

OBSERVER-BASED AND REGRESSION MODEL-BASED DETECTION OF EMERGING FAULTS IN COAL MILLS

Peter Fogh Odgaard* Bao Lin**
Sten Bay Jørgensen**

* *Department of Control Engineering, DK-9220 Aalborg
University, Aalborg, Denmark, odgaard@control.aau.dk*

** *CAPEC, The Technical University of Denmark,
DK-2800 Kgs. Lyngby, Denmark, {bal,sbj}@kt.dtu.dk*

Abstract

In this paper three different approaches for fault detections are compared on example with coal mill used at a power plant. The compared methods are based on: an optimal unknown input observer, static and dynamic regression model-based detections. These approaches are compared on data from a coal mill, where a fault emerges during the test set. The conclusion on the comparison is that observer-based scheme detects the fault 13 samples earlier than the dynamic regression model-based method, and that the static regression based method is not usable due to generation of far too many false detections.

Keywords: Fault Detection, Coal-fired Power Plants, Optimal Unknown Input Observer, Dynamical Regression model-based detection

1. INTRODUCTION

Due to increased performance requirements of power plants, reliability of these are of increased importance. An important method to improve reliability of the power plant, is to detect the faults as they emerge and then handle the detected faults. An example of such a fault is an unexpected increase in the moisture content of the coal, which might result in a decreased coal flow out of the coal mill. In most coal mills it is not possible to measure the pulverized coal flow out of the mill. The coal flow into the mill is only indirectly measurable implying that the early fault detection based on detection of drops in the coal flow of the coal mill, is not so simple as it seems. A coal mill is illustrated in Fig. 1.

Dynamic modeling of these coal mills have been the topic of numerous of publications. Some ex-

amples dealing with modeling of coal mills are (Rees and Fan 2003), (Zhang *et al.* 2002) and (Tigges *et al.* 1998). High order dynamic models and observer design for coal mills are the topics in (Fukayama *et al.* 2004). In (Odgaard and Mataji 2005) a method for detecting emerging changes in the coal moisture content is designed. The suggested scheme is based on a optimal-unknown-input observer scheme, see (Chen and Patton 1999).

However, if models are not available of a system, (in this case the coal mill), one could use regression based methods, see (Venkatasubramanian *et al.* 2003). Modern measurement technologies enable a large amount of operating data to be collected and stored, thereby rendering data-driven approach a viable alternative for fault detection. A two-step procedure is commonly employed for data-driven fault detection. A model that repre-

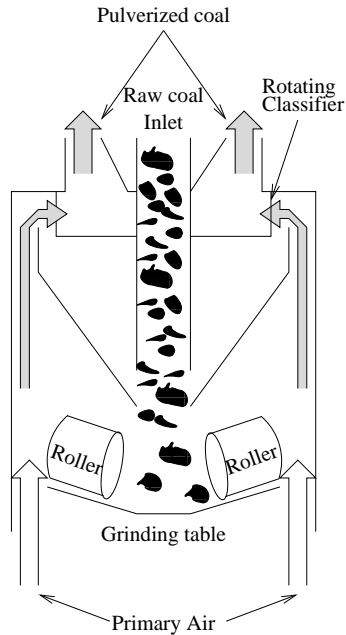


Figure 1. An illustration of the coal mill.

sents the normal operation condition (NOC) is first developed; then faults are detected according to the residual information between model prediction and process measurements.

This paper is organized as follows. Section 2 describes the coal mill process. Observer-based fault detection method is presented and designed in Section 3. Section 4 describes the multivariate data-driven fault detection approaches, including static and dynamic PCA/PLS models, followed by an application of both observer-based and regression-based fault detection schemes to a coal mill process.

2. THE COAL MILL

The work presented in this paper, is based on a Babcock MPS 212 coal mill used at Elsam's Nordjyllandsværket Unit 3. However, the method proposed in the paper is so generic that it can be applied to other types of coal mills. The coal mill is illustrated in principles in Fig. 1. The coal is fed to the coal mill through the central inlet pipe. The coal is pulverized on the rotating grinding table by the rollers. The pulverized coal is then blown up and the moisture content is evaporated by the hot primary air. The primary air is mixed by cold outside air and heated outside air, which is heated by the furnace. The ratio of these air flows are used to control the temperature of the primary air flow. Coal particles which in the pulverizing have been small enough will pass through the classifier and out through the outlet pipes into the boiler. References to coal flow and primary air flow are given by the general power plant controller, as well as rotational speed of the classifier. The temperature of the primary air is used to control the temperature in the coal mill at

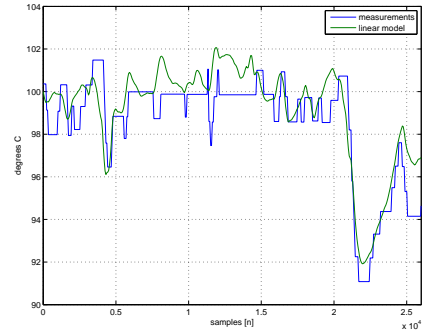


Figure 2. A plot of the linear model response compared with measurements of a step response on the coal mill.

the classifier. The temperature controller is often required to keep temperature constant at 100°C in order to evaporate the moisture content in the coal. A coal mill is a harsh environment in which it is difficult to perform measurements, this means that all the variables are not measurable, e.g. the actual coal flows in and out of the coal mill are not measurable. However, the primary air flow and temperature are, as well as the temperature at the classifier.

A number of different faults can occur in the coal mill, and if the fault leads to a decrease in the output coal flow from the coal mill, this can lead to drop out of the entire power plant unit. In this paper the fault in question is due to increased coal moisture content.

3. OBSERVER BASED DETECTION SCHEME

This specific fault detection scheme for detecting increased coal moisture content is derived in (Odgaard and Mataji 2005). It consists of an optimal unknown input observer, see (Chen and Patton 1999) based on simple energy balance model of the coal mill. The observer estimates a variable corresponding to energy lack due to the emerging fault.

3.1 Coal mill energy model

A simple energy balance model of the coal mill is derived in (Odgaard and Mataji 2005), this model is based on a more detailed model found in (Rees and Fan 2003). In this model the coal mill is seen as one body with the mass m_m . The following variables are defined: $T(t)$ is the temperature in the coal dust flow in the mill, $\dot{m}_{PA}(t)$ is the primary air mass flow, $T_{PA}(t)$ is the temperature of the primary air flow, $\dot{m}_c(t)$ is the mass flow of the coal flow, and $\gamma(t)$ is the moisture content of the coal flow. In (Odgaard and Mataji 2005) the linear model is derived, and can be seen in (1).

$$\dot{\bar{T}}(t) = \mathbf{A}\bar{T}(t) + \mathbf{B} \cdot \begin{bmatrix} \bar{m}_{\text{PA}}(t) \\ \bar{T}_{\text{PA}}(t) \\ \bar{m}_{\text{c}}(t) \\ \bar{\gamma}(t) \end{bmatrix} + \mathbf{q}(t), \quad (1)$$

$$\bar{T}_{\text{m}}(t) = \mathbf{C}\bar{T}(t) + \mathbf{r}(t), \quad (2)$$

where a given signal \circ is linearized by $\bar{\circ} = \circ - \circ_{\text{o}}$, \circ_{o} is the operation point of \circ , $\mathbf{q}(t)$ is the normal distributed process disturbances, $\mathbf{r}(t)$ is the normal distributed measurement noises, $T_{\text{m}}(t)$ is the measured temperature and

$$\mathbf{A} = \left[\frac{(-\dot{m}_{\text{PA},\text{o}} \cdot C_{\text{air}} - \dot{m}_{\text{c},\text{o}} \cdot (C_{\text{c}} + \gamma_{\text{o}} \cdot H_{\text{st}}))}{m_{\text{m}} \cdot C_{\text{m}}} \right], \quad (3)$$

$$\mathbf{B} = \begin{bmatrix} \frac{C_{\text{air}} \cdot (T_{\text{PA},\text{o}} - T_{\text{o}})}{m_{\text{m}} \cdot C_{\text{m}}} \\ \frac{C_{\text{air}} \cdot \dot{m}_{\text{PA},\text{o}}}{m_{\text{m}} \cdot C_{\text{m}}} \\ \frac{C_{\text{c}} \cdot (T_{\text{s}} - T_{\text{o}}) + \gamma_{\text{o}} \cdot (C_{\text{w}} \cdot T_{\text{s}} - H_{\text{st}} \cdot T_{\text{o}})}{m_{\text{m}} \cdot C_{\text{m}}} \\ \frac{\dot{m}_{\text{c},\text{o}} \cdot (C_{\text{w}} \cdot T_{\text{s}} - H_{\text{st}} \cdot T_{\text{o}})}{m_{\text{m}} \cdot C_{\text{m}}} \end{bmatrix}^T, \quad (4)$$

$$\mathbf{C} = \mathbf{I}. \quad (5)$$

The model parameters are found by partly by data sheets as wells as experiments. The response of this model is compared with measurement in Fig 2, more informations on the parameters can be found in (Odgaard and Mataji 2005).

3.2 Residual generation

The energy balance given by (1-6) would in case of a fault deviate from the coal mill which indicates an unbalance in the model. The unbalance in the model can be represented by an additional fault input. In order to estimate the need energy/fault signal it is more convenient to represent it by an internal state in the model. The simplest representation is a first order low-pass filter which is included in the model. This results in the linear model given by (7-12),

$$\begin{bmatrix} \dot{\bar{T}}(t) \\ \text{bar}Q(t) \end{bmatrix} = \mathbf{A}_{\text{q}} \begin{bmatrix} \bar{T}(t) \\ \bar{Q}(t) \end{bmatrix} + \mathbf{B}_{\text{q}} \begin{bmatrix} \bar{m}_{\text{PA}}(t) \\ \bar{T}_{\text{PA}}(t) \\ \bar{m}_{\text{c}}(t) \end{bmatrix} + \mathbf{E}_{\text{q}}Q_{\text{n}}(t) + \mathbf{q}(t), \quad (7)$$

$$\bar{T}_{\text{m}}(t) = \mathbf{C}_{\text{q}} \begin{bmatrix} \bar{T}(t) \\ \bar{Q}(t) \end{bmatrix} + \mathbf{r}(t), \quad (8)$$

where $Q(t)$ is the estimated residual (need energy flow to balance the model), $Q_{\text{n}}(t)$ is the generic

unknown input which is low-pass filtered in order to represent the residual, and

$$\mathbf{A}_{\text{q}} = \begin{bmatrix} \mathbf{A} & \frac{1}{p} \\ 0 & -p \end{bmatrix}, \quad (9)$$

$$\mathbf{B}_{\text{q}} = \begin{bmatrix} \mathbf{B}_{1 \times (1 \dots 3)} \\ \mathbf{0} \end{bmatrix}, \quad (10)$$

$$\mathbf{C}_{\text{q}} = [\mathbf{C} \ 0], \quad (11)$$

$$\mathbf{E}_{\text{q}} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad (12)$$

where p is the pole of internal residual model, $\mathbf{B}_{1 \times (1 \dots 3)}$ denotes the three first elements of \mathbf{B} . The model represented by (7-12) is discretized before an observer is designed to estimate the states in the model, and hereby the residual, see (13-14).

$$\begin{bmatrix} \bar{T}[n+1] \\ \bar{Q}[n+1] \end{bmatrix} = \mathbf{A}_{\text{d}} \begin{bmatrix} \bar{T}[n] \\ \bar{Q}[n] \end{bmatrix} + \mathbf{B}_{\text{d}} \begin{bmatrix} \bar{m}_{\text{PA}}[n] \\ \bar{T}_{\text{PA}}[n] \\ \bar{m}_{\text{c}}[n] \end{bmatrix} + \mathbf{E}_{\text{d}}Q_{\text{n}}[n] + \mathbf{q}[n], \quad (13)$$

$$\bar{T}_{\text{m}}[n] = \mathbf{C}_{\text{d}} \begin{bmatrix} \bar{T}[n] \\ \bar{Q}[n] \end{bmatrix} + \mathbf{r}[n], \quad (14)$$

where \mathbf{A}_{d} , \mathbf{B}_{d} , \mathbf{C}_{d} , \mathbf{E}_{d} are the discretized representations of \mathbf{A}_{q} , \mathbf{B}_{q} , \mathbf{C}_{q} , \mathbf{E}_{q} . This model is a system with an unknown input, disturbances and measurement noises, i.e. an optimal unknown input observer would be an obvious observer to use for estimating the residual.

3.3 Optimal unknown input observer

The optimal unknown input observer is described in (Chen and Patton 1999). For discrete time systems with unknown inputs and disturbances which can be represented by

$$\mathbf{x}[n+1] = \mathbf{A}_{\text{n}}\mathbf{x}[n] + \mathbf{B}_{\text{n}}\mathbf{u}[n] + \mathbf{E}_{\text{n}}\mathbf{d}[n] + \mathbf{q}[n], \quad (15)$$

$$\mathbf{y}[n] = \mathbf{C}_{\text{n}}\mathbf{x}[n] + \mathbf{r}[n], \quad (16)$$

an optimal unknown input observer of the following form can be derived

$$\mathbf{z}[n+1] = \mathbf{F}_{n+1}\mathbf{z}[n] + \mathbf{T}_{n+1}\mathbf{B}_{\text{n}}\mathbf{u}[n] + \mathbf{K}_{n+1}\mathbf{y}[n], \quad (17)$$

$$\hat{\mathbf{x}}[n+1] = \mathbf{z}[n+1] + \mathbf{H}_{n+1}\mathbf{y}[n+1]. \quad (18)$$

The specific observer is designed in (Odgaard and Mataji 2005). The fault is detected by a simple threshold rule, meaning if the fault estimate is larger than the threshold β a fault is detected at the specific sample. The threshold is chosen such that it do result in false detection eg. from high frequently variations in the residual.

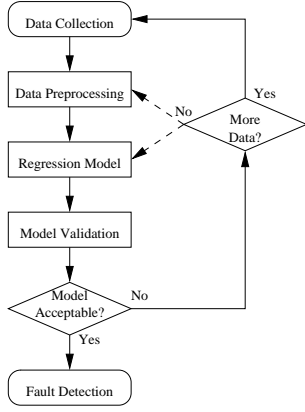


Figure 3. Generic framework for data-driven fault detection.

4. DATA-DRIVEN APPROACHES FOR FAULT DETECTION

The key component of a data-driven fault detection system is a NOC model that is developed through multivariate regression techniques. Then, fault subspace is defined based on the residual between the regression model and process measurements. If deviations are larger than a predefined confidence level, the process is deemed to be operated outside the range of normal condition, and a fault is detected. Both principal component regression (PCR) and partial least squares (PLS) have been extensively employed to develop an empirical model from operational data. The selection of appropriate number of variables (LVs) maximizes process variations while minimizing noise content. One-step-ahead prediction residual sum of squared errors (PRESS) between the model, $\{\hat{y}(i)\}$, and the process measurement, $\{y_m(i)\}$, evaluated on validation data is used to choose the number of LVs

$$\text{PRESS} = \sum_{i=1}^{N_v} (\hat{y}(i) - y_m(i))^2, \quad (19)$$

where N_v is the total number of samples during the validation period.

Process measurements are often contaminated with data points that deviate significantly from real values due to instrument failure or changes of operating conditions, which significantly degrade the regression model. Therefore, robust statistical approaches were employed to pre-process process measurements in order to provide reliable results in the presence of abnormal observation (Lin *et al.* 2005). The generic approach for data-driven fault detection can be summarized in Fig. 3. 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.

$$\mathbf{X} = \sum_{i=1}^k \mathbf{t}_i \mathbf{p}_i^T + \mathbf{E}, \quad (20)$$

where \mathbf{E} is the residual matrix, \mathbf{p}_i and \mathbf{t}_i are loading and score vectors. 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 a data point and the model is calculated using the residual vector, \mathbf{e}_i .

The magnitude, $\mathbf{Q}_i = \mathbf{e}_i \mathbf{e}_i^T$, indicates how well a sample fits to the PCA model. (Jackson and Mudholkar 1979). Q -statistic measures the distance of a sample from 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.

PLS is a multivariate statistical approach for relating input data matrix, \mathbf{x} and dependent data block \mathbf{y} . A regression model is determined by choosing the appropriate number of latent variables (LVs). The deviation between the process measurement and the NOC model prediction is used to detect faults in the process.

The fundamental assumption of the PLS approach is that the data matrix is not correlated in time. However, operating data commonly show auto-correlation due to process dynamics and feedback loops. The standard PLS approach only constructs a linear static mapping from the data matrix, thereby it can not reveal the dynamic relations between process variables, which is required to represent the dynamical system. A dynamic PLS (DPLS) model is obtained by augmenting the original data block with time lagged variables. Similarly, fault detection using dynamic PCA model is also investigated.

5. EXPERIMENTS

The experimental work in this paper consists of two parts. I) experiments with optimal-unknown observer scheme, and II) design and experiments of the PLS/PCA model based methods, in the end the results on these methods are compared. In practice only data of one given fault type is available. In this case the moisture content is dramatically increased. The sample frequency in this experiment is 60s. By visual inspection of the data from the coal mill the beginning of this fault can be detected at sample 160. The applied methods are consequently compared on this data set.

5.1 Experiments with the observer based approach

The detection scheme is designed in Section 3. The introduced fault detection methods based on the estimated residual can detect a number of different faults in the coal mill. The estimated residual $\hat{Q}[n]$, can for this given fault be seen in Fig. 4. Using this method the fault is detected at sample 147, i.e. 13 samples and whereby 13 minutes before the visual inspection indicates.

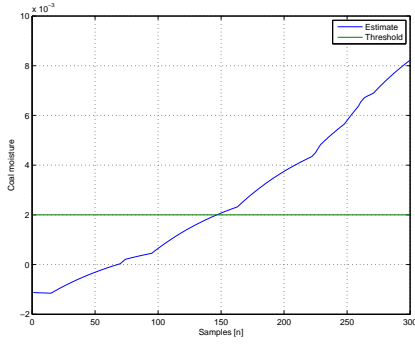


Figure 4. Fault detection based on unknown input observer estimate of $\hat{Q}[n]$.

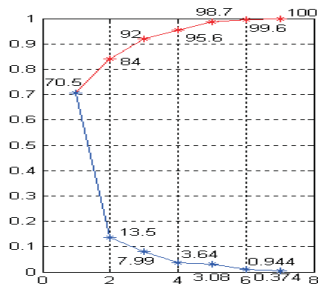


Figure 5. Percentage of variations captured by static PCA model.

5.2 Experiments with and design of the regression model-based approach

Operating data from a coal mill is used to compare the fault detection observer-based method and PCA/PLS models based approach. There are 13 process measurements available representing different temperature, mass flows, pressures, speed etc in the coal mill.

The measurement is not updated, if the variation is less than 1%. The variations of $T(t)$ is in the major part of the operational time inside this interval. Therefore, it is not suitable to be chosen as the predictor variable. However, the variations can be extracted from the $T_{PA}(t)$, which is used to control the temperature of the mill. Therefore, the PLS model is developed with the temperature of the mill as the dependent variable. In addition 6 of the other variables are chosen as regressors since there is barely information in the remainder.

600 samples that represent normal conditions are used to develop the NOC model. 300 observations are used for choosing number of LVs by cross-validation. The test dataset consists of 300 samples.

A static PCA model is first developed, which captures around 99% of variations with 5 PCs (see Fig. 5), which indicates strong collinearity among regressors. As shown in Fig. 6, both Q and T^2 statistics (with 95% confidence level) of the static PCA model are noisy, which potentially lead to false alarms. A static PLS model with 2 LVs achieves the minimal PRESS (see Fig. 7), which is applied to the test dataset. Fig. 8 shows

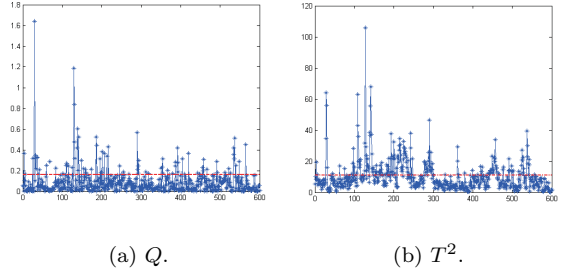


Figure 6. Q and T^2 statistics of the static PCA model with 5 PCs.

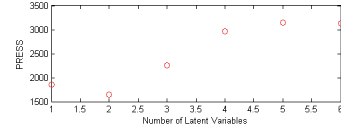


Figure 7. PRESS vs. the number of latent variables of static PLS model.

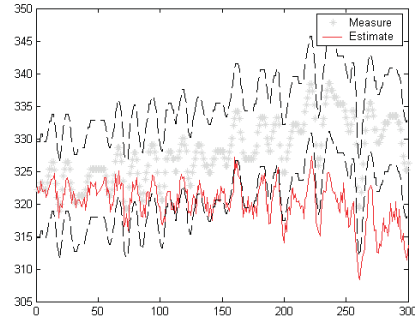


Figure 8. Fault detection with the static PLS model of 2 LVs.

the comparison between process measurement and the static PLS model prediction, together with the 95% confidence level. The process gradually drifts away from the NOC model, which eventually moves beyond the threshold around the sample 150. Due to the noise involved in the prediction signal, the estimation moves in and out the threshold from 110 till 200, when it is clearly out of the confidence level. Both Figs. 6 and 8 reveal that static PCA and PLS models may lead to false alarms due to the noisy estimation. In addition, process measurements are commonly auto-correlated, this behaviour is expected since the coal mill runs dynamical. Thus, dynamic models are developed by including time lagged process measurements, to address the issue of auto-correlations and reduce the possibility of false alarms due the none modelled dynamics.

Including time lagged terms enhance the NOC model by including historical data. However, time lagged terms also introduce additional noise into the modelling data block. For example, including $n + 1$ time lagged terms might lead to poorer validation performance than the model with n terms due to measurement noise. Therefore, PRESS is used to choose appropriate number of time lagged terms for a dynamic PLS model.

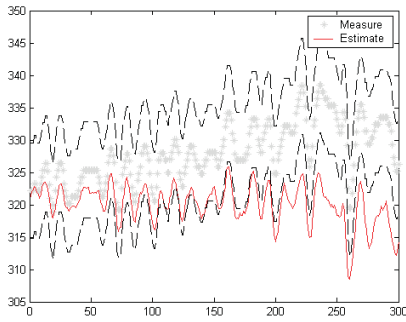


Figure 9. Fault detection with dynamic PLS model of 3 LVs using 8 time lagged terms.

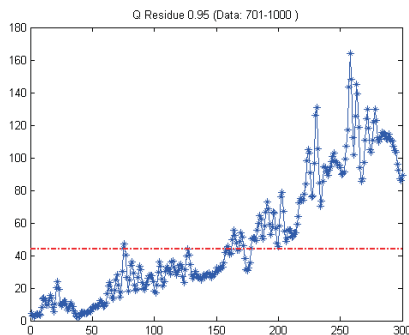


Figure 10. Fault detection with dynamic PCA model of 2 PCs using 8 time lagged terms.

The predictive ability of the PLS model is improved with the inclusion of time lagged terms. The PRESS decreases from 1.645 to the minimal value of 1.142, which is obtained with a dynamic PLS of 3 LVs using 8 time lagged terms. The application of the dynamic PLS model to the test data reveals that the fault occurs in the process around sample 160. Fig. 9 also shows a much smoother prediction such that the possibility of false alarms is significantly reduced. A dynamic PCA model is developed by the inclusion of 8 time lagged terms. The number of PCs is chosen as 2 through cross-validation, which explains 70.6% of process variations. The Q statistic of the dynamic PCA model is shown in Fig. 10, the fault is detected around 160 samples, which is consistent with the dynamic PLS model.

5.3 Summary of experimental work

These experiments show that a static method is of no good use, since it generates too many false detections. Both dynamic methods detect the fault as it emerges, however, the observer based method detects this specific fault earlier, approximately 13 samples earlier than the regression model-based method. This might depend on the chosen parameters in the two methods. In addition one should notice the model order this is as well significant lower for the observer based method compared with the regression model-based approach.

6. CONCLUSION

This paper compares some different approaches for fault detection on a specific application: detection of faults in a power plant coal mill. The compared methods are: optimal-unknown-input observer, static and dynamic regression model-based approaches. The methods are applied to a specific data set with an emerging fault in the coal mill. From this test it can be seen that a dynamic approach is required. The two dynamical method detect the fault as emerge, however, the observer based scheme detects the fault 13 samples earlier. The observer based scheme do as well require a lower model order, than the dynamical regression approach. But the data-driven dynamical regression model based approach is a good alternative if a model of the system is not aviable.

7. ACKNOWLEDGMENT

The authors acknowledge the Danish Ministry of Science Technology and Innovation, for support to the research program CMBC (Center for Model Based Control), grant no 2002-603/4001-93.

REFERENCES

- Chen, Jie and R. J. Patton (1999). *Robust model-based fault diagnosis for dynamic systems*. first ed.. Kluwer academic publishers.
- Fukayama, Y., K. Hirasawa, K. Shimohira and H. Kanemoto (2004). An adaptive state estimator for pulverizer control using moments of particle size distribution. *IEEE Transaction on Control System Technology* **12**, 797–811.
- Jackson, J.E. and J.E. Mudholkar (1979). Control procedures for residuals associated with principal component analysis. *Technometrics* **21**, 341–349.
- Lin, B., B. Recke, J. Knudsen and S.B. Jørgensen (2005). Robust statistics for soft sensor development. In: *Proceedings of 7th World Congress of Chemical Engineering*. Glasgow, UK.
- Odgaard, P.F. and B. Mataji (2005). Estimation of moisture content in coal in coal mills. Submitted for publication.
- Rees, N.W. and F.Q. Fan (2003). Modelling and control of pulverised fuel coal mills. In: *Thermal power plant simulation and control* (D. Flynn, Ed.). first ed. Institution of Electrical Engineers.
- Tigges, K.D., W. Bischoff and T. Steinhage (1998). Walzenschüsselmühlen als komponenten moderner feuerungstechnik. *VGB KraftwerksTechnik* **78**, 77–88.
- Venkatasubramanian, V.R., R. Rengaswamy, K. Yin and S.N. Kavuri (2003). A review of process falt detection and diagnosis part i: Quantitive model-based methods. *Computers & Chemical Engineering* **27**, 293–311.
- Zhang, Y.G., Q.H. Wu, J. Wang, G. Oluwande, D. Matts and X. X. Zhou (2002). Coal mill modeling by machine learning based on onsite measurements. *IEEE Transactions on Energy Conversion* **17**, 549–555.