Fault Diagnosis for Blast Furnace Ironmaking Process Based on Two-stage Principal Component Analysis

Tongshuai ZHANG,1) Hao YE,1)* Wei WANG1) and Haifeng ZHANG2)

1) Tsinghua National Laboratory for Information Science and Technology (TNList), Department of Automation, Tsinghua University, Beijing, 100084 China. 2) Liuzhou Iron & Steel Co., LTD, Liuzhou, 545002 China.

(Received on February 13, 2014; accepted on June 30, 2014)

Monitoring an ironmaking process is a very challenging task as it often fluctuates frequently and lacks of direct measurements. Principal component analysis (PCA) technique has been widely used in various industrial fields, mainly due to its advantage of not requiring the information about the principle knowledge of the process and faults. However, the PCA based application results in ironmaking process are still limited. In this paper, based on the dataset collected from a real blast furnace with a volume of 2,000 m³, a fault diagnosis method by incorporating the PCA technique in two stages will be presented. To overcome the adverse effects of the peak-like disturbances caused by switching between two distinct hot-blast stoves, they are identified and removed from the dataset through the first-stage PCA. Experimental results show that our method outperforms the existing algorithm and the operators’ monitoring in detecting the getting cold accident of the blast furnace.

KEY WORDS: blast furnace; ironmaking process; principal component analysis; fault diagnosis; process monitoring.

1. Introduction

Blast furnaces are a class of huge high-temperature reactors producing molten iron for primary steelmaking in modern iron and steel industry.1–3) In a typical ironmaking process,4–6) iron-bearing materials (including iron ores, sinters, pellets, etc.), cokes and flux (such as limestone and dolomite) are continuously dumped into the blast furnace from the top, while dry air, enriching oxygen, moisture, fuels like tar or pulverized coal are blasted into the furnace from the bottom. The output consisting of slag and molten iron flows out of the furnace from the bottom. Thus the chemical reactions in the blast furnace are quite complex with one stream downward and the other upward.

In order to keep a blast furnace working at its normal status steadily,1,3) prompt detection of different abnormality conditions is vitally important. However, monitoring the status of a blast furnace is exceptionally difficult due to the lacks of its accurate mathematical model and the direct measurements for its internal states.

Most of the existing fault diagnosis methods for the blast furnaces are based on expert systems.4–6) But they could deliver desired performances only when comprehensive rules and sufficient historical fault information are available. In recent years, some data-driven methods have been proposed, including Support Vector Machine (SVM),7–12) neural network4,13–15) and state space model16) assisted approaches. However, most of them require that a large amount of a priori information about the faults can be obtained,16) which is not an easy task in real ironmaking process.

It is known that multivariate statistical process control (MSPC) is an important class of data-driven methods. The main advantage of such schemes with comparison to the aforementioned approaches is that neither a priori fault information nor the system model is necessary. Although there are some results reported on MSPC based applications in ironmaking field, they are mainly focused on the prediction of silicon content17–19) and modeling20–22) problems.

It is worth mentioning that Gamero’s work23) and Vanhatalo’s work24) are the representative of limited works on fault diagnosis of the ironmaking process by adopting PCA, which is a typical MSPC technique. In Gamero’s work, the qualitative trend analysis was addressed.23) In Vanhatalo’s work, an experimental blast furnace was considered.24) In fact, PCA based methods have been widely used for process control and fault diagnosis in chemical industry. MacGregor et al. first used PCA for process monitoring in industrial applications.25) Dunia and Qin employed PCA technique to realize sensor fault diagnosis.26) In the field of steel production, Dudzic et al. applied PCA to process monitoring and quality control.27–29) Similar to the advantages of MSPC methods as discussed above, PCA technique is an attractive tool due to the following two reasons: i) it needs less principle knowledge and causal relationship of the process compared with those model-based approaches; ii) in contrast to the expert systems and SVM based methods, it does not require a priori information of the faults.

Inspired by related literature,23,24) and the successful applications of PCA in other process control fields, we...
adopt the PCA technique for solving the problem of detecting the getting cold accidents for a real blast furnace with a volume of 2 000 m³. Note that the data collected from the blast furnace in this paper is disturbed by the switching between any two distinct hot-blast stoves for blowing the hot air into the blast furnace. This constitutes the main difference between our work and earlier work. 23, 24] To discriminate the disturbances caused by stove switching and the real abnormal status, we adopt a two-stage PCA based monitoring method. The first-stage PCA is used to identify, locate and remove the switching disturbances, while the second-stage PCA adopts the new dataset by removing the stove switching invoked disturbances to build the process model. Then the Hotelling’s T^2 and SPE statistics are utilized to monitor the process online. It is known that T^2 statistic reflects the variation in each observation within the model by calculating squared Mahalanobis distance 30) projected in the principal space between the observation and sample center. SPE statistic refers to the squared prediction error from the residual space, which measures the difference between an observation and its projection into the principal space retained in the PCA model.

2. A Brief Introduction to PCA Based Fault Diagnosis

Let X ∈ R^{p×p} be a data matrix, where the L rows are the observations and the p columns denote the variables. Normally in practice, the kth row of X represents the measurement of all variables collected at the kth sampling instant. Suppose that each column is mean centered and scaled, 25) i.e. it is obtained by subtracting the sample mean from each individual sample and then dividing the difference by the sample standard deviation. With the PCA technique, X can be decomposed as follows: 31)

\[ X = TP^T = \sum_{k=1}^{s} t^k p^k T + E, \quad \text{.............. (1)} \]

where P ∈ R^{p×s} is a loading vector which is defined as the right singular vector of X corresponding to the s largest singular value; t_i is a score vector which is defined as the values of the s principal components for the L observation vector (i.e., \( t_i = X p^T, i = 1, 2, \cdots, p \)). E is the residual matrix. s can be determined by

\[ \sum_{j=1}^{s} \sigma_j^2 \times 100\% > 95\%, \quad \text{.............. (2)} \]

where \( \sigma_j \) is the jth singular value of X. Inequality (2) indicates that more than 95% of the total variance of the data is covered by the first s principal loadings. 31]

Hotelling’s T^2 statistic and SPE statistic are two statistics, which have been widely used in the PCA based online process monitoring methods. They are defined, respectively, as

\[ T^2 = x^T \Sigma_x^{-1} x \quad \text{.............. (3)} \]

and

\[ SPE_i = r^T_i r_i, \quad \text{with} \quad r_i = x_i - \hat{x}_i, \hat{x}_i = P_i p_i x_i, \quad \text{...... (4)} \]

where \( x_i \) denotes a new-coming observation at sampling instant \( k \), \( P_i \) is a matrix composed of the first s columns of \( P \) (i.e. \( p_i \) for \( 1 \leq i \leq s) \), \( \Sigma_s \) is a diagonal matrix composed of the s largest singular values of \( P_i \).

Given a confidence level \( \alpha \), the thresholds for monitoring \( T^2 \) and SPE statistics can be calculated as below: 25, 31)

\[ \begin{align*}
T^2_{\alpha} &= \frac{s(L-1)\alpha(L+1)}{L(L-s)} F_{\alpha}(s, L-s) \quad \text{.............. (5)} \\
SPE_{\alpha} &= \theta_1 \left[ c_{i}(2\theta_2 h_0)_{\alpha}^{1/2} + 1 + h_0(h_0 - 1)^{-1/2} \right]^{1/2}, \quad \text{.............. (6)}
\end{align*} \]

where \( F_{\alpha}(s, L-s) \) is the critical value of the F-distribution with \( s \) and \( L-s \) degrees of freedom for the \( \alpha \) significance level, \( c_{i} \) is the normal deviate corresponding to the upper \((1-\alpha)\) percentile, \( \theta_i = \sum_{j=1}^{a} \sigma_j^2 \) for \( i = 1, 2,3 \) and \( h_0 = 1 - 20 \theta_1 / (3\theta_1) \).

Note that when the SPE statistic exceeds its threshold, the contribution plot 25) which reflects the contribution of each variable to deviations in the SPE statistic within a given sampling interval, can be used to diagnose which variables may be the cause of the faults. The contribution of the \( i^{th} \) variable to the SPE statistic at instant \( t \) is defined as:

\[ c_i(t) = (x_i(t) - \hat{x}_i(t))^2 = r_i(t)^2, \quad i = 1, \cdots, p. \quad \text{.............. (7)} \]

The contribution of the \( i^{th} \) variable to the SPE statistic between the instants \( k_1 \) and \( k_2 \) can be further calculated as

\[ C_i = \sum_{t=k_1}^{k_2} c_i(t). \quad \text{............... (8)} \]

3. Methodology and Case Study

3.1. Description of the Data Set

In this paper, part of the historical data collected from No. 3 Blast Furnace with a volume of 2 000 m³ in Liazhou Iron&Steel Co. Ltd. is used as the data set. The dataset includes 103 680 samples (corresponding to 12 days’ data with the sampling time of 10 seconds) for 29 main monitored variables, which are now listed in Table 1. An overview of the data is given in Fig. 1. It can be seen that the data set is divided into two parts, i.e. Part 1 and Part 2. Part 1 with 40 000 samples were collected under normal condition and will be used as the training set to build the PCA based process-in-control model. Part 2 with 63 680 samples will be used as the test set. According to the event log provided by the blast furnace operators, the data collected from the 85 553rd sampling instant to the 103 680th sampling instant (i.e. the part between the two adjacent red dotted lines in Fig. 1) corresponds to the process of a getting cold accident of the blast furnace. The remaining samples in Part 2 were collected under normal condition.

Remark 1: It is worth mentioning that all the 29 variables are plotted in Fig. 1 with their original amplitudes. Thus some variables appear overlapped at the bottom of Fig. 1, because their amplitudes are relatively small with comparison to the red, black, blue, pink and green ones, which correspond to blast furnace bosh gas volume, enriching oxygen flow, cold blast flow, theoretical combustion temperature
and hot blast temperature, respectively. For the same reason, only the five variables are marked in the legend of Fig. 1. The units of the rest variables can be found in Table 1. Nevertheless, the data will be mean centered and scaled before they are used in the modeling and monitoring procedures. More details will be given in Section 3.2.

Different from the experimental blast furnace mentioned in literature, there are four stoves blowing the hot air into the blast furnace in turn in Liuzhou Iron&Steel Co., Ltd. Each switching between two distinct stoves would cause a peak-like disturbance as shown in the measured signals in Fig. 2, which enlarges the data of Part 1 in Fig. 1 from sampling instants \( k = 52500 \) to \( k = 55000 \). Since the switching time has not been recorded in the event log and there is no signal in the historical database can directly indicate the switching operation, the starting and ending time of the aforementioned disturbances are unknown to us. These switching invoked disturbances may severely influence the performance of process monitoring.

3.2. Two-stage PCA Based Process Monitoring for Ironmaking
3.2.1. Basic Idea

To overcome the effects of the aforementioned peak-like disturbances due to stove switching, a two-stage PCA based method is presented in this paper. In the first stage, those strong disturbances are identified based on PCA and a new training dataset is constructed by removing the identified switching disturbances. In the second stage, the standard PCA based process monitoring method is utilized. More concretely, a PCA model for the normal ironmaking process of the blast furnace is firstly built based on the training dataset constructed in the preceding stage. Then the \( T^2 \) and \( SPE \) statistics for each newly arrived sample in the test dataset are calculated and monitored to simulate a real online monitoring procedure. It is worth noting that the presented method in this paper is different from the standard PCA based method on the decision rule for generating fault alarms as will be discussed in Remark 4.

3.2.2. Stage 1: PCA Based Method for Removing the Data Corresponding to Stove Switching Operations

As mentioned above, a PCA based algorithm is proposed in the first stage to remove the peak-like disturbances from the training dataset which are caused by stove switching operations. The details, are summarized as follows.

**Algorithm 1:**

- **Step 1:** Pre-process the raw training data set \( \tilde{X}_{tr,ori} \) to obtain the mean centered and scaled measurements \( X_{tr,ori} \).
- **Step 2:** Let \( X = X_{tr,ori} \), build the PCA model for \( X_{tr,ori} \) according to (1) and select the number of principal components, \( s \) according to (2).
- **Step 3:** For each sample \( x_j \) in \( X_{tr,ori} \), where \( j \) denotes the sampling instant, derive the \( T^2 \) statistic, \( T^2_j \), from (3).
- **Step 4:** For each instant \( k \), calculate the mean value \( \mu_k \) and standard deviation \( \sigma_k \) of \( T^2_j \) for \( k \leq j \leq k + 50 \).
- **Step 5:** If \( T^2_k < \mu_k + 3\sigma_k \), regard \( x_k \) as a normal sample; otherwise, identify \( x_k \) as a sample related to a stove switching operation. The starting and ending time of the switching operation are determined as follows. Find the nearest instants \( k - m \) and \( k + n \) from \( k \) such that \( T^2_{k-m} < \mu_k + \sigma_k \) and \( T^2_{k+n} < \mu_k + \sigma_k \) are satisfied, respectively, where \( m \) and \( n \) are both positive constants and \( m < k \). The data sampled between instants \( k - m \) and \( k + n \) are regarded as peak-like disturbances.
- **Step 6:** Construct a new training data set \( (X_{tr,new}) \) by first removing the data corresponding to the disturbances identified in Step 5, and then re-processing the data to obtain the mean centered and scaled measurements.

**Remark 2:** Note that \( \mu_k + 3\sigma_k \), which is adopted at the beginning of Step 4 in Algorithm 1, is normally selected as the threshold to judge whether the \( T^2 \) statistic has an obvious change in the sense of hypothesis-test. In contrast to this, \( \mu_k + \sigma_k \) is used for locating the starting and ending points of the peak. This indicates that when \( T^2_k \) statistic falls below \( \mu_k + \sigma_k \), the samples would be treated to be normal. The effectiveness of such selections will be demonstrated by the

<table>
<thead>
<tr>
<th>Variable list of the data set.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Oxygen enrichment rate (%)</td>
</tr>
<tr>
<td>2. Blast furnace permeability index</td>
</tr>
<tr>
<td>3. CO volume (%)</td>
</tr>
<tr>
<td>4. H2 volume (%)</td>
</tr>
<tr>
<td>5. CO2 volume (%)</td>
</tr>
<tr>
<td>6. Blast velocity at tuyere of blast furnace (m/s)</td>
</tr>
<tr>
<td>7. Enriching oxygen flow (m³/h)</td>
</tr>
<tr>
<td>8. Cold blast flow (10⁴ m³/h)</td>
</tr>
<tr>
<td>9. Blast momentum (kJ)</td>
</tr>
<tr>
<td>10. Blast furnace bosh gas volume (m³)</td>
</tr>
<tr>
<td>11. Theoretical combustion temperature (°C)</td>
</tr>
<tr>
<td>12. Blast furnace top gas pressure (1) (°C)</td>
</tr>
<tr>
<td>13. Blast furnace top gas pressure (2) (°C)</td>
</tr>
<tr>
<td>14. Blast furnace top gas pressure (3) (°C)</td>
</tr>
<tr>
<td>15. Blast furnace top gas pressure (4) (°C)</td>
</tr>
<tr>
<td>16. Actual blast velocity (m/s)</td>
</tr>
<tr>
<td>17. Cold blast pressure (MPa)</td>
</tr>
<tr>
<td>18. Total pressure drop (Kpa)</td>
</tr>
<tr>
<td>19. Hot blast pressure (MPa)</td>
</tr>
<tr>
<td>20. Actual blast pressure (MPa)</td>
</tr>
<tr>
<td>21. Hot blast temperature (°C)</td>
</tr>
<tr>
<td>22. Top temperature (1) (°C)</td>
</tr>
<tr>
<td>23. Top temperature (2) (°C)</td>
</tr>
<tr>
<td>24. Top temperature (3) (°C)</td>
</tr>
<tr>
<td>25. Top temperature (4) (°C)</td>
</tr>
<tr>
<td>26. Drag coefficient</td>
</tr>
<tr>
<td>27. Coal injection set value (T/h)</td>
</tr>
<tr>
<td>28. Actual coal injection rate (T/h)</td>
</tr>
<tr>
<td>29. Actual coal injection in last hour (T)</td>
</tr>
</tbody>
</table>
experimental results given in Section 3.3.

Remark 3: In processing the historical data, we find that the stove switching operation would cause large fluctuations in $T^2$ statistic. This implies that a switching may change the relationships of the variables in a process and it can be described by $T^2$ statistic. However, it is hardly to be described by a single variable. Therefore, although some variables including blast velocity at tuyere and hot blast pressure are strongly related to the stove switching operations, $T^2$ statistic is adopted in Algorithm 1 of this paper rather than a single variable to identify the peak-like disturbances invoked by stove switching operations. Moreover, it will be observed from the results of historical data processing presented in Section 3.3 that the $SPE$ statistic is not as sensitive as the $T^2$ statistic to the switching disturbances.

3.2.3. Stage 2: PCA Based Status Abnormality Monitoring

We now use the standard PCA based method presented in Section 2 to monitor the occurrences of abnormality during the ironmaking process. The following two procedures will be involved: (i) the PCA based modeling; (ii) the $T^2$ and $SPE$ statistics based monitoring.

Note that the modeling procedure is similar to Step 2 of Algorithm 1, i.e. let $X = X_{tr\_new}$, build the PCA model for $X_{tr\_new}$ according to (1), and select the number of principal components $s$ according to (2).

In the following, the $T^2$ and $SPE$ statistics are calculated and monitored for each sample in the test dataset to sim-
ulate the real online monitoring process for each newly arrived sample.

**Algorithm 2:**

- **Step 1:** Based on $X_{tr, new}$, set the threshold, i.e., the upper control limit (UCL), $T_{UCL}^2$ for the $T^2$ statistic and the threshold $SPE_{UCL}$ for the $SPE$ statistic according to (5) and (6), respectively. Set another threshold $W$ from the distribution of the width of the identified peak-like switching disturbances obtained by performing Algorithm 1.
- **Step 2:** Recalling that there are 40,000 samples in the original training set $X_{tr, ori}$. From (3) and (4), calculate $T_k^2$ and $SPE_i$ for each sample $x_k$ for $k = 40,000 + 1, ..., 40,000 + W$ (i.e., the first $W$ samples in the test dataset $X_{test}$).
- **Step 3:** Calculate $T_k^2$ and $SPE_i$ for each newly arrived sample $x_k$, for $k = 40,000 + W + 1, ..., 103,680$. Then make a decision according to the following rules.
  (i) If $T_k^2 ≥ T_{UCL}^2$ and $SPE_i ≥ SPE_{UCL}$ hold simultaneously, for $i = k - W + 1, k - W + 2, ..., k$, it is concluded that the blast furnace has been operated under an abnormal condition.
  (ii) If $T_k^2 ≥ T_{UCL}^2$ and $SPE_i ≥ SPE_{UCL}$ hold simultaneously only for the time instants $k$ smaller than $W$, it is concluded that the abnormal $T^2$ and $SPE$ statistics are caused by the peak-like switching disturbances.
  (iii) Otherwise, it is concluded that the ironmaking process has been operated under normal condition.
- **Step 4:** If the blast furnace is judged to be operated under an abnormal condition, i.e., the conditions in Step 3 (i) are satisfied, a fault alarm is generated and the contributions for $SPE$ statistic are further calculated from (8).

**Remark 4:** In most of the existing PCA based monitoring algorithms, the process is judged as in abnormal condition regarded that $T^2$ and/or $SPE$ statistics exceed their respective thresholds. In contrast to this, Step 3 (i) of Algorithm 2 in this paper indicates that only if both the $T^2$ and $SPE$ statistics have exceeded their thresholds for a period longer than $W$, it is concluded that an abnormal status has occurred in the blast furnace. This will inevitably result in a detection delay of $W$ sampling instants. However, such delay may last only several minutes, thus it is negligible with comparison to the settling times of an ironmaking process which are normally several hours. Moreover, $W$ in Algorithm 2 can be selected according to the distribution of the widths of the identified switching disturbances by Algorithm 1. In the experiments of this paper, it is set as 0.9 $W_m$ where $W_m$ is the maximal width of the identified disturbances.

### 3.3. Experimental Results

#### 3.3.1. Results of Peak-like Disturbances Removing

In this part, the experimental results of the first stage of our PCA based process monitoring method are provided. By performing Algorithm 1 on the original training dataset $X_{tr, ori}$, a new training dataset $X_{tr, new}$ is obtained with the peak-like disturbances removed.

**Figure 3** shows the percentages of the variance, which are accounted for by 29 principal components. It can be seen that more than 95% of the variance in the data set are captured by the first 12 principal components. Thus the number of principal components, $s$, is set to 12. In other words, only the first 12 principal components are adopted to build the PCA model in Step 2 of Stage 1.

We select the confidence level $\alpha$ as $\alpha = 1 \times 10^{-5}$. According to (5) and (6), $T_{UCL}^2$ and $SPE_{UCL}$ are set as $T_{UCL}^2 = 45.1095$ and $SPE_{UCL} = 6.9504$.

The $T^2$ statistic of $X_{tr, ori}$ obtained by performing Step 3 of Algorithm 1 is given in **Fig. 4** where the peak-like disturbances identified by Algorithm 1 are also indicated.

To verify that the $T^2$ statistic outperforms the $SPE$ statistic in identifying the switching disturbances, we also calculate $SPE_i$ for each sample in $X_{tr, ori}$ as shown in **Fig. 5**. Clearly, the $SPE$ statistic is not as sensitive as the $T^2$ statistic to the switching disturbances.

#### 3.3.2. Results of PCA Based Status Abnormality Monitoring

In this part, a PCA based model is firstly rebuilt based on the new training dataset $X_{tr, new}$ obtained in Stage 1. Then Algorithm 2 is performed on the test data $X_{test}$ to simulate the real online monitoring process for each newly arrived sample.

By following similar procedures of selecting $s$ in Stage 1, $s$ is also set to 12 in the second stage of PCA modeling. And the thresholds $T_{UCL}^2$ and $SPE_{UCL}$ for process monitoring are set to 45.1109 and 6.2149 for the confidence level $\alpha = 1 \times 10^{-5}$ according to (5) and (6), respectively.

The distribution of the widths of the switching disturbances identified by Algorithm 1 is plotted in **Fig. 6**, where “frequency” in vertical axis is specifically the number of times that the widths fall within a given interval. It can be seen that the maximal width $W_m$ is 40 instants. Then the threshold $W$ is set as 0.9$W_m$, i.e., 36 instants which correspond to 6 minutes. Therefore, in Step 3 of Algorithm 2, if both the two statistics have exceeded their respective thresholds for a period longer than 36 instants, an alarm for abnormality will be generated. If both the statistics exceeds the thresholds, whereas the lasting period is shorter than 36 instants, the samples will be treated as disturbances caused by stove switching.

An overview of the $T^2$ statistic and $SPE$ statistic for the
test set $X_{\text{test}}$ by performing Step 2 and Step 3 of Algorithm 2 in this paper is provided in Figs. 7(b) and 7(c). The blue, green and red parts of the line are judged as normal status, disturbances and abnormal status, respectively. The grey dotted line represents the thresholds. Besides, the yellow dotted vertical line indicates the starting time of the accident reported in the event log, while the red dotted vertical line indicates the time of fault alarm generated by Algorithm 2. It can be seen that with the monitoring method presented in this paper, the alarm of the getting cold accident can be generated in advance of the time reported in the event log. Furthermore, there is no false alarm beforehand. Note that the red part of the line near time instant 70 000 alarmed by our method corresponds to a scheduled blast furnace blowing down.

By enlarging part of Fig. 7 from $k = 84 600$ to $k = 85 600$ in Fig. 8 the advantages of our two-stage PCA monitoring method in generating the accident alarm become more apparent. From Figs. 8(a), 8(b) and 8(c), it can be seen that the alarm is generated 396 instants (approximately 1.1 hours) earlier than the instant recorded in the event log. In fact, most of the serious faults are evolved gradually from tiny faults, which leads to the gradual increase of the $T^2$ and $SPE$ statistic.
SPE statistics. This constitutes the main reason that PCA based monitoring method can detect a fault in its incipient stage.

Observing Figs. 8(d) and 8(e), we find the $T^2$ statistic of the direct PCA method generates the fault alarm at almost the same time as the two-stage PCA based method. However, its SPE statistic fails to generate an alarm earlier than the time recorded in the event log. As mentioned in Remark 4, different from our presented monitoring algorithm, the operational status of the process is normally judged according to $T^2$ and/or SPE statistics in direct PCA method. Thus the advantages of the two-stage PCA can be concluded based on Fig. 8 from the following two aspects:

1) If $T^2$ statistic is adopted in direct PCA to monitor the process, it can be seen that the direct PCA generates the alarm 3 instants (30 seconds) earlier than the two-stage PCA method. However, such negligible superiority in alarm generation time, is gained at a cost of higher false alarm rate. More concretely, direct PCA method leads to false alarms of 106 instants (17.7 minutes), whereas no false alarm is generated by the two-stage PCA method.

2) If SPE statistic is employed in direct PCA to avoid false alarms, the two-stage PCA method alarms at the 85 156th instant, and the direct PCA method alarms at the 85 553rd instant. Thus the two-stage PCA method alarms 397 instants (approximately 1.1 hours) earlier than the direct PCA method.

As mentioned in Section 2, once an alarm is generated, contribution plot can be further used to analyze which variables contributes the most to the change of the SPE statistic. The contributions of the 29 main variables to the SPE statistic given by our two-stage PCA model and the direct PCA model are plotted in Figs. 9(a) and 9(b), respectively. We focus on the time interval from instant 85 156 to 85 300 due to the following reasons. 1) During this period, the fault had developed gradually. However, it has not become so severe that the PCA model needs to be adjusted. 2) There is no stove switching procedure during this period.

### Fig. 8. Comparison of the alarm time by different monitoring methods.

![Comparison of the alarm time by different monitoring methods.](image)

In Figs. 9(a) and 9(b), the contribution plots of the 2-stage PCA and direct PCA methods are given, respectively. In (b), the 12th variable (corresponding to theoretical combustion temperature) contributes the most to the deviation of the SPE statistic. However in (a), the 6th and 8th variable (corresponding to blast velocity at tuyere of blast furnace and cold blast flow) contribute more than theoretical combustion temperature.

The main reason to such difference of the contribution plots is that in contrast to direct PCA, switching disturbances have been removed in the training stage of the two-stage PCA. Since the blast velocities at tuyere of blast furnace and cold blast flow are hidden in the switching disturbances, the SPE statistic of direct PCA is much less sensitive to the two variables than that of two-stage PCA.

**Remark 5:** Note that the contribution plot of the two-stage PCA method is consistent with the operators’ empirical judgments and the analysis in the accident report provided by Liuzhou Iron&Steel Company.
4. Conclusion

This paper focuses on process monitoring and fault diagnosis for the blast furnace ironmaking process. A two-stage PCA method is used to monitor the process. In the first stage, the peak-like disturbances caused by the stove switchings are removed. In the second stage, the monitoring model is first rebuilt based on dataset with the switching disturbances removed. Then the process is monitored based on the $T^2$ statistic and the SPE statistic with a modified decision logic. Experiment results based on the historical data collected from a real blast furnace with a volume of 2 000 m$^3$ validate the advantages of our two-stage PCA method in detecting the getting cold accident with comparison to the direct PCA method and operators’ monitoring.

Acknowledgements

The work was supported by National Natural Science Foundation of China under grant 61290324 and 61203068.

REFERENCES

22) D. Noskivečová: Metalurgija, 48 (2009), 281.