A Modified Dynamic PLS for Quality Related Monitoring of Fractionation Processes

Xue Xu  Qiang Liu  Jinliang Ding

State Key of Laboratory of Synthetical Automation for Process Industries, Northeastern University, Shenyang 110819, China

(e-mail: xuxueneu@126.com; liuq@mail.neu.edu.cn; jlding@mail.neu.edu.cn).

Abstract: The fractionation process is a typical dynamic process, and practitioners highly pay attention to the quality-related abnormal in the real refining processes. In this paper, a modified dynamic PLS (MDPLS) modeling method and the corresponding process monitoring strategy are proposed. The main contributions of the proposed method are in the following. First, a clear dynamic relation is captured between process data and quality indices. Moreover, the process and quality space are comprehensively divided into dynamic quality-related subspace, static quality-unrelated subspace as well as the residual space for improving the performance of monitoring. Finally, the effectiveness of the proposed algorithm is demonstrated with the data from a real fractionation process.

Keywords: Fractionation process, Dynamic modeling, Quality related monitoring

1. INTRODUCTION

Fluid catalysis and cracking unit (FCCU) is an important process for converting heavy oil fractions into more valuable light production in petroleum processing. Fractionation column is an indispensable unit that produces gasoline and diesel oil in fluid catalysis and cracking processes. The two kinds of oil are major products of FCCU. Endpoint is a critical quality index of gasoline and diesel. Therefore, it is playing a dominant role in monitoring endpoint-related faults with the fractionation column to help improve product quality and operation safety.

The fractionation processes with multiple process variables are highly dynamics and correlated that make it be difficult to build first principles model accurately [1-3]. As a result, model-based process monitoring and fault diagnosis approaches are not applicable. Fortunately, due to the extensive application of distributed control system in recent years, a large number of valuable data with production process have been collected. The researchers rely on the available data to achieve data-based operation optimization, process monitoring as well as fault diagnosis [4-10]. Especially, multivariate statistical process monitoring (MSPM) is the most popularly utilized [11-15]. The principal component analysis (PCA), partial least squares (PLS) method project high-dimensional and correlated process variables into lower-dimensional latent space in chemical industry.

As for the fractionation processes, Thomas C et al. [16] employed PCA method for extracting latent variables, with two models based on hourly and minute-by-minute for monitoring distillation column of FCCU. Taking into account the time-varying behavior, such as changes of crude oil property, decay of catalyst activity in reactor, operation transform, Alghazzawi A et al. [17] proposed a recursive multi-block PCA monitoring strategy which can quickly identify and isolate abnormality in the crude distillation unit. In the real industry, auto-correlation and lag cross-correlation or dynamics exist among large-scale process variables. Therefore, Gao et al. [18] proposed indiscernibility dynamic kernel PCA, where two parameters—the indiscernibility and the degree of cross are defined for reducing data dimension. The algorithm that can extract the dynamics is utilized to moniter distillation column with lower missed alarm rate than the conventional method. Process monitoring based on PCA algorithm monitors only the process variables in unique layer, while abnormal variability may not affect product quality, due to compensation from feedback control in the fractionation processes. Unlike PCA, the PLS method is a powerful approach for establishing relationships of data from two layers [19, 20]. PLS-based process monitoring approaches extract latent factors according to the maximum co-variance criterion via nonlinear iterative partial least squares algorithm [21]. The work of [22] developed an improved multiscale PLS method, which combined PLS algorithm and wavelet analysis to build the model, with using generalized likelihood ration (GLR) testing in the residual space. The monitoring performance of this approach outperforms than PLS-based algorithm.

However, the above some methods only consider the static latent variables, which cannot capture the dynamics. Then it will result in high missed alarm in process monitoring. Although a few methods took into account the dynamics in terms of augmented data, which cannot make accurate description of the dynamic and static structure in the data. Therefore, in this paper, a new dynamic PLS method for
quality-related monitoring is proposed and applied into fractionation processes. The contributions of the work are in the following. First, the process data space is decomposed into dynamic quality-related and static quality-unrelated subspace, respectively. For quality-unrelated space, it can contain a few large variations that is further divided by performing PCA decomposition to improve the monitoring performance.

The remainder of this paper is organized as follows. In section 2, fractionation process of FCCU is described. In section 3, the PLS method is reviewed briefly and a modified dynamic PLS approach is presented, and the corresponding monitoring strategy is given in later. The proposed algorithm is applied in the real fractionation processes in section 4. Finally, the conclusions are concluded and the following work is discussed in the section 5.

2. FRACTIONATION PROCESS DESCRIPTION

Fractionating column is an important unit in catalytic cracking processes, whose process structure diagram is shown in Figure 1. The high-temperature oil-gas mixture enters the fractionation column from bottom column, and then are separated several fractions according to different boiling temperatures. Among them, overhead product is gasoline and side-lines product is diesel. Gasoline and diesel are the main products of FCCU. The endpoint is an important quality index of gasoline and diesel. The major process variables relevant endpoint of gasoline are top tower temperature, top pressure, flow of top reflux, the 1st pump-around reflux flow, that of diesel are top pressure, column bottom temperature, the 1st pump-around reflux temperature, respectively. According to gasoline, diesel oil as the major fuel, endpoint of gasoline is extremely high, which will make it difficult to evaporate and burn the gasoline completely. The endpoint of diesel will affect the mobility under different climate conditions. Further, the endpoint also has an effect on product distribution in the production processes.

Therefore, a great deal of researchers focus on anomalies monitoring of fractionation process. In the early researches, researchers look forward to establish first principle model for solving the abnormal monitoring, whereas several assumptions are made to simplify the model because of difficulties in the solution of quite complex mechanism model. However, the above strategy will lead to a great difference between the built model and the real process [15]. As a result, considering the above drawbacks, multivariate statistical process monitoring method provides a data-driven framework to monitor the fractionation process. The fractionation process is a typical dynamic process, with dynamic characteristics derived from the interactions among complex recycle reflux. Therefore, a quality related modified dynamic PLS monitoring method is proposed for monitoring the fractionation processes.

3. A MODIFIED DYNAMIC PLS MODELING FOR PROCESS MONITORING

3.1 Reviews of PLS

Collect normal process and quality data and construct inputs $X \in \mathbb{R}^{n \times m}$ containing $n$ samples with $m$ variables, and outputs $Y \in \mathbb{R}^{n \times r}$ containing $n$ samples with $r$ variables. Then, the scaled data is projected to low-dimensional space as the following PLS model [23]:

$$
X = \sum_i t_i p_i^T + E = T P^T + E
$$

$$
Y = \sum_i t_i q_i^T + F = T Q^T + F
$$

where $T \in \mathbb{R}^{n \times i}$ is the score matrix of inputs $X$, $P \in \mathbb{R}^{i \times m}$ is the loading matrix of inputs $X$, $Q \in \mathbb{R}^{i \times r}$ is the loading matrix of outputs $Y$, $i$ is the number of latent factors, $E$ and $F$ are residuals for inputs and outputs, respectively. The detailed PLS method is described in literatures [24, 25]. The matrix $W$ is weight matrix of deflation matrix $X = X - t_i p_i^T$. In order to represent $t_i$ in terms of original data $X$, the following equation is utilized.

$$
T = X R
$$

where $R = W (P^T W)^{-1}$ from [26].

3.2 A modified dynamic PLS (MDPLS) modelling algorithm

Generally, both auto-correlations and cross-correlations exist inside real industrial data simultaneously. The above static PLS model cannot capture dynamics in variables. As a result, missed alarm rate may increase when static model is applied to monitor dynamic processes. In this paper, a new dynamic latent variable monitoring algorithm is proposed and applied to quality-related fault detection.
The explicit dynamic and static relations between variables is modeled by the dynamic inner PLS (DiPLS) algorithm[27]:

\[
\begin{align*}
\mathbf{X} &= \mathbf{T} \mathbf{P}^T + \mathbf{E}, \\
\mathbf{Y} &= \mathbf{G} (z^{-l}) \mathbf{t}^*_l \mathbf{q}^*_l + \ldots + \mathbf{G} (z^{-1}) \mathbf{t}^*_1 \mathbf{q}^*_1 + \mathbf{F}
\end{align*}
\]  

(2)

where \( \mathbf{G} (z^{-1}) = \mathbf{a}^T z^{-1} \), \( z \in [1, z^{-1}, \ldots, z^{-l}] \) describes the dynamic relationship between the quality data and the dynamic scores \( t_l \), \( z^{-1} \) denotes delay operator. \( l \) is the number of latent variables, which is obtained by cross-validation method.

However, the above approach focuses on building a regression model, rather than supply a statistical model for process monitoring. Further, the procedures of DiPLS algorithm is similar to that of PLS, thus scores \( T \) include variations orthogonal to \( Y \). In order to overcome the aforementioned shortcomings, a modified monitoring strategy is proposed. According to DiPLS method, the predicted quality variable can be shown [27]:

\[
\hat{\mathbf{y}} = \mathbf{T}_0 Q^T - \mathbf{T}_0 \mathbf{B} \mathbf{Q}^T = \mathbf{Z}_{\text{new}} \mathbf{R} \mathbf{B} \mathbf{Q}^T
\]  

(3)

where \( \mathbf{B} = \begin{bmatrix} \beta_{11}, \beta_{12}, \ldots, \beta_{1m} \end{bmatrix}^T \), \( \mathbf{T}_0 = \begin{bmatrix} \mathbf{T}_1 & \mathbf{T}_2 & \ldots & \mathbf{T}_{m-1} \end{bmatrix} \) , with \( \mathbf{T}_l = \mathbf{X}_l \mathbf{R} \).

The dynamic predictable quality variable \( \hat{y} \) is performed singular value decomposition as following:

\[
\hat{y}_l = \mathbf{L}_l C_l H_l^T = \mathbf{L}_l Q_{\text{new}}^T
\]  

(4)

where \( Q_{\text{new}} = \mathbf{H}_l C_l \) contains \( l \) nonzero singular values in descending order and the corresponding right singular vectors.

\[
\mathbf{L}_l = \hat{y}_l \mathbf{H}_l C_l^{-1} = \mathbf{Z}_{\text{new}} R B \mathbf{Q}^T H_l C_l^{-1} = \mathbf{Z}_{\text{new}} R_l
\]  

(5)

where \( \mathbf{R}_l = \mathbf{R} \mathbf{Q}^T H_l C_l^{-1} \) For the remaining static quality-irrelevant process variables via projecting onto the orthogonal of Span \{ \mathbf{R}_l \} , \( \tilde{\mathbf{x}}_l = \mathbf{x} - \mathbf{K}_l \mathbf{R}_l \) ,where \( \mathbf{R}_l = (\mathbf{R}^T \mathbf{R}_l)^{-1} \mathbf{R}^T \). Due to the large variation in the residual subspace, PCA decomposition is further realized on \( \tilde{\mathbf{x}}_l \) with \( l \) components

\[
\begin{align*}
\mathbf{x}_l &= \mathbf{T}_l \mathbf{P}_l + \mathbf{E}_l, \\
\mathbf{y} &= \tilde{\mathbf{t}}^*_l \mathbf{Q}^T + \mathbf{F}
\end{align*}
\]  

(6)

where \( \mathbf{T}_l \) is the principal component score matrix that is useless to predict outputs, \( \mathbf{E}_l \) represents residuals, respectively. According to the proposed MDPLS method, the outer model of input and output spaces are decomposed into the following form:

\[
\begin{align*}
\mathbf{X} &= \mathbf{L}_l \mathbf{R}_l + \mathbf{T}_l \mathbf{P}_l + \mathbf{E}_l, \\
\mathbf{Y} &= \mathbf{T}_l \mathbf{Q}^T + \mathbf{F}
\end{align*}
\]  

(7)

3.3 MDPLS-Based Monitoring

Given a new sample \( \mathbf{x}_{\text{new}} \), it can be calculated as follows:

\[
\begin{align*}
t_l &= \mathbf{R}_l^T \mathbf{x}_{\text{new}}, \\
\mathbf{t}_l &= \mathbf{P}_l^T \tilde{\mathbf{x}}_l, \\
\mathbf{e}_l &= (1 - \mathbf{P}_l^T \mathbf{P}_l) \tilde{\mathbf{x}}_l
\end{align*}
\]  

(8)

The score vector \( t_l \), can be monitored via Hotelling’s \( T^2 \) statistic:

\[
T^2_l = t_l^T \mathbf{A}_l t_l
\]  

(9)

where \( \mathbf{A}_l \) is the covariance of \( t_l \).

\( \mathbf{e}_l \) can be monitored via \( Q \) statistic:

\[
Q_{\mathbf{e}_l} = \left\| \mathbf{e}_l \right\|
\]  

(10)

The \( t_l \) is stationary time independent series, which is not suitable to monitor. As a consequence, it can be described as a stationary time series model with auto-regressive mode as following:

\[
t_l(k) = y_l t_l(k-l) + \ldots + y_n t_l(k-n) + \nu(k)
\]  

(11)

where \( y_l \) is the parameters of auto-regressive model, \( \nu \) is model order, it can be obtained by Bayesian Information Criterion. The parameters are solved by the least squares algorithm as following:

\[
\phi = \begin{bmatrix} \sum_{i=1}^n \phi(i) \phi(i) \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^n \phi(i) t_l(i) \end{bmatrix}
\]  

(12)

where \( \phi = [y_1, \ldots, y_n] \), \( \phi(k) = [t_l(k-1), \ldots, t_l(k-n)] \).

As \( \mathbf{v}_l \) can be time independent, it can be monitored by statistic \( T^2 \).

\[
T^2_l = \mathbf{v}_l^T \mathbf{A}_l^T \mathbf{v}_l
\]  

(13)

where \( \mathbf{A}_l \) is covariance of \( \mathbf{v} \), \( \mathbf{v} = [v_1, \ldots, v_n]^T \).

**Table 2 Monitoring statistics and control limits**

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Calculation</th>
<th>Control limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>( T^2_l )</td>
<td>( t_l^T \mathbf{A}_l^{-1} t_l )</td>
<td>( \frac{l_n(n^2-1)}{n-l_n} F_{n-l_n,n-l_n} )</td>
</tr>
<tr>
<td>( T^2_l )</td>
<td>( \mathbf{v}_l^T \mathbf{A}_l^{-1} \mathbf{v}_l )</td>
<td>( \frac{n(n^2-1)}{n-l_n} F_{n-l_n,n-l_n} )</td>
</tr>
<tr>
<td>( Q_{\mathbf{e}_l} )</td>
<td>( \left| \mathbf{e}_l \right| )</td>
<td>( (S_\nu / 2 \mu_\nu) \chi^2(2 \mu_\nu / S_\nu) )</td>
</tr>
</tbody>
</table>

In this paper, the monitoring statistics and corresponding control limits are listed in the Table 2. Among them, where \( l_n, l_r \) are the number of the latent variables, \( n \) is the number of training samples, \( F_{n-l_n,n-l_n} \) indicates \( F \)-distribution with \( n-l_n, n-l_r \) degrees of freedom under \( \alpha \) confidence coefficient. \( \chi^2(2 \mu_\nu / S_\nu) \) represents \( \chi^2 \)-distribution with \( 2 \mu_\nu, S_\nu \) degrees of freedom, where \( S_\nu \) means the variance of the training samples, and \( \mu_\nu \) represents mean of training samples, respectively [28, 29].

4. THE SCENARIO STUDIES IN QUALITY RELATED MONITORING OF FRACTIONATION PROCESSES

This section applies the proposed MDPLS method to endpoint relevant monitoring in fractionation processes. The collected data contain 28 process variables and 2 quality variables (endpoint of gasoline and endpoint of diesel). In this paper, a number of 600 samples were collected, of which
300 samples under fault-free condition, and additional 300 samples under fault condition. The prediction performance of two quality variables with PLS and dynamic PLS can be seen in the Fig. 3 and Fig. 4, respectively. The dominant process variables are shown in the Fig. 5. It is obvious that the dynamic PLS have better prediction accuracy than PLS.

![Gasoline endpoint prediction](image1)

![Diesel endpoint prediction](image2)

**Fig. 3. Quality prediction of PLS**

![Gasoline endpoint prediction](image3)

![Diesel endpoint prediction](image4)

**Fig. 4. Quality prediction of dynamic PLS**

To demonstrate the effectiveness of the algorithm, quality related faults 1 and faults 2 are modeled and monitored, comparing with conventional dynamic PLS (DPLS) methods. In the fractionation processes, 1st pump-around flowrate and the top tower pressure with too low are common faults. The former (fault 1) can lead to a higher temperature with the top and middle of the fractionating column, even further lead to the phenomenon of punching column, which will make the endpoint of gasoline and diesel rise, and also affect the heat source supply for the downstream devices. The latter (fault 2) will also have impact on the product quality.

The monitoring performance with traditional DPLS and the MDPLS under fault scenario 1 is described in Fig. 6. It is seen from $r^2$ statistic with Fig. 6 (b) that the endpoint relevant fault is detected precisely from the 183th to 267th sample and have a trending of returning to the normal condition after the 267th samples because of the closed-loop controller to decrease the impact of the fault on process. Nevertheless, the $r^2$ statistic from Fig. 6 (a) cannot detect the influence of feedback control on quality indexes, which still above the control limit after the 267th sample considering as the false alarms. Even if $r^2$ statistic in Fig 6 (b) above the control limit, which is a process-specific variation considering as
normal situation when concentrating on the relevant endpoint fault. Therefore, the effectiveness of proposed monitoring method is demonstrated via the above experimental results. When the fractionation process under fault 2 situation, the monitoring results with DPLS and the MDPLS algorithms are displayed in Fig 7(a) and Fig (b), respectively. The $r^2_i$ statistic with Fig 7(b) tends to exceed the control limit from the 145th sample, which denotes the quality-related faults can be detected successfully. However, the $T$ statistic with DPLS method from Fig.7 (a) have missed alarms among 165-185, 231-240,267-280 samples.

Fault 1:

![Graphs showing monitoring results for fault 1]

Fault 2:

![Graphs showing monitoring results for fault 2]

Fig. 6 Fractionation process monitoring results (faulty 1)
5. CONCLUSIONS

This paper proposed a modified dynamic PLS quality-related monitoring approach for fractionation processes. It provides a simple implementation to achieve comprehensive monitoring. The effectiveness of the proposed method is illustrated by the application results on a real fractionation processes. The corresponding fault diagnosis strategy of the proposed monitoring scheme can be developed for localizing the faults in the future work.

ACKNOWLEDGEMENTS

This work was financially supported in part by the National Natural Science Foundation of China under Grants 61590922, 61525302, 61673097, 61573022, and 61490704, the Project of Ministry of Industry and Information Technology of China under Grand 20171122-6, the Projects of Shenyang under Grand Y17-0-004, and the Fundamental Research Funds for the Central Universities under Grand N160801001, N161608001.

REFERENCES


(26) De Jong S, SIMPLS: an alternative approach to partial least squares regression, Chemometric and Intelligent Laboratory Systems, 1993, 18, 251–263.

