PCA-SDG Based Process Monitoring and Fault Diagnosis: Application to an Industrial Pyrolysis Furnace

Xianyao Han\textsuperscript{a}, Shengwei Tian\textsuperscript{a}, Jose A. Romagnoli\textsuperscript{b}, Hui Li\textsuperscript{c}, Wei Sun\textsuperscript{a}\textsuperscript{*}

\textsuperscript{a}Beijing Key Lab of Membrane Science and Technology, College of Chemical Engineering, Beijing University of Chemical Technology, 100029 Beijing, China

\textsuperscript{b}Cain Department of Chemical Engineering, Louisiana State University, Baton Rouge, LA 70808, United States

\textsuperscript{c}Shanghai SupeZET Engineering Technology Co., Ltd., 200335, Shanghai, China

Corresponding Author: Wei Sun, sunwei@mail.buct.edu.cn

\textbf{Abstract}: Ethylene cracking furnace is a key unit in the ethylene production process, whose operating condition is subject to changes with fluctuation in feed condition, the aging of equipment, and other possible disturbances. To promptly and correctly identify the root causes of process changes, an online process monitoring system based on Principal Component Analysis (PCA) and Signed Directed Graph (SDG) method is proposed for multiple operational conditions monitoring and fault detection. Active process adjustments or passive fluctuations are first differentiated, then the root cause is isolated by SDG reasoning based on the contribution percentage of principal variables in PCA. The True Positive Rate (TPR) of fault detection is 98\%, and False Alarm Rate (FAR) is 1.56\%. The root causes of fault match very well with operation records.

\textbf{Keywords}: multiple operational condition, active process adjustment identification, statistical monitoring

1. INTRODUCTION

Ethylene cracking furnace is a key unit in ethylene production process, which is expected to maintain a long-term steady operation (Wang and He, 2000). During the operation of a furnace, process variables are adjusted by operators quite often due to the changes of feedstock, production planning and equipment aging, which leads to the frequent switches of operating conditions. It is hard to find the exact same situation within the available historic record, which makes it quite challenging for data-based monitoring methods, i.e., there are no sufficient historical data for developing models to monitor the current operation. On the other hand, once a significant deviation is detected, the root cause needs to be promptly isolated and operators can make adjustment accordingly to bring the operation back to expected state.

It can be seen that, for the process monitoring of ethylene cracking operations, works in following three aspects need to be conducted: how to correctly identify active process adjustment or passive fluctuation, how to realize the real-time multiple operational conditions monitoring, and how to isolate the root cause of the unexpected process deviation under different operational conditions.

In ethylene cracking process, manipulated variables include total naphtha mass flow rate, dilution steam mass feed rate, and feed ratios among six coils. Any active process adjustment will be associated with the changes of these manipulated variables.

Most process deviation can be identified efficiently by Multivariate Statistical Monitoring approaches given enough historical data under one certain process operation condition (Cinar et al., 2007). Among them, PCA is an efficient method, especially in dimension reduction and latent feature extraction of high-dimension data (Venkatasubramanian et al., 2003). To monitor an industrial process with multiple operating conditions, some approaches based on PCA method have been proposed, such as multiple PCA modes (Zhao et al., 2004), moving window PCA (Wang et al., 2005) and recursive PCA (Li et al., 2000). However, for multiple PCA modes, a high resolution can be given to known modes but its application is limited by available operational condition data. For recursive PCA and moving window PCA, they fail to distinguish the active process adjustment from process deviation, and tend to neglect process drift if it is monotone in one direction.

Furthermore, fault diagnosis is concerned in determining the root cause of the observed out-of-control status (Chiang et al., 2000). PCA approach can be used for fault diagnosis by analysing variable contributions. However, maximal contribution variable may not be the actual root cause of fault. SDG is a knowledge-based fault diagnosis method, which can efficiently determine the fault root cause. PCA-SDG-based fault diagnosis algorithm was proposed by Vedam and Venkatasubramanian (1999), which combined advantages of both methods, and has shown good performance in the Amoco Model IV Fluidized Catalytic Cracking Unit. In addition, for multiple chemical operating
units system, fault diagnosis can be realized by using causal map and multivariate statistics (Chiang et al., 2015).

In this work, historical data of an industrial ethylene cracker are analysed. An online process monitoring system based on PCA method is developed for multiple operational condition monitoring and fault detection, in which active process adjustment or passive fluctuation of operation process are first differentiated by a condition recognizer, then the root cause is isolated by PCA based SDG reasoning.

2. METHODOLOGY

2.1 Principal Component Analysis (PCA)

PCA is a standard multivariate method proposed by Pearson (1901), which has a good performance on feature extraction and dimension reduction. Considering data matrix \(X \in \mathbb{R}^{n \times m}\), a normalized data matrix of \(n\) samples and \(m\) variables, the covariance matrix of \(X\) is defined as follows,

\[
\text{Cov}(X) = \frac{X^T X}{n-1}
\]

\(X\) can be decomposed as follows into a score matrix \(T\) and a loading matrix \(P\) whose columns are the right singular vectors of \(X\) plus a residual matrix \(E\). \(k\) is the number of the principal components

\[
X = TP^T + E = t_1p_1^T + t_2p_2^T + \ldots + t_kp_k^T + E
\]

2.2 Similarity analysis among different PCA models

According to linear algebra, the eigenvector of a matrix can be found as following:

\[
Ax = \lambda x
\]

If \(A\) is the loading matrix \(P\) in Eq. (2), then there will exist a \(\lambda\) for the \(P\), which gives an equivalent transformation result of \(P\) regarding variable vector \(x\). Thus, the similarity analysis among different PCA models can be conducted by calculating the correlation coefficients among \(\hat{\lambda}\)s of loading matrices.

2.3 Fault detection based on PCA method

PCA based fault detection method mainly utilizes \(T^2\) statistic, SPE statistic, and their control limits, which are defined as follows,

\[
T^2 = s^T S^{-1} s
\]

\[
T^2_a = \frac{k(n-1)(n+1)}{n(n-k)} F_a (k, n - k)
\]

\[
SPE = e e^T = \sum_{i=1}^{m} (x_i - \hat{x}_i)^2
\]

\[
SPE_a = \theta_1 \left[ \frac{\hat{\alpha} \sqrt{2m}}{\theta_1} + 1 + \frac{\alpha h_a - 1}{\theta_1^2} \right]^{1/h_a}
\]

where \(t\) is the score vector, \(S\) is a diagonal matrix composed by the eigenvalues of matrix \(X\), \(\hat{\alpha}\) is an estimated value reconstructed by score vector and loading vector, and \(e\) is the deviation of real time data matrix \(X\) and reconstructed matrix \(\hat{\alpha}\), and \(\alpha\) is the significance level. \(\theta_1 = \sum_{j=a+1}^{m} a^2\), \(h_0 = 1 - \frac{2\theta_1}{2\theta_1^2}\), and \(c_\alpha\) is the normal deviate corresponding to the \((1 - \alpha)\) percentile.

2.4 PCA-SDG fault diagnosis

SDG is a knowledge-based fault diagnosis method, which represents the causal relationships between variables by graphs. SDG model is a network diagram which consists of nodes and directed arcs between nodes. The nodes can express system physical variables, control variable or instrument, or an event. A simple example of a SDG model is shown in Fig. 1, where A is the cause node, and B is the effect node. The algorithm tries to find a single consistent path from the root node to all the abnormal measured nodes.

![Fig. 1. Diagram of SDG](image-url)

In SDG, single-variable statistic method is used to determine the node thresholds, where relationships among variables haven’t been taken into account. \(2m\) thresholds must be determined, where \(m\) represents the number of procedure variable. PCA is an efficient method in multivariate statistic, by which correlations among variables are well considered. By combining PCA and SDG methods, the node thresholds in SDG can be determined by SPE statistic \(SPE_a\) based on PCA, which is computed by Eq. (6). There is only one value \(SPE_a\) that needs to be computed to determine all nodes thresholds, which can greatly reduce the amount of involved calculations. Generally, the high threshold of node in SDG model is selected in the range of \(\sqrt{SPE_a/m} \sim \sqrt{SPE_a}\), the low threshold is in \(-\sqrt{SPE_a/m} \sim -\sqrt{SPE_a}\). In this paper, the high and low thresholds of node are set to \(\sqrt{SPE_a/m}\) and \(-\sqrt{SPE_a/m}\) respectively. When a fault is detected based on \(T^2\) statistic by Eq. (4) and Eq. (5), the residual contributions are generated. Nodes in the SDG import values of (0), (+) and (-) representing the variables are within threshold values, higher and lower than threshold values, respectively. SDG is then used to find root cause by finding the consistent path. The PCA-SDG procedure is shown in Fig. 2.
3. FAULT DETECTION AND FAULT DIAGNOSIS

As mentioned before, the first step of this work is to correctly identify active process adjustment or passive fluctuation, followed by real-time multiple operational condition monitoring, and root cause isolation as shown in Fig. 3.

3.1 Condition recognizer based on manipulated variables

As discussed above, active process adjustment is always associated with manipulate variable changes, which gives a nudging point for operation condition recognition.

For the total naphtha mass flow rate and each coil dilution steam mass feed rate, wavelet filters are first used to eliminate the effects of outliers, and then two moving windows are used for the condition recognition. The long moving window (300 samples at 1 minute sampling frequency) is used to calculate the mean of the state. The short moving window (10 samples) is used to calculate the current mean. The two moving windows are separated by a certain distance in order to improve the sensitivity. Condition changes occur when the average of the small window exceeds ±3 times standard deviation of the large window and holds 10 sampling points. Then, a trend recognizer is used to identify the new operational condition. The procedure is shown in Fig. 4.

For the feed ratio among six coils, Euclidean distances are used to determine process operating condition. The stable time of adjusting naphtha mass flow rate needs 30 minutes, so the Euclidean distance is composed by parameters between current time and the last 30 minutes. It’s considered as a new stable condition when Euclidean distance exceeds a value which is selected by historical data and continues over 5 samples.

New process condition is determined when all manipulated variables are under new stable condition.
3.2 Online monitoring system based on Principal Component Analysis (PCA) method

Four different condition data sets with different data mean and variance are selected from two different production cycles to analyse the similarity of PCA projections. Four PCA models are established by their respective normalization centers. For PCA model, data features are extracted in the loading matrix \( P \), which can be inferred from its eigenvalue \( \lambda \) based on Eq. (3).

The results are shown in Table 1. Over 96.55% similarity can be found among \( \lambda_s \) from different data sets, which shows that the projection orientations of different data sets are very close.

Thus, the same PCA loading matrix can be used in different operational conditions by adjusting the normalization center. In online process monitoring, condition recognizer is used first to recognize current data condition and identify normalization center.

**Table 1. The correlation coefficients in loading matrix \( P \) among different conditions’ data**

<table>
<thead>
<tr>
<th></th>
<th>Condition1</th>
<th>Condition2</th>
<th>Condition3</th>
<th>Condition4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Condition1</td>
<td>1</td>
<td>0.9955</td>
<td>0.9686</td>
<td>0.9655</td>
</tr>
<tr>
<td>Condition2</td>
<td></td>
<td>1</td>
<td>0.9964</td>
<td>0.9695</td>
</tr>
<tr>
<td>Condition3</td>
<td></td>
<td></td>
<td>1</td>
<td>0.9703</td>
</tr>
<tr>
<td>Condition4</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>

3.3 SDG model and variable information

In this work, an industrial process of naphtha steam cracking is analysed. In pyrolysis furnace, the mixture of naphtha and steam are preheated in the convection section and take place cracking reaction in the radiation section. Then the cracking reaction is terminated by rapid cooling in heat exchanging section. The energy of the reaction is provided by the fuel gas, which is burned in multiple nozzles. In cracking process, COT determines the component distribution in final product. By naphtha, diluted steam and fuel gas feed adjustment, total COT, which is calculated by the mean of 6 coils (4 tubes in each coil), can be controlled to a steady value. The SDG diagram of this process is shown in Figure 5. Variable information for process monitoring and fault diagnosis is shown in Table 2.

**Fig. 5. SDG model for ethylene cracking operation**

<table>
<thead>
<tr>
<th>Variables</th>
<th>Description</th>
<th>Quantity</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1</td>
<td>Naphtha mass flow rate</td>
<td>7</td>
</tr>
<tr>
<td>T1</td>
<td>Naphtha temperature</td>
<td>1</td>
</tr>
<tr>
<td>P1</td>
<td>Naphtha pressure</td>
<td>1</td>
</tr>
<tr>
<td>F2</td>
<td>Diluted steam mass flow rate</td>
<td>7</td>
</tr>
<tr>
<td>T2</td>
<td>Diluted steam temperature</td>
<td>1</td>
</tr>
<tr>
<td>T3</td>
<td>Crossover section temperature</td>
<td>6</td>
</tr>
<tr>
<td>P2</td>
<td>Crossover section pressure</td>
<td>6</td>
</tr>
<tr>
<td>Ta</td>
<td>Temperature in furnace A side</td>
<td>1</td>
</tr>
<tr>
<td>Tb</td>
<td>Temperature in furnace B side</td>
<td>1</td>
</tr>
<tr>
<td>Ql</td>
<td>Total fuel gas calorific value</td>
<td>1</td>
</tr>
<tr>
<td>Cot</td>
<td>Coil outlet temperature</td>
<td>25</td>
</tr>
<tr>
<td>P3</td>
<td>Outlet pressure</td>
<td>6</td>
</tr>
</tbody>
</table>

**Table 2. Variables information**

4. APPLICATION RESULTS

Data for a complete production cycle, from November 6, 2015 to February 6, 2016 with one-minute interval, are selected to test the applicability of the proposed fault detection and diagnosis strategy.

4.1 Fault detection results

For a known load adjustment due to the downstream reactor switching, the condition recognizer results, total naphtha mass flow rate, and naphtha mass flow rate among six coils are shown in Fig. 6. In Fig. 6(a), straight line indicates total naphtha mass flow rate, dash dot line indicates the mean of subsequent monitoring model, and pentagonal mark indicates the condition recognizer results. When the condition recognizer confirms that the process parameters are adjusted by operators, the value is assigned to 1. Otherwise, the value is assigned to 0. In Fig. 6, process experienced operator adjustments during 44311th to 44387th and 44510th to 44604th sample. The condition recognizer confirmed that process has reached a new steady condition on the 44409th sample and then the mean of monitoring model was updated. Based on condition recognizer results, real-time monitoring can be realized by updating model parameters.

**Fig. 6. Condition recognizer results**
The monitoring result for the complete production cycle (93 days) is shown in Fig. 7. As known, cracking process cannot be kept identical for a long-term. In the process initial stage, operators make parameters adjustments in order to optimize operational condition. In later stage, parameters are adjusted due to the non-uniform coking condition of coils. It can be seen that the process monitoring results are in consistence with the actual process.

In the industrial process, the coil outlet temperature (COT) represents the cracking reaction temperatures, which are extremely important control variables. Operators need to make proper parameters adjustments once coil outlet temperatures (COT) is over ±3°C range of setting value. Samples which are over COT permissible range and last more than 2 minutes are considered as faults. Red samples in plot show the samples which are out of the COT permissible range.

![Fig. 7. Monitoring results](image)

By proposed monitoring system, the TPR of these faults is 98% with alarm in advance and FAR is 1.56%. As shown, it can alarm 12 minutes early in Fig. 8(a) and 3 minutes early in Fig. 8(b).

![Fig. 8. Detail monitoring results](image)

Parameters adjustments occurred in Fig. 9 due to process deviation. ‘x’ mark samples in plot show the samples which are under parameters adjustment process by operator and not yet reach new steady condition. Process reach a new steady operational condition after 57239th sample. In Fig. 9(a), the condition recognizer is not used and continuous false warnings are triggered. In Fig. 9 (b), when new operational condition is identified by condition recognizer, the model parameters are updated to realize new operational condition process monitoring. As shown, monitoring system can promptly adapt new steady operational condition.

![Fig. 9. Monitoring results without (a) and with (b) condition recognizer](image)

It can be seen that the monitoring system can efficiently detect the abnormal working condition in advance, and realize online monitoring under multiple operational conditions.

### 4.2 Fault diagnosis results based on PCA-SDG

In this monitoring system, deviations, which are over $T^2$ statistic value and last more than 3 minutes, are considered as fault here. Thus, following root cause is isolated by measured variable contributions and SDG reasoning based on the fourth fault sample in PCA.

The fault detected in Fig. 8(a) is analysed. As shown, the COT is over permissible range on the 35581th sample in red. Faults are detected since the 35569th samples. Contribution plot and fault diagnosis results based on the 35573th are shown in Fig. 10. According to Fig. 10(a), naphtha mass flow rate in coil 5 (number 6) and outlet pressure in coil 2 (number 53) are under high contribution rate. 16 variables exceed the 1/63 contribution rate. Apparently, some of them cannot be the root cause. Their residuals are recorded and input into SDG network diagram (Fig. 5). Fault diagnosis result based on SDG indicates that COT is low in the second tube of coil 1 due to high naphtha mass flow rate in coil 1. COT is high in the second tube of coil 6 due to high total fuel gas calorific value. Crossover pressure is high because of high naphtha mass flow rate in coil 6.

By combining contribution plot and SDG analysis, naphtha mass flow rate fluctuated in coil 1, coil 5, and coil 6 while total naphtha mass flow rate didn’t. Although the total fuel gas feed reflects COT in the second tube of coil 6, the contribution rate of this fault path is not the most prominent. Thus, this fault is mainly caused by naphtha feed ratios deviation. In addition, in the SDG network there is no fault output path about outlet pressure in coil 2, which is under high contribution rate. Therefore, deviation of outlet pressure in coil 2 is caused by coke layer peeling.

![Fig. 10. Fault diagnosis results based on PCA-SDG](image)
Fig. 10. Contribution plot in (a) and fault diagnosis result in (b).

Furthermore, the monitoring result for a period of process deviation is shown in Fig. 11. Faults are detected since the 65578th sample. Contribution plot and fault diagnosis result based on 65582th sample are shown in Fig. 12.

![Monitoring results](image)

**Fig. 11. Monitoring results**

![Contribution plot](image)

**Fig. 12. Contribution plot in (a) and fault diagnosis result in (b)**

According to result in Fig. 12(a), naphtha mass flow rate in coil 6 (number 7), COT in the first and fourth tube of coil 3 (number 36 and 39), and outlet pressure in coil 2 (number 53) display high contribution rates. 16 variables residuals are input into SDG network diagram (Fig. 5).

Fault diagnosis result based on SDG in Fig. 12(b) indicates that crossover temperature is low due to high naphtha mass flow rate in coil 6 and high COT in whole coil 6 which were caused by high temperature in furnace A side. Generally, high naphtha feed will result in low COT. However, in 65582th samples moment, naphtha feed and COT are high in coil 6 at the same time and COT is low in coil 3 without naphtha feed changing in coil 3. Temperature in furnace A side is high. Thus, this fault root cause is the uneven heating in furnace which cause by non-uniform fuel gas feed in nozzles.

As shown above, according to variable contribution rate analysis in PCA, the fault result and cause are mixed together, with root cause unrevealed sometimes. With SDG the connection among variables which have contribution to the process deviation can be further recognized.

5. CONCLUSIONS

In this paper, process monitoring and fault diagnosis based on PCA-SDG is implemented and tested during the operation of an ethylene cracking unit. It can not only correctly identify active process adjustment or passive fluctuation, but also realize the real-time multiple operational condition monitoring with 98% true positive rate. Root causes are also isolated by SDG reasoning based on PCA, and matching perfectly with operation records.

REFERENCES


