ROBUST STATISTICS FOR SOFT SENSOR DEVELOPMENT IN CEMENT KILN

Bao Lin[1], Bodil Recke[2], Philippe Renaudat[2]
Jørgen Knudsen[2], Sten Bay Jørgensen[1]

[1] *CAPEC, Department of Chemical Engineering, DTU, Denmark*
[2] *FLS Automation, Valby 2500, Denmark*

Abstract: This paper presents a systematic approach of developing data-driven soft sensor using robust statistical technique. Data preprocessing procedures are described in detail. First, a template defined with a key process variable is used to handle missing data. Second, a univariate, followed by a multivariate approach, principal component analysis (PCA), is used to detecting outlying observations. Then, regression technique is employed to derive an inferential model. The proposed methodology is applied to a cement kiln system for realtime estimation of free lime, demonstrating improved performance over a standard multivariate approach.

Keywords: Regression analysis, Soft sensing, Statistics

1. INTRODUCTION

Soft sensors have been developed as supplement to online instrument measurements for process monitoring and control. Both model-based and data-driven soft sensors have been developed. If a first principle model (FPM) describes the process sufficiently accurately, a model-based soft sensor can be derived. However, a soft sensor based on detailed FPM is computationally intensive for real time applications. Modern measurement technique enables a large amount of operating data to be collected and stored, thereby rendering data-driven soft sensor development a viable alternative.

A data-driven soft sensor is an inferential model developed from process observations. Multivariate regression techniques have been extensively employed to develop an empirical model. Multivariate linear regression (MLR) suffers from numerical problems as well as degraded models when a data set is strongly collinear. Both principal component regression (PCR) and partial least squares (PLS) solve this issue by projecting the original process variables into a smaller number of orthogonal latent variables.

Early work on soft sensor development assumed that a process model was available. The inferential model is developed using Kalman filter (Joseph and Brosilow, 1978). In case the process mechanisms are not well understood, empirical models, such as neural network (Qin and McAvoy, 1992; Radhakrishnan and Mohamed, 2000), multivariate statistical methods are used to derive the regression model (Kresta *et al.*, 1994; Park and Han, 2000; Zhao, 2003).

Process measurements are often contaminated with data points that deviate significantly from the real values due to human errors, instrument failure or changes of operating conditions. Since outlying observations spoil the regression model, robust statistical approaches have been developed to provide reliable results in the presence of abnormal observations. This paper presents a systematic approach for building a soft sensor. The application example is estimation of free lime for cement kilns using robust multivariate methods.

The paper is organized as follows. Section 2 describes both univariate and multivariate approaches to detect outlying observations. The robust PCR and PLS approaches are presented in section 3, followed by the illustrative application on development of a free lime soft sensor for a cement kiln.

2. DATA PREPROCESSING

Outliers are commonly defined as observations that are not consistent with the majority of the data (Pearson, 2002; Chiang *et al.*, 2003), including missing data points or blocks, and observations that deviate significantly from the normal values. Popular multivariate approaches of building data-driven soft sensor, such as PCR and PLS, assume a linear relationship between variables. The performance of derived model deteriorates even in the presence of a single abnormal observation. Process data from cement plants are commonly contaminated, e.g. by abnormal values, which may lead to model misspecification. Therefore, outlier detection constitutes an essential prerequisite step for a data-driven soft sensor design.

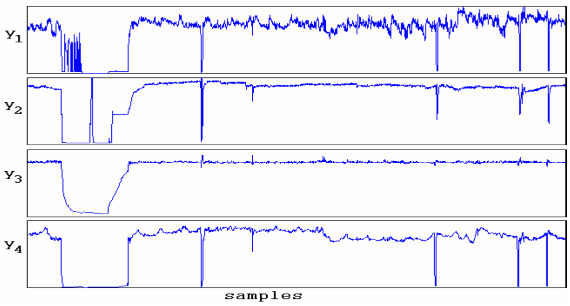

Fig. 1. Segment of operating data from a cement kiln

A heuristic procedure has been implemented in the paper to handle missing data. Figure 1 shows a segment of four process measurements from a cement kiln. A systematic pattern of missing data is observed, which can be identified with a template defined by a key measurement in the dataset. In case a small block (i.e., less than 2 hour) of data is missing, interpolated values using neighbor observations will be inserted. If a large segment of missing data is detected, these blocks will be marked and not be used to build the soft sensor.

Missing data is one type of outliers. The second type denotes abnormal operation conditions. For example, the malfunction of process equipment might cause a change in process measurements that may affect several successive samples. Both univariate and multivariate approaches have been developed to detect these outlying process observations.

A popular approach to detect outliers is the *3σ edit rule* (Ratcliff, 1993). This method labels outliers when data points are three or more standard deviations from the mean. Unfortunately, this procedure often fails in practice because the presence of outliers tends to inflate the variance estimation, causing too few outliers to be detected. The *Hampel identifier* (Davies and Gather, 1981) replaces the outlier-sensitive mean and standard deviation estimates with the outlier-resistant median and median absolute deviation from the median (MAD). The MAD scale estimate is defined as:

$$MAD = 1.4826 \; median \left\{ \left| x_i - x^* \right| \right\} \qquad (1)$$

where x^* is the median of the data sequence. The factor 1.4826 was chosen so that the expected value of MAD is equal to the standard deviation σ for normally distributed data.

Since process measurements from the cement kiln system are not independent from each other, detecting outliers using univariate diagnostics is not sufficient, resulting in *masking* and *swamping*. Masking refers to the case that outliers are incorrectly identified as normal samples; while swamping is the case that normal samples are classified to be outliers. Effective outlier detection approaches are expected to be based on multivariate statistical techniques.

Principal component analysis (PCA) is a multivariate analysis tool. Given a data matrix $\mathbf{X}$ constructed by *m* observations of *n* variables, PCA projects it to a lower dimensional space that explains a large fraction of variability in the original data.

$$\mathbf{X} = \mathbf{t}_1\mathbf{p}_1^T + \mathbf{t}_2\mathbf{p}_2^T \ldots + \mathbf{t}_k\mathbf{p}_k^T + \mathbf{E} = \sum_{i=1}^{k} \mathbf{t}_i\mathbf{p}_i^T + \mathbf{E} \qquad (2)$$

where $\mathbf{E}$ is the residual matrix. The orthogonal loading vectors, $\mathbf{p}_i$, are eigenvectors of the scatter matrix, $\frac{1}{m-1}(\mathbf{X}^{\mathrm{T}}\mathbf{X})$. The score vector, $\mathbf{t}_i$, is the projection of $\mathbf{X}$ along the direction of $\mathbf{p}_i$. The loading vectors corresponding to the *k* largest eigenvalues are retained to optimally capture the variations of the data and minimize the effect of random noise.

The fitness between data and the model can be calculated using the residual, $\mathbf{e}_i$, defined as:

$$\mathbf{e}_i = \mathbf{x}_i - \hat{\mathbf{x}}_i = \mathbf{x}_i(1 - \mathbf{P}_k\mathbf{P}_k^T) \qquad (3)$$

where $\mathbf{e}_i$ is the i^{th} row of $\mathbf{E}$ and $\mathbf{P}_k$ is the matrix of the first *k* loading vectors. The magnitude, $Q_i = \mathbf{e}_i\mathbf{e}_i^T$, indicates how well a sample fits to the PCA model. The significance level for the *Q* statistic is:(Jackson and Mudholkar, 1979)

$$Q_a = \Theta_1 \left[\frac{c_\alpha \sqrt{2\Theta_2 h_0^2}}{\Theta_1} + 1 + \frac{\Theta_2 h_0 (h_0 - 1)}{\Theta_1^2} \right]^{\frac{1}{h_0}} \qquad (4)$$

where

$$\Theta_i = \sum_{j=k+1}^{n} \lambda_j^i, \quad i = 1, 2, 3$$

$$h_0 = 1 - \frac{2\Theta_1\Theta_3}{3\Theta_2^2}$$

λ_i is the eigenvalue

c_α is the normal deviate corresponding to the upper $1-\alpha$ percentile.

Q-statistic measures the variation of a sample outside of the PCA model. Hotellings T^2 statistic provides an indication of variability within the normal subspace. The T^2 value of a sample is equal to the sum of squares of the adjusted (unit variance) scores:

$$T^2 = \sum_{i=1}^{k} \left(\frac{t_i}{\lambda_i} \right)^2 \qquad (5)$$

It indicates the distance that the estimated sample by the PCA model is from the multivariate mean of the data. The statistical significance level for T^2 can be calculated with *F*-distribution:(Wise, 1991)

$$T^2_{k,m,\alpha} = \frac{k(m-1)}{m-k} F_{k,m-k,\alpha} \qquad (6)$$

where α is the standard normal deviate. The combined *Q*- and T^2-tests are used to detect abnormal observations (Chiang *et al.*, 2003). Measurements with $Q > Q_\alpha$ or $T^2 > T^2_{k,m,\alpha}$ are classified as outliers. In this paper the significance level alpha has the same value in the two tests, however finding a compromise between accepting large modelled disturbances and rejecting large unmodelled behaviours for outlier detection clearly needs further investigation.

3. ROBUST STATISTICS

Scaling is an important step in PCA. Since numerically large values are associated with numerically large variance, appropriate scaling methods are introduced such that all variables will have approximately equal weights in the PCA model. *Autoscaling* is commonly used in the absence of a prior knowledge about relative importance of process variables. First, each measurement is mean-centered by subtracting the mean value of the variable. Then the measurement is divided by standard deviation to be scaled down to unit variance. As previously mentioned, both the mean value and the standard deviation are inflated by the outlying observations. Autoscaling is not suitable in handling data which are especially noisy. For such cases, a robust scaling has been suggested (Chiang *et al.*, 2003):

$$\bar{x}_{ij} = \frac{x_{ij} - x_j^M}{x_j^{MAD}} \qquad (7)$$

where x_j^M is the median and x_j^{MAD} the MAD of the variable. This paper applies robust scaling to cement kiln data before performing PCA.

There are two types of approaches for rendering PCA robust. The first is detecting and removing outliers using a univariate approach then carry out a classic PCA on the new data set; the second is multivariate and is based on robust estimation of covariance matrix. In this study, elliposidal multivariate trimming (MVT) (Devlin *et al.*, 1981) approach is used. It iteratively detects bad data based on the squared *Mahalanobis distance*:

$$d_i^2 = (x_i - x_i^*)^T S^{*-1} (x_i - x_i^*) \qquad (8)$$

where x_i^* is the current robust estimation of the location and S^* is the robust estimation of the covariance matrix. 95% of measurements with smallest Mahalanobis distance will be used for the estimation of the covariance matrix in the next iteration. Since the data set has been preprocessed with a *Hampel identifier*, 95% of data are retained. The iteration proceeds till both x_i^* and S^* converge. In this paper, the iteration stops at the 10th iteration such that at least 60% of the data is retained for the estimation of covariance matrix.

Since outlying observations disrupt the covariance matrix from the initialization step. Chiang *et al.* (2003) proposed the closest distance to center (CDC) approach. The most consistent observations are first identified based on their deviation from the center of the data. The *m*/2 observations with smallest distance are used to calculate the mean value, x_i^*. Thus, a PCA model is developed using the samples identified by the CDC/MVT procedure. A regression model is derived with score vectors and free lime measurements from the lab.

Principal component regression (PCR) is a natural extension of PCA to derive an inferential model. The measurement of free lime from lab analysis is assigned to be vector $\mathbf{y}$. Given the decomposition of PCA, $\mathbf{y}$ can be regressed against the matrix of score vectors, $\mathbf{T}_k$ using multivariate linear regression method.

$$\mathbf{b} = (\mathbf{T}_k^T \mathbf{T}_k)^{-1} \mathbf{T}_k^T \mathbf{y} \qquad (9)$$

The coefficients between original data, $\mathbf{X}$ and $\mathbf{y}$ can be obtained straightforwardly as:

$$\mathbf{c} = \mathbf{P}_k \mathbf{b} \qquad (10)$$

PCA model identifies outlying observations. During the regression step, zero weights are assigned to these outlying observations; a weight value of one is assigned to normal data.

PLS is a multivariate statistical approach for relating input data matrix, $\mathbf{X}$ and dependent data block $\mathbf{y}$. PLS can be thought of as a simultaneous decomposition of the $\mathbf{X}$ and $\mathbf{y}$ matrices using PCA. The $\mathbf{X}$-matrix is projected onto a *k*-dimensional hyper-plane such that the coordinates are good predictors of $\mathbf{y}$. The existence of outlying observations can be the source of error in the PLS model that is based on the assumption of linear relationship between input and output matrices. Therefore, the outlying measurements identified with a PCA model is downweighted before PLS analysis.

The proposed approaches, robust PCR and weighted PLS, are applied to a data set collected from the log system of a cement kiln.

In summary, the systematic procedure of applying robust statistical techniques for soft sensor development consists of the following steps:

1. Handle missing data using a template defined with the key process measurement;
2. Detect outliers with a univariate approach (*Hampel identifier*) followed by a multivariate approach (robust PCA) using Q and T^2 tests;
3. Derive regression model with weighted PLS;
4. Validate the soft sensor on other process data.

4. CEMENT KILN SYSTEM

The rotary kiln is the most operationally complex and energy consuming equipment in the cement industry. For most dry processes (as shown in Figure 2), the feed materials are preheated by hot gas from the rotary kiln. A fuel combustion chamber, called precalciner, is integrated in the preheating tower to improve energy efficiency. The mixture of preheated and precalcined materials enters the rotary kiln, where fuel together with air enter from the opposite end.

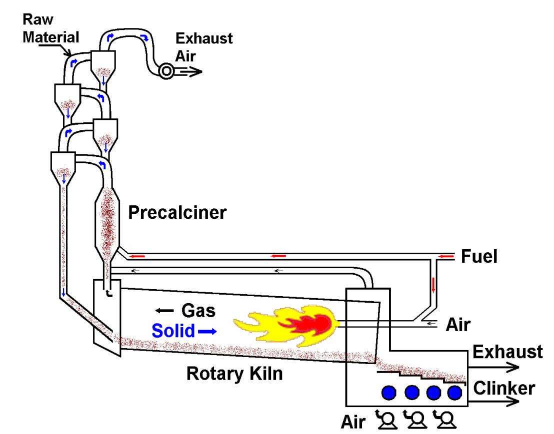

Fig. 2. A typical modern dry kiln system

Several exothermic and endothermic reactions take place in both solid and gas phases. The solid feed is heated to an extremely high temperature (about 1500°C) in the burning zone such that raw materials react and form the nodular clinker. The clinker exits the kiln at about 1200°C, then is cooled down by cross-flowing air in a separate clinker cooler. Partial heat integration is achieved by feeding part of the heated air back into the kiln and part to the precalciner. The operating data of the precalciner and the kiln are used to derive a soft sensor of free lime in the clinker.

The quality of the kiln product is indicated by the amount of free lime. The direct measurement is generally only available with a delay time of about an hour. In addition, the measurement also suffers from the perturbations within the kiln and the cooler, which result in uncertain indication of the average quality. It is desirable to develop a soft sensor that is able to accurately predict the content of free lime in real time, and can be employed for effective quality control.

5. CASE STUDY

The data from a cement kiln log system is used in this study. There are 13 process measurements available, including kiln drive current, kiln feed, fuels to calciner and kiln, plus several temperature measurements within the kiln system. Kiln fuel is the

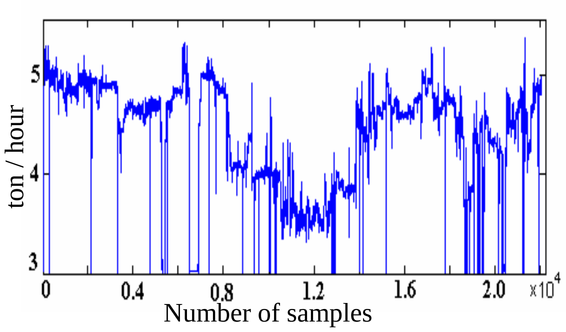

Fig. 3. Kiln fuel flowrate during validation period

main manipulated variable for the kiln system, which varies from between 5 and 3.5 ton per hour. As shown in Figure 3, kiln fuel measurement is heavily contaminated by the missing values and outlying observations. Following the procedure mentioned previously, a template is defined by the kiln drive measurement to identify the pattern of missing observations,

The standard measurements are logged every 10 min, whereas the laboratory analysis of free lime content of the clinker is logged approximately every 2 hours. A data block of 22000 samples is selected in this study. the data block between 8000 and 15500 is used to derive the model and samples between 18000 and 21500 for validation.

The one-step ahead prediction residual sum of squared errors (PRESS) between the model and measured lime content is used to select the number of principal components (PCs):

$$PRESS = \sum_{i=1}^{N_V} \left(\hat{y}(i) - y_m(i) \right)^2 \qquad (11)$$

where N_V is the total number of samples during the validation period. It is calculated only when a new lab measurement is available. The PRESS of the PCR model using 6 PCs has the minimum PRESS (see Figure 4). The 5th PC does not contribute positively to the prediction error, which suggests that this factor is not relevant for predictive ability of the model.

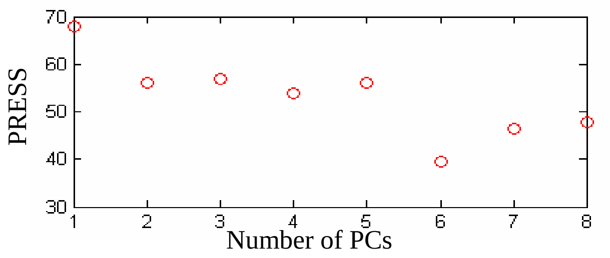

Fig. 4. PRESS of robust PCR model

Improved predictive ability is achieved by including the 6th principal component at the expense of an increased noise level. The validation of PCR models with 6 PCs (PRESS = 39.607) is shown in Figure 5.

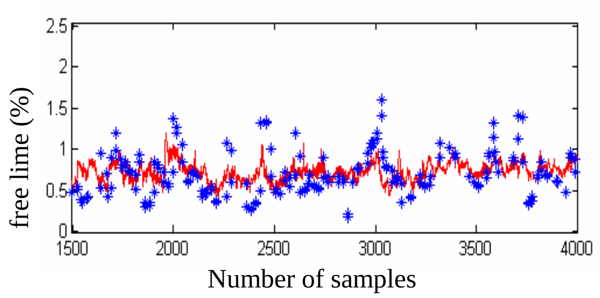

Fig. 5. Validation of robust PCR model with 6 PCs (* - lab measurements; solid line - robust PCR model)

Given the PCA decomposition, zero weights are assigned to abnormal points to downweight these observations before a regression model is derived between score vectors and free lime results. 95% significance level is commonly used for Q- and T^2-tests. The lower the significance level, the higher the chance to reject outlying points, at the risks of rejecting essential process dynamics. Thus, the optimal value of significance level for Q- and T^2-tests should be selected considering the quality of the data. Figure 6 shows the PRESS of robust PCR model using 6 PCs with varied Q- and T^2- test significance level, varying from 100% to 95%. It is obvious that downweighting outlying observations improves the predictable ability of the regression model (see Figure 6). The minimum (PRESS = 38.362) is obtained when the significance level is selected as 97.5%.

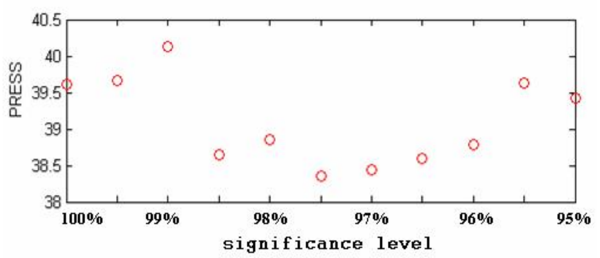

Fig. 6. PRESS of robust PCR model of 6 PCs with significance level varied from 100% to 95%

Another regression model is derived with PLS approach for the same block of operating data. The relation between PRESS and the number of latent variables (LV) is shown in Figure 7.

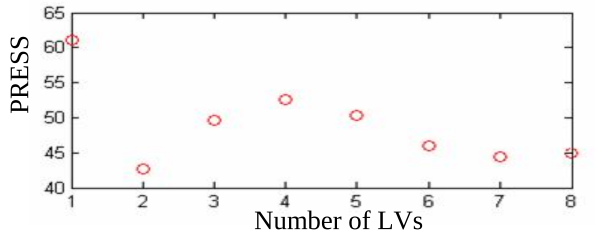

Fig. 7. PRESS of PLS model

The PLS analysis shows a minimum of PRESS at 2 LVs. PLS finds LVs that describe a large amount of variation in $\mathbf{X}$ and are correlated with dependent variables, $\mathbf{Y}$, while the PCs in PCR approach are selected only on the amount of variation they explain in $\mathbf{X}$. Therefore, PLS model is able to capture more of the relevant information than PCR model with a smaller number of LVs. Comparisons of the PLS model with lab measurements during modeling and validation periods are shown in Figure 8 and 9 respectively.

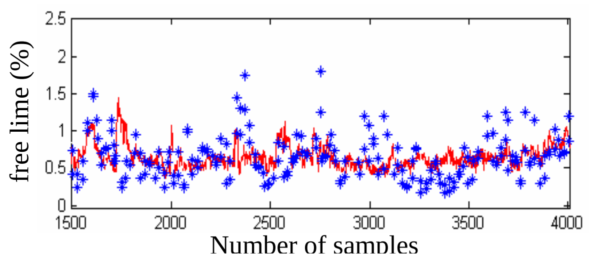

Fig. 8. PLS model with 2 LVs versus lab measurements during modelling period (* - lab measurements; solid line - PLS model)

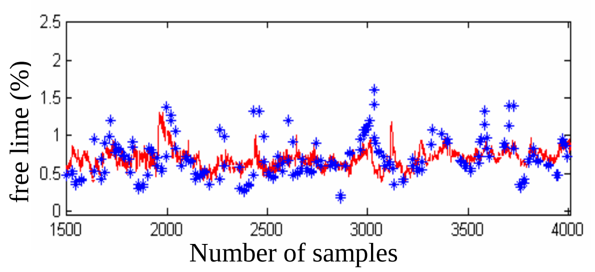

Fig. 9. PLS model with 2 LVs versus lab measurements during validation period (* - lab measurements; solid line - PLS model)

The study with robust PCR models shows enhanced performance using downweight scheme based on the PCA analysis of the data matrix, $\mathbf{X}$. Since PLS model is based on the assumption of linear relationship, the result of PLS analysis is affected by outlying observations in the dataset. Downweighting these data points is beneficial for the PLS model.

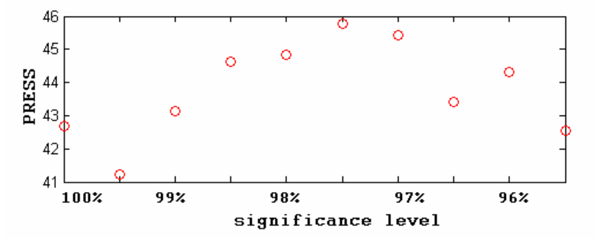

Fig. 10. PRESS of weighted PLS model with 2 LVs significance level varied from 100% to 95%

Then, the weighting vectors obtained during the PCR analysis is applied to the PLS model. The weighting vectors vary with the selection of significance level. Figure 10 shows the PRESS of weighted PLS model versus the change of significance level from 100% to 95%. The minimum is obtained with the significance level 99.5% with a PRESS of 41.236, about 10% less than that of the standard PLS model (PRESS = 45.004).

Results of applying standard and robust PCR, standard and weighted PLS are summarized in Table 1. With a larger number of principal components employed in the development of regression model, sum of squared errors (SSE) during the modeling period of PCR approaches are comparable to those of PLS models. Among all the approaches investigated in this paper, standard PLS model is worst regarding predictive ability. Due to the contamination of outlying observations, PRESS of standard PLS (45.004) is higher than the weighted model (41.236) derived utilizing the weighting vector obtained by PCA analysis. The robust PCR model (38.362) performs slightly better than the PLS model at the cost of using 4 more principal components.

Table 1. Comparison of PCR and PLS models

Approach	*No. of PCs*	*SSE*	*PRESS*
Std. PCR	7	87.658	43.630
Robust PCR	6	87.235	38.362
Std. PLS	4	87.003	45.004
Weighted PLS	2	87.520	41.236

6. CONCLUSIONS

This paper presents a systematic approach to build a soft sensor using robust statistical techniques. The proposed methodology is applied to predict free lime of cement kiln systems. Due to the low signal-to-noise ratio in operating data, data preprocessing demonstrates to be an essential step for development of data-drive soft sensor.

A case study demonstrates the improved performance of robust PCA model in the detection outliers. The real-time estimation of free lime can be obtained with the data-driven soft sensor, which shows some potential to be used in closed loop control. However, the performance of soft sensor depends substantially on the data used to build the inferential model. Due to the contamination of process measurements in operating data, downweighting outlying observations is beneficial to enhance the predictive ability of a regression model. The case study indicates the existence of a optimal downweighting vector determined by the significance level of Q- and T^2-statistics. The issue of finding the optimal significance levels for regression model and integrating the information from irregularly-sampled low quality measurements into the weighting vector needs to be performed in the future.

REFERENCES

Chiang, L.H., R.J. Pell and M.B. Seasholtz (2003). Exploring process data with the use of robust outlier detection algorithms. *Journal of Process Control*, **13**, (5): 437-449.

Davies, L. and U. Gather (1981). The identification of multiple outliers. *Journal of the American Statistical Association*, **88**, 782-801.

Devlin, S.J., R. Gnanadesikan and J.R. Kettenring (1981). Robust estimation of dispersion matrices and principal components. *Journal of the American Statistical Association*, **76**, 354-362.

Jackson, J.E. and G.S. Mudholkar (1979). Control procedures for residuals associated with principal component analysis. *Technometrics*, **21**, (3): 341-349.

Joseph, B. and C. Brosilow (1978). Inferential control of processes. *AIChE J.*, **24**, 485-509.

Kresta, J.V., T.E. Marlin and J.F. MacGregor (1994). Development of inferential process models using PLS. *Computers and Chemical Engineering*, **18**, (7): 597-611.

Park, S. and C. Han (2000). A nonlinear soft sensor based on multivariate smoothing procedure for quality estimation in distillation columns. *Computers and Chemical Engineering*, **24**, 871-877.

Pearson, R.K. (2002). Outliers in process modeling and identification. *IEEE Transactions on Control Systems Technology*, **10**, (1): 55-63.

Qin, S.J. and T.J. McAvoy (1992). Nonlinear PLS modeling using neural networks. *Computers and Chemical Engineering*, **16**, (4): 379-391.

Radhakrishnan, V.R. and A.R. Mohamed (2000). Neural networks for the identification and control of blast furnace hot metal quality. *Journal of process control*, **10**, (6): 509.

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin*, **114**,: 510-532.

Wise, B.M. (1991). Adapting multivariate analysis for monitoring and modeling dynamic systems, *PhD Thesis*, University of Washington.

Zhao, Y.-H. (2003). A soft sensor based on nonlinear principal component analysis. *2003 International Conference on Machine Learning and Cybernetics*.