSION CONTROL CHART INTEGRATED WITH PRINCIPAL COMPONENT ANALYSIS³

C.1. Introduction

Most industrial processes involve many variables that significantly affect quality of the output characteristics. Moreover, when predictor variables are related to each other, multicollinearity among them is said to exist. In the multicollinearity case, direct analysis resulting from procedures such as linear regression is not reliable. In this paper, Principle Components Analysis (PCA) is used to alleviate the problem of multicollinearity and control chart is built through multiple regression modeling.

The earliest paper on regression control chart in the quality control literature is by DiPaola [1]. Mandel [2] uses the regression control chart in an example of post office department. Jackson [3] illustrates the use of principal components analysis in detail. Numerous considerations and rules on principal components are discussed by Jolliffe [4]. Mastrangelo *et al.* [6] review methodologies for multivariate process monitoring and show the effectiveness of principal components. Hawkins [5] discusses the case when there is a distributional change in each variable by using different regression adjusted scales. Since the regression model is rarely known in practice, Shu *et al.* [7] study the run length performance of regression control charts with estimated parameters for Shewhart and exponentially weighted moving average (EWMA) control charts.

C.2. Methodology

The proposed methodology for multivariate autocorrelated process control uses multiple regression control chart in conjunction with principal component analysis. First,

Content in this appendix was previously published in Proceeding of the Industrial Engineering Research Conference, May 19-23, 2007, Nashville, Tennessee and was written by Canan Bilen-Green, Xianzhe Chen, Anakaorn Khan and Om Prakash Yadav. Anakaorn Khan was responsible in writing the draft of the paper and analyzing data.

Principal Component Analysis (PCA) is used to identify significant independent process variables. Then, white noise series generated by autoregressive error correction model is used to construct a control chart for future process monitoring.

C.2.1. Multiple regression model

Suppose a process of interest involves k independent variables, x_1, x_2, \dots, x_k , then the multiple regression model is given as follows:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki} + \varepsilon_i, \quad (i = 1, 2, \dots, n) \quad (1)$$

where y is the response variable $\varepsilon_i \sim N(0, \sigma^2)$. One key assumption of multiple regression model is that any one value of y is statistically independent of any other value of y . Therefore, checking the validity of the independence assumption is an essential step before further analysis of the data. The multiple regression model in (1) can be written in matrix notation as

$$Y = XB + E \quad (2)$$

where $Y = [y_1 \ y_2 \ \dots \ y_n]'$, $E = [\varepsilon_1 \ \varepsilon_2 \ \dots \ \varepsilon_n]'$, $B = [\beta_0 \ \beta_1 \ \beta_2 \ \dots \ \beta_k]'$ and

$$X = \begin{bmatrix} 1 & x_{11} & \dots & x_{k1} \\ 1 & x_{12} & \dots & x_{k2} \\ \vdots & \vdots & & \vdots \\ 1 & x_{1n} & \dots & x_{kn} \end{bmatrix}_{n \times (k+1)}$$

Regression coefficient estimates are given by $\hat{\beta} = (X'X)^{-1} X'Y$ and the variance-covariance matrix of the estimated regression coefficients are represented by $Var(\hat{\beta}) = \sigma^2 (X'X)^{-1}$.

C.2.2. Standardizing

Since the predictors in the process often have different units, we should standardize them before further analysis. Consider equation (1), the independent variables are transformed

as $x_{ji}^s = (x_{ji} - \bar{x}_j) / s_j$, where x_{ji} is the i^{th} observation for the j^{th} variable (x_j) and $s_j = \sqrt{\sum_{i=1}^n (x_{ji} - \bar{x}_j)^2 / (n-1)}$.

Hence, equation (1) can be rewritten as

$$y_i = \beta_0^s + \beta_1^s \left[\frac{x_{1i} - \bar{x}_1}{s_1} \right] + \beta_2^s \left[\frac{x_{2i} - \bar{x}_2}{s_2} \right] + \dots + \beta_k^s \left[\frac{x_{ki} - \bar{x}_k}{s_k} \right] + \varepsilon_i \quad (3)$$

If we separate the first column from matrix X , then we will have

$$Y = \beta_0^s I + X^s B^s + E \quad (4)$$

where I is a vector of ones. It is noted that $X^s X^s$ is proportional to the correlation matrix for the predictor variables.

C.2.3. Principal components analysis

Principal Components Analysis (PCA) is a method for overcoming multicollinearity, which occurs when predictor variables are related to each other. Under multicollinearity, the analytic results from many procedures such as linear regression are not useful as they can be misleading. While PCA can be used to transform correlated variables into uncorrelated variables by using certain selection rules, in the meantime the dimension of a dataset can be reduced while explaining most of the variability in a process. Let A be $k \times k$ matrix which consists of the normalized eigenvector such that $AA^s = I$. Therefore, equation (4) can be written as

$$Y = \beta_0^s I + X^s AA^s B^s + E = \beta_0^s I + Z\gamma + E \quad (5)$$

where $Z = X^s A$ is a $n \times k$ matrix of principal components and $\gamma = A^s B^s$ is a $k \times 1$ vector of new coefficients. The principal components are orthogonal to each other since

$$Z^s Z = (X^s A)^s (X^s A) = A^s X^s X^s A = \text{diag}(\lambda_1 \quad \lambda_2 \quad \dots \quad \lambda_k) \quad (6)$$

where $\lambda_1, \lambda_2, \dots, \lambda_k$ are the eigenvalues of the matrix $X^s X^s$, furthermore $\sum_{j=1}^k \lambda_j = k$. Then, j^{th} principal component can be computed as

$$z_j = a_{1j}x_1^s + a_{2j}x_2^s + \dots + a_{kj}x_j^s \quad (7)$$

where $[a_{1j}, a_{2j}, \dots, a_{kj}]$ is eigenvector associated with eigenvalues λ_j and z_j accounts for λ_j of the total variance. The least squares estimator for regression coefficients of vector λ is given by $\hat{\gamma} = (Z^s Z)^{-1} Z^s Y$ and the variance covariance for the estimated coefficients of $\hat{\gamma}$ can be computed as

$$\text{Var}(\hat{\gamma}) = \hat{\sigma}^2 (Z^s Z)^{-1} = \hat{\sigma}^2 \text{diag}(\lambda_1^{-1}, \lambda_2^{-1}, \dots, \lambda_k^{-1}) \quad (8)$$

A comprehensive guideline for multivariate process analysis, monitoring, and control using the developed methodology is illustrated in Figure C.2.

C.3. Application

The sugar beet pulp drying process is shown in Figure C.1. The historical data for a period of two months for 24 hours a day are used in the analysis. Twelve input variables are adjusted according to time at t , $t+1/2$ and $t+1$.

The basic steps of the process data analysis are as follows: 1) use PCA to get independent predictor variables; 2) construct a regression model to get a set of white noise residuals; 3) build control charts on the white noise residuals for monitoring future process.

Step 1: The observed pulp process data are somewhat messy with a few missing points.

Therefore, in order to construct the stationary control chart, 228 consistent observations, showing no obvious outliers, collected in the middle of the process are selected from a total 1,349 observations. As the independent variables were in different units, these variables are standardized according to the procedure described above in the methodology section. Then SAS statistical analysis software is used to conduct PCA of the standardized data. The eigenvalues are given in Table C.1.

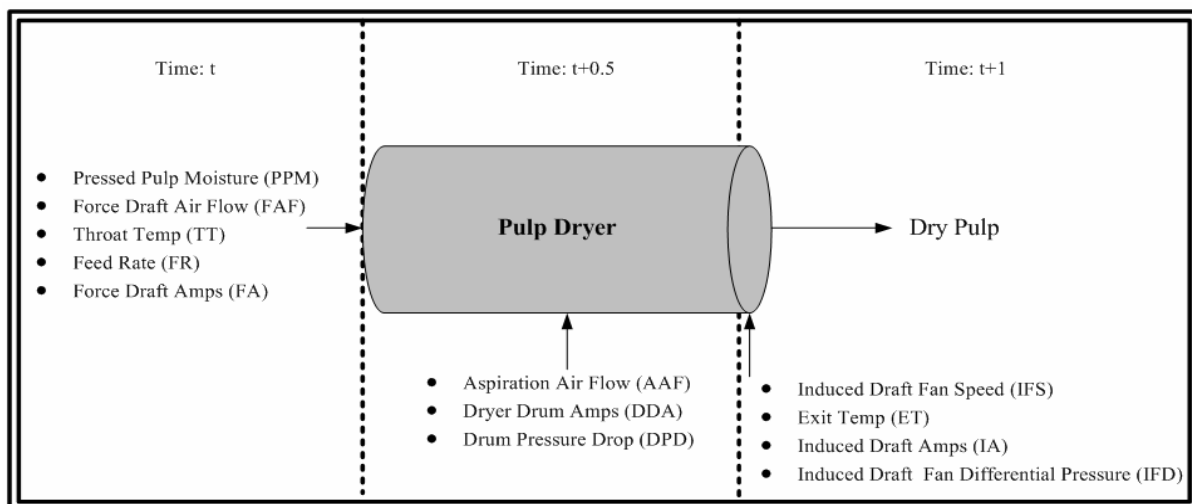


Figure C.1. Pulp drying process according to time

If a significant proportion of the total variance can be attributed to the first few variables, then these components can be used to explain the process without much loss of information. There are several rules for selecting principal components. According to [4], one popular rule is to choose the set of largest r components which first sums up to 85% or $\sum_{j=1}^r \lambda_j / k = 0.85$, another widely used rule states that the proportion of one component should be larger than the inverse of the number of variables, or $1/k = 1/12 = 0.083$. We selected the principal components according to both rules so that for the pulp drying process the first four largest principal components are selected for further analysis.

It is informative to determine the correlations of the variables with the principal components in order to interpret the importance of the k^{th} variable to the i^{th} principal component. The correlation coefficient between component z_i and variable x_k can be computed as $\rho_{z_i, x_k} = e_{ik} \sqrt{\lambda_i} / \sigma$. However, this correlation coefficient does not measure the importance of an individual x_k to a component y in the presence of other independent variables. It is recommended in the PCA literature that only the coefficients e_{ik} (in absolute value) should be used to interpret the components.

The first four eigenvectors associated with eigenvalues are presented in Table C.2. Notable observations are that variable z_{aaf} receives the greatest weight in principal component 1; z_{faf} , z_{tt} and z_{fr} are almost equally important to principal component 2; z_{ifd} and z_{dpd} contribute almost the same to principal component 3; the most important variable to principal component 4 is z_{dda} .

Step 2: Since the response variable (dried pulp moisture) is recorded every one hour and it shows that autocorrelation exists, it violates the independence assumption. It is

important to remove autocorrelation before the model is built. The method often used to deal with autocorrelation is to fit an ARIMA model for the input series so that it is sufficient to reduce the residuals to white noise. However, there are no ARIMA models that reasonably fit well for the response data since the residuals produced from the ARIMA model to build regression model will give an R-square value of almost zero. Therefore, we use the autoregressive error model.

One key assumption of ordinary regression analysis is that the errors are independent of each other. However, with time series data, the ordinary regression residuals are usually correlated over time. Hence, it is not desirable to use ordinary regression analysis for time series data. Therefore, we build a regression model with an autoregressive series for the random error, accounting for the autocorrelation of the errors. Instead of the usual regression model, the following autoregressive error model is used:

$$\begin{aligned}
 y_t &= XB + v_t \\
 v_t &= -\phi_1 v_{t-1} - \phi_2 v_{t-2} - \dots - \phi_m v_{t-m} + \varepsilon_t \\
 \varepsilon_t &\sim i.i.d.N(0, \sigma^2)
 \end{aligned}
 \tag{9}$$

This kind of regression analysis is called autoregressive error correction or serial correlation correction. Again with SAS, the following model coefficients are estimated by the maximum likelihood method:

$$\begin{aligned}
 y_t &= 9.6302 - 0.3199 * prin1 + 0.1652 * prin2 - 0.8376 * prin3 + 0.4932 * prin4 + v_t \\
 v_t &= 1.1188v_{t-1} - 0.5804v_{t-2} + 0.3945v_{t-3} + \varepsilon_t \\
 \varepsilon_t &\sim i.i.d.N(0, \sigma^2)
 \end{aligned}
 \tag{10}$$

The Durbin-Watson test for autocorrelation in residuals confirms that the residuals are white noise. And the intercept, components 1 to 4 and the coefficients for AR1, AR2 and AR3 are all significant indicating that the variables associated with these coefficients are important and meaningful to the model. The R-square is 0.6137, which is not as high as desired. However, various factors,

such as instrumentation, gauge accuracy, process interruption by operators, and data recording errors affect data collection and the production process. Therefore, even a model with a moderate R-square is informative about the process and the relative importance of numerous variables to the process.

Table C.1. Eigenvalues of the correlation matrix

PC	Eigenvalue	Proportion	Cumulative
1	4.6290	0.3858	0.3858
2	2.6333	0.2194	0.6052
3	1.6352	0.1363	0.7415
4	0.9960	0.0830	0.8245
5	0.6784	0.0565	0.8810
6	0.4533	0.0378	0.9188
7	0.3803	0.0317	0.9505
8	0.2762	0.0230	0.9735
9	0.1783	0.0149	0.9883
10	0.0694	0.0058	0.9941
11	0.0511	0.0043	0.9984
12	0.0195	0.0016	1.0000

Step 3: From the analysis results, the mean of the process is -.0016 and the standard deviation is 0.6933. So the control limits for Shewhart chart is computed as:

$$-0.0016 \pm 3 * 0.6933 = (-2.0815, 2.0783) \quad (11)$$

According to the analysis in step 1, when detecting an out of control signal, the sequence for checking process variables according to their priority should be: 1) AAF; 2) FAF, TT, and FR; 3) IFD and DPD; 4) DDA.

It is informative to interpret the results from the physical prospective of the process. For example, the aspiration air flow (AAF) is the air that causes different pressure between high temperature at the beginning and low temperature at the end of the drum. If there is high pressure in the drum, then the pulp will have less moisture content and vice versa. Thus, AAF contributes the majority of variation in the process. The second priority can be thought of as input factors affecting the variation of the process. Furthermore, the feed rate (FR) and FD air flow (FAF) affect the duration of the pulp inside the drum. If the feed rate and FD air flow are increased, the pulp will have less time inside the drum which directly leads to high moisture content in the pulp. Overall, these are the four important variables identified, out of the initial twelve, which should be included in the analysis.

Table C.2. Correlation between principal components and variables

Variable	Prin1	Prin2	Prin3	Prin4
z_ppm	-.333218	0.146091	0.194914	0.118675
z_faf	-.166576	-.539317	0.000217	-.024384
z_tt	0.202653	0.505909	0.056680	-.115743
z_fr	0.111124	0.512892	0.147625	-.057863
z_fa	0.293420	-.248280	0.414488	-.282509
z_ifs	0.380466	-.148433	0.119992	0.213348
z_et	-.341491	0.231609	0.072179	-.090648
z_ia	-.225141	-.119583	0.395806	0.254624
z_ifd	0.338596	0.000655	0.466796	-.275830
z_aaf	-.439290	0.088861	0.107443	-.094088
z_dda	0.067763	0.085509	0.353278	0.772074
z_dpd	0.312140	0.044331	-.484442	0.298605

C.4. Conclusion

This paper develops a framework for analyzing multivariate processes. Use of different rules for principal component determination will lead to different choices of components. It is important to construct control charts for stationary processes. Therefore, the process first should be checked to determine if it is stationary before further analysis.

C.5. References

1. DiPaola, P. P., 1945, "Use of Correlation in Quality Control," *Industrial Quality Control*, Vol. 2, No.1, 10-14.
2. Mandel, B. J., 1969, "The Regression Control Chart," *Journal of Quality Technology*, Vol. 1, No. 1, 1-9.
3. Jackson, J. E., 1980, "Principal Components and Factor Analysis: Part I – Principal Components," *Journal of Quality Technology*, Vol. 12, No. 4, 201-213.
4. Jolliffe, I. T., 1982, "A Note on the Use of Principal Components in Regression," *Appl. Statist.*, 31, No. 3, 300-303.
5. Hawkins, D. M., 1993, "Regression Adjustment for Variables in Multivariate Quality Control," *Journal of Quality Technology*, Vol. 25, No. 3, 170-182.
6. Mastrangelo, C. M., Runger, G. C. and Montgomery, D. C., 1996, "Statistical Process Monitoring with Principle Components," *Quality and Reliability Engineering International*, Vol. 12, 203-210.
7. Shu, L., Tsung, F., and Tsui, K., 2004, "Run-Length Performance of Regression Control Charts with Estimated Parameters," *Journal of Quality Technology*, Vol. 36, No. 3, 280-292.

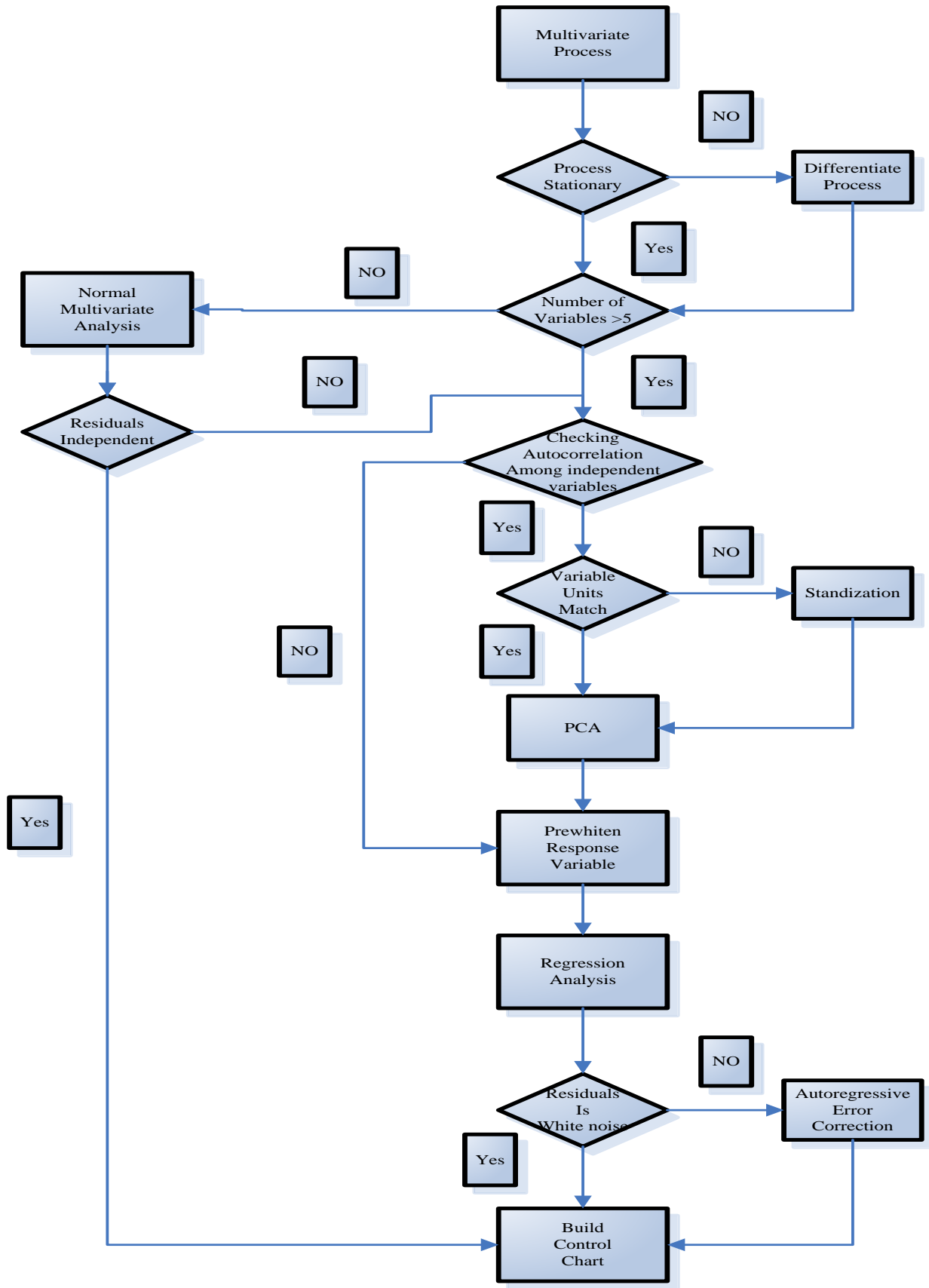


Figure C.2. Flow chart for analyzing multivariate processes

APPENDIX D. MULTIVARIATE AUTOCORRELATED PROCESS MONITORING WITH MULTIPLE REGRESSION AND PRINCIPAL COMPONENTS⁴

D.1. Introduction

In most chemical and process industries, there are often many variables involved in process monitoring, and these variables are frequently related to each other which creates a multicollinearity problem. Besides, the capability of increasing sampling frequency has been developed tremendously in recent decades which causes the sampling data from the process inherit autocorrelation effect. Therefore, the traditional control charts and direct method in establishing the relationship between inputs and outputs are no longer effective in process monitoring. In general, the quality of the process output is directly affected by many uncontrollable input variables. Mandel [9] attempted to monitor the process output and capture the relationship between inputs and outputs by combining linear regression analysis with control charts. This regression control chart is applied as performance measurement and control tool in detecting assignable causes for process improvement by using a regression line to establish the relationship between a single output and single input. However, in most cases, many process variables need to be monitored at the same time with a separate control chart for each variable. The difficulty in monitoring many process variables can be remedied with the regression adjustment control chart, especially for cascade process [2, 3]. The performance of the regression adjustment control chart has been explored in terms of detecting an out-of-control signal for various shift sizes of each variable. The difficulty arises when some input variables are related to each other. The multicollinearity is said to exist among them which fails the assumption of independent regressors in regression analysis. This impacts the response variables and estimation of the regression parameter. Montgomery et al. [11] solved this issue

⁴ Content in this appendix was previously published in Proceeding of the Industrial Engineering Research Conference, May 19- June 3, 2009, Miami, Florida and was written by Canan Bilen-Green, Anakaorn Khan and Om Prakash Yadav. Anakaorn Khan was responsible in writing the draft of the paper and analyzing data.

by a simulation study and provides some guidelines of unbiased estimation method for the regression parameter. Schall et al. [13] recommended using principal component analysis in handling the multicollinearity effect for the case where the number of input variables is less than the number of output variables and simultaneously reducing the number of both input and output variables in multivariate processes.

The traditional control chart performance is also affected by the autocorrelation effect. This issue has been studied extensively due to its effect in reducing the performance of traditional control charts. Some researchers have suggested extensions of traditional multivariate control chart for monitoring autocorrelated processes [5, 7]. Loredó et al. [8] developed the extension of reference [2, 3] by suggesting a model-based control chart for monitoring multivariate autocorrelated processes. The application of principal component analysis can also be used in monitoring multiple variables, when the process is autocorrelated. Runger [12] presented multivariate control statistic based on principal component decomposition. Mastrangelo et al. [10] explored the application of principal components in multivariate autocorrelated process by plotting principal components on multivariate control chart for process monitoring. The integration of principal component analysis (PCA) and multiple regression is seldom explored in establishing the relationship between input and output that exhibits both multicollinearity and autocorrelation effects. Besides, selection rules for determining the number of principal components for quality monitoring and principal components interpretation are often not discussed. Bilen et al. [1] presented a general framework for monitoring multivariate autocorrelated processes by combining PCA in conjunction with multiple regression.

In this paper, the integration of PCA and multiple regression is done by using PCA to relieve the multicollinearity effect and to satisfy independent regressors assumption for multiple regression. PCA also assists in reducing the number of input variables to fewer

significant principal components and regressed to get residuals. Because multiple regression residuals are not white noise due to the autocorrelation in the output variable, the autoregressive error correction model is applied to get white noise residuals so that traditional control charts can be used for process monitoring. This paper is an extension of the multivariate autocorrelated monitoring framework to identify significant principal components by integrating three standard variation-based selection rules. Each rule has a different criterion for selecting the number of principal components to be retained. By combining these criteria, significant principal components can be properly identified to ensure development of an accurate multiple regression model. The interpretation of out-of-control signals is also discussed. The proposed methodology is applied to sugar beet pulp drying process that involves a large number of inputs and a single output, where the variables exhibit both autocorrelation and multicollinearity. The general form of the pulp drying sugar process is shown in Figure D.1. The goal is to monitor the output that is affected by many input variables and to investigate the root cause of the out-of control signal.

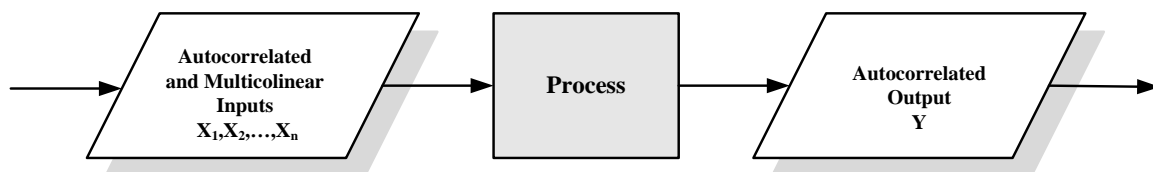


Figure D.1. Process description

D.2. Description of the Method

The methodology for multivariate autocorrelated processes monitoring consists of using multiple regression in conjunction with principal component analysis. Principal component analysis (PCA) with proposed selection rule is used to reduce the number of variables involved in the process. Subsequently, significant principal components are regressed, and the autoregressive error correction model is applied to the multiple regression

equation to generate white noise residual series for process monitoring. The overall methodology consists of four phases which are discussed next.

Phase I. Process Assessment

Individual time series plots are used to determine whether each process variable is stationary. When the joint distribution of any variable, $x(t_1), \dots, x(t_n)$, is the same as the joint distribution of $x(t_1 + \tau), \dots, x(t_n + \tau)$ for all t_1, \dots, t_n, τ the process is said to be stationary. For non-stationary process variables, differentiation of the data series is needed before further analysis. The multicollinearity effect is assessed through scatter plots, and autocorrelation is examined through the autocorrelation function (ACF) and partial autocorrelation function (PACF). If multicollinearity among process variables and autocorrelation within variables do not exist, simple multiple regression can be applied directly if the number of variables in the process is not large.

Phase II. Principal Component Analysis and Selection Rule

Principal component analysis is a method for eliminating multicollinearity and gives better results than the ordinary least squares regression when used successfully. In PCA, the original process variables are transformed into a new set of orthogonal or uncorrelated variables called principal components. The PCA transformation ranks the new orthogonal variables in the order of their importance. A number of principal components are selected to explain the major variability of the process, while others are eliminated to reduce the dimension of the variables.

All variables in PCA are ensured to be in the same unit by standardization. The variables are standardized by $x_{ji}^s = (x_{ji} - \bar{x}_j) / s_j$, where x_{ji} is the i^{th} observation for the j^{th} variable x_j and

$s_j = \sqrt{\frac{\sum_{i=1}^n (x_{ji} - \bar{x}_j)^2}{(n-1)}}$. The principal component is given by

$$z_j = a_{1j}x_1^s + a_{2j}x_2^s + \dots + a_{kj}x_j^s \quad (1)$$

where $[a_{1j}, a_{2j}, \dots, a_{kj}]$ is the eigenvector associated with eigenvalues λ_j and z_j accounts for λ_j of the total variance. The least squares estimator for regression coefficients of vector λ is given by $\hat{\gamma} = (Z'Z)^{-1}Z'Y$ and the variance covariance of the estimated coefficients of $\hat{\gamma}$ is given by

$$Var(\hat{\gamma}) = \hat{\sigma}^2 (Z'Z)^{-1} = \hat{\sigma}^2 \text{diag}(\lambda_1^{-1}, \lambda_2^{-1}, \dots, \lambda_k^{-1}) \quad (2)$$

The principal components (PCs) are ranked according to their eigenvalues. The highest eigenvalue is assigned to be the first PC and the remainder of eigenvalues is listed in descending order with the increasing rank of PC. In principal component analysis, it is critical to find the optimal number of PCs to be included in the multiple regression model without major loss of information. There are many rules for selecting PCs. Each rule has different criteria which leads to different number of PCs retained. Therefore, it is difficult for practitioner to pinpoint the optimal number of PCs. For this reason, we use a PC selection rule derived from the integration of two standard PC selection rules, found in the literature, to increase confidence in choosing the number of PCs. These two rules work well in practice [6]. The first rule is based on the cumulative percentage of total variation of selected PCs, t_m , less than a predetermine cut-off point, t^* . The first rule leaves the unsolved issue of deciding the best cut-off point, t^* . The second rule is based on the size of variances of PCs, λ_j , PC with variance less than 0.7 contains less information than the original variables, thus it is discarded. The sampling variation is already taken into account by the second rule which is not included by the first rule.

The proposed selection rule is integration between modified first rule and second rule. The modification of the first rule is on the cut-off point. The range of cumulative percentage of total variation of 70% to 90% instead of using the exact percentage of cumulative percentage of total variation is used to keep the potential number of PCs as much as possible without loss of major information of the data. Therefore, the number of PCs retained is in

range as well. The range of PCs is simultaneously satisfied with the second rule which is already taken sampling variation into account, so the final number of PCs retained by using the combination of both rules is more accurate without much loss of information. The proposed selection rule is shown.

$$t_m = \frac{100}{k} \sum_{j=1}^m \lambda_j, \quad 70\% \leq t_m \leq 90\% \text{ and } \lambda_j = 0.7 \quad (3)$$

where k is the original number of PCs, m is the number of PCs to be retained and t_m is cumulative percentage of total variation of selected PCs. For example, if the first 3-6 PCs variation is combined and it is in the range of 70% to 90%, then 3-6 PCs can be retained. Furthermore, if the first four PCs have variances or eigenvalues larger than 0.7, then the final number to be retained are only the first four PCs.

Phase III. Multiple Regression and Autoregressive Error Correction Model

A regression model is applied to get a set of white noise residuals. Suppose a process of interest involves k independent variables, x_1, x_2, \dots, x_k , then the multiple regression model is given by

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki} + \varepsilon_i, \quad (i = 1, 2, \dots, n) \quad (4)$$

where y is the response variable and $\varepsilon_i \sim \text{i.i.d. } N(0, \sigma^2)$. In multiple regression, errors are assumed to be independent of each other, but multiple regression on time series data usually results in correlated residuals. The autoregressive error correction is applied to the multiple regression model and accounted for the autocorrelated errors. The autoregressive error correction model is given by

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \dots + \beta_k x_{kt} + v_t \quad (5)$$

$$v_t = -\phi_1 v_{t-1} - \phi_2 v_{t-2} - \dots - \phi_m v_{t-m} + \varepsilon_t \quad (6)$$

where y_t is the response variable, $|\phi_1| < 1$ and $\varepsilon_t \sim \text{i.i.d. } N(0, \sigma^2)$.

Phase IV. Control Charting and Correlation Matrix

Traditional control charts on the white noise residuals are constructed to investigate the out-of-control signals and to monitor future process. The correlation of the variables with the principal component is necessary for interpreting the importance of the k^{th} variable to the i^{th} principal component [4]. The correlation between component z_i and variable x_k is calculated as

$$\rho_{z_i, x_k} = e_{ik} \sqrt{\lambda_i} / \sigma_k \quad (7)$$

where e_{ik} is the coefficient, λ_j is eigenvalue and σ_k is standard deviation. The correlation of the variables with principal components (correlation matrix) are simplified to -1, +1, (-1) and (+1) according to the weight of each variable with respect to the maximum weight variable in each principal component. If the weight of variables in absolute value is greater than half the maximum weight variable, they are indicated as -1 and +1 depending on their signs. If the weight of variables in absolute value is in between a quarter and a half the maximum weight variable, they are indicated as (-1) and (+1).

D.3. Example

In this section, the methodology is used to monitor the sugar beet pulp drying process discussed in Bilen et al. [1]. In the sugar beet pulp drying process, it is critical to monitor the moisture content of the dried pulp that is affected by 12 measurable input variables, pressed pulp moisture (ppm), forced draft amps (fa), induced draft fan speed (ifs), exit temperature (et), induced draft amps (ia), induced draft fan differential pressure (ifd), aspiration air flow (aaf), drum pressure drop (dpd), force draft air flow (faf), throat temperature (tt), feed rate (fr), and dryer drum amps (dda). A detailed discussion of dried pulp quality monitoring at each phase of the methodology is given next.

Phase I: Time series plots of all variables indicate a stationary process. The ACF and PACF plots show that all process variables are autocorrelated. The multicollinearity among input variables is observed from the scatter plots.

Phase II: Due to different units of predictor variables, variables are standardized. SAS statistical analysis software is used to conduct PCA of the standardized data. The eigenvalues for each principal component and correlation between principal components and standardized variables (correlation matrix) are generated [1]. Based on the proposed selection rule, only first four principal components are selected for further analysis. These principal components are used to explain the process without much loss of information of all 12 variables. Obviously, this reduction in the dimension of input variables reduces the complexity of the problem.

Phase III: Since the response variable (dried pulp moisture) is autocorrelated, it violates the independence assumption. So it is important to remove autocorrelation before the model is built by regressing four significant principal components. One key assumption of ordinary regression analysis is that the errors are independent of each other. However, with time series data, the ordinary regression residuals are usually correlated over time. Hence, it is not desirable to use ordinary regression analysis for time series data. Therefore, we build a regression model with an autoregressive series for the random error, accounting for the autocorrelation of the errors.

Phase IV: From the analysis results, the mean of the process is -0.0016 and the standard deviation is 0.6933 . The control limits for the Shewhart chart are computed as $(-2.0815, 2.0783)$. The Shewhart individual control chart of dried pulp residuals is shown in Figure 2. The correlation matrix gives some information regarding the weight of the variables in the principal component, but further interpretation is still needed for understanding the process nature and importance of each variable. A simplified version of the correlation matrix is created to interpret significant variables as shown in Table D.1. From principal component 1 (PC1), eight variables, ppm, fa, ifs, et, ia, ifd, aaf, and dpd, are considered to be important based on their weights with maximum weight variables. Three variables, faf, tt, and fr, are significant in principal component 2 (PC2), even though their importance is minor compared to principal

component 1. Principal component 3 (PC3) shows five variables, fa, ia, ifd, dpd , and dda, as important, and only dda is significant in principal component 4 (PC4). Since all four principal components are orthogonal to each other, the combination of these variables, aaf, ifs, et, ifd, ppm, dpd, fa, ia, faf, fr, tt, and dda, according to their degree of significance, should be investigated further.

Table D.1. Simplified correlation between principal components and standardized variables

z	PC1	PC2	PC3	PC4
ppm	-	(+)	(+)	
faf	(-)	-		
tt	(+)	+		
fr	(+)	+	(+)	
fa	+	(-)	+	(-)
ifs	+	(-)		(+)
et	-	(+)		
ia	-		+	(+)
ifd	+		+	(-)
aaf	-			
dda			+	+
dpd	+		-	(+)

There are several out-of-control signals in the Shewhart control chart given in Figure D.2:

- 1) Points 64 and 228 are beyond control limits. The out-of-control signal at point 64 might be due to process variables fa, fr, faf and tt. At point 228, et and ifd might be the cause of the out-of-control signal.

- 2) At points 181 and 182, nine points consecutively are on the same side of the center line.
This out-of-control signal may be due to variables ifs, dpd, and fa.
- 3) There is an out-of-control signal at point 67 as four out of five points are more than one standard deviation away from the center line (on one side of CL). The signal causes might be due to variables ia, faf, fr, and tt.
- 4) At points 104, 105, 106, 107, 178, and 202, 15 points in a row are within one standard deviation of center line (above and below CL), possibly due to variables aaf, ifs, et, ifd, dpd, fa, and ia.

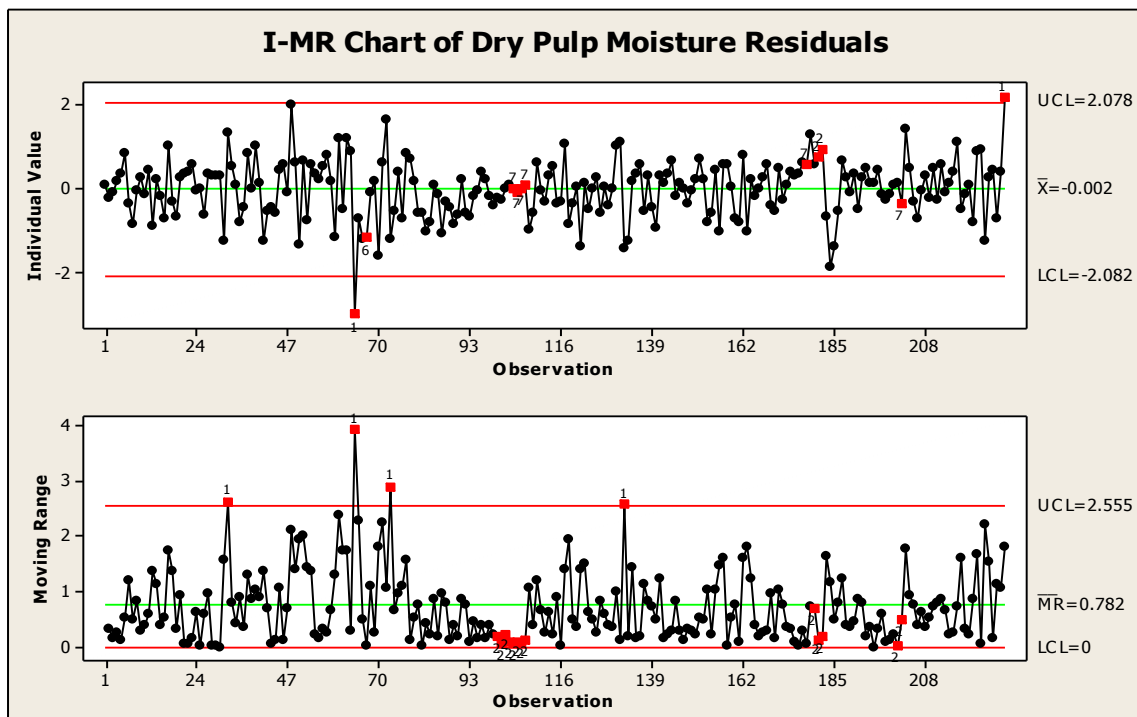


Figure D.2. Shewhart control chart for dry pulp moisture residuals

It is difficult to pinpoint a single source for each of the out-of-control signals since most variables are correlated to each other. A physical interpretation of the results is informative; for example, the aspiration air flow (aaf), is the air that is affected by the pressure difference between temperature at the beginning and temperature at the end of the drum. If there is high pressure in the drum, then the pulp will have less moisture content, and vice versa. Exit temperature (et), is the temperature at the end point of the drum. If the exit temperature is

higher, then the moisture in the pulp will be decreased and throat temp (tt), is the temperature at the beginning of the drum which behaves the same way as exit temperature. Furthermore, the feed rate (fr) and force draft air flow (faf) will affect the duration of the pulp inside the drum, if the feed rate and force draft air flow are increased, the pulp will have less time inside the drum which directly lead to high moisture content in the pulp.

D.4. Conclusion

This paper presented a framework for monitoring multivariate autocorrelated processes with guidelines in handling multicollinearity and autocorrelation effect. The PCA reduced the number of input variables and combated the multicollinearity effect. The autoregressive error correction model along with the multiple regression is applied to get the relationship between output variables and the principal components of original input variables and generate the independent residuals in monitoring the quality characteristic of interest. A real industrial example is considered to demonstrate the applicability of the framework. Use of different rules for principal component determination will lead to different choices of components. Additionally, interpretation of out-of-control signals is a challenge as each principal component is a linear combination of other principal components, and due to the fact that principal components have no direct physical meaning. Comprehensive interpretation of principal components requires prior knowledge of the process which is difficult to obtain in practice. Further study of simulated data is necessary to validate the method and the usage of the calculated control limits for monitoring future process data. Analysis of the effect of the number of components selected and comprehensive physical interpretation of principal components selected are interesting areas for adopting.

D.5. References

1. Bilen, C., Chen, X., Khan, A. and Yadav, O. P., 2007, "Multiple Regression Control Chart Integrated with Principal Component Analysis". Proceeding of the 2007 Industrial Engineering Research Conference, May 19-23, Nashville, TN, USA.
2. Hawkins, D. M., 1991, "Multivariate Quality Control Based on Regression-Adjusted Variables," *Technometrics*, Vol. 33, No. 1, 61-75.
3. Hawkins, D. M., 1993, "Regression Adjustment for Variables in Multivariate Quality Control," *Journal of Quality Technology*, Vol. 25, No. 3, 170-182.
4. Jackson, J. E., 1980, "Principal Components and Factor Analysis: Part I – Principal Components," *Journal of Quality Technology*, Vol. 12, No. 4, 201-213.
5. Jiang, W., 2004, "Multivariate Control Charts for Monitoring Autocorrelated Processes," *Journal of Quality Technology*, Vol. 36, No. 4, 367-379.
6. Jolliffe, I. T., 2002, "Principal Component Analysis," Springer, New York, NY.
7. Kalgonda, A.A. and Kulkarni, S.R., 2004, "Multivariate Quality Control chart for Autocorrelated Processes," *Journal of Applied Statistics*, Vol. 31, No. 3, 317-327.
8. Loredó, E. N., Jearkpaporn, D. and Borrór, C. M., 2002, "Model-Based Control Chart for Autoregressive and Correlated Data," *Quality and Reliability Engineering International*, Vol. 18, 489-496.
9. Mandel, B. J., 1969, "The Regression Control Chart," *Journal of Quality Technology*, Vol. 1, No. 1, 1-9.
10. Mastrangelo, C. M., Runger, G. C. and Montgomery, D. C., 1996, "Statistical Process Monitoring with Principal Components," *Quality and Reliability Engineering International*, Vol. 12, 203-210.
11. Montgomery, D.C. and Friedman, D. J., 1993, "Prediction using Regression Models with Multicollinear Predictor Variables," *IIE Transaction*, Vol. 25, No. 3, 73-85.

12. Runger, G.C., 1996, "Multivariate Statistical Process Control for Autocorrelated Processes," *International Journal of Production Research*, Vol. 34, 1715-1724.
13. Schall, S. and Chandra, J., 1987, "Multivariate Quality Control Using Principal Components," *International Journal of Production Research*, Vol. 25, 571-588.