A survey of the application of basic data-driven and model-based methods in process monitoring and fault diagnosis

S. X. Ding∗ P. Zhang† T. Jeinsch** E. L. Ding***
P. Engel**** W. Gui†

* Institute for Automatic Control and Complex Systems (AKS), University of Duisburg-Essen, 47057 Duisburg, Germany
** Lausitz University of Applied Sciences, 01968 Senftenberg, Germany
*** University of Applied Sciences Gelsenkirchen, 45877 Gelsenkirchen, Germany
**** PC-Soft GmbH, 01968 Senftenberg, Germany
† School of Information Science and Engineering, Central South University, Hunan, P.R. China

Abstract: Basic data-driven and model-based process monitoring and fault diagnosis methods are surveyed from the application viewpoint. The main objective is to study the needed modifications and/or combined use of these methods under different real operating conditions.

Keywords: Process monitoring; fault diagnosis; PCA; PLS; observer-based FDI; Parity space methods; GLR; adaptive methods; recursive computation.

1. INTRODUCTION

Process monitoring and fault diagnosis (PM-FD) are currently receiving considerably increasing attention both in the application and research domains. There are a great number of PM-FD methods, most of them are well described and included in the monographs by Gertler [1998], Mangoubi [1998], Chen and Patton [1999], Patton et al. [2000], Russell et al. [2000], Gustafsson [2000], Chiang et al. [2001], Blanke et al. [2003], Simani et al. [2003], Isermann [2006], Ding [2008]. The recent surveys by Kinnert [2003], Venkatasubramanian et al. [2003a,b], Qin [2003], Zhang and Ding [2008], Mangoubi et al. [2009], Qin [2009], Hwang et al. [2010] provide the reader with the excellent review and analysis of the current development of the advanced and sophisticated PM-FD schemes and their interconnections.

In this paper, we shall survey basic data-driven and model-based methods, including principal component analysis (PCA), partial least squares (PLS), generalized likelihood ratio (GLR), Kalman filter (KF), fault detection filter (FDF), diagnostic observer (DO) and parity space (PS) schemes, from the viewpoint of their application. Thanks to their simple forms and less restrictive requirements on the design and engineering efforts, these PM-FD methods are widely used in many industrial sectors (Russell et al. [2000], Gustafsson [2000], Chiang et al. [2001], Isermann [2006], Qin [2009], Hwang et al. [2010]). In our work with industrial and research projects, we notice that modifications on these methods are often helpful. We also observe that some of these methods may suffer a considerable loss in the PM-FD performance if the required assumptions/conditions cannot be satisfactorily met in the industrial environment. These observations motivate us to review these methods aiming at understanding their original ideas, the needed conditions for a successful application and their limitations. In this study, we have identified questionable applications of some of these methods and the needs for a combined use of data-driven and model-based PM-FD methods. Motivated by it, we have further reviewed the (existing) modifications of the above-mentioned PM-FD methods and proposed some alternative solutions and schemes for a combined use of some data-driven and model-based PM-FD methods. Our goal is to improve the applicability, capacity and efficiency of these basic PM-FD methods in the industrial processes and plants without loss of their simplicity. It is the main objective of this paper to report our results and to initial a discussion.

The paper is organized as follows. In Section 2, the basic PM-FD methods are reviewed in their standard form. Section 3 is dedicated to the study on the application of the basic PM-FD methods in their original form and under ideal operation conditions. In Section 4, problems with the system dynamics and uncertainties met by the application of the PM-FD methods under industrial conditions are addressed. The focus is on the combined use of the data-driven and model-based methods and the application of the adaptive and recursive computation techniques.

Notations The notations adopted throughout this paper are fairly standard. \( R^n \) denotes the \( n \)-dimensional Euclidean space and \( R^{n \times n} \) the set of all \( n \times n \) real matrices. The superscripts \( "T" \), 
\( -1 \) and 
\( ^+ \) respectively stand for the transpose, inverse and pseudo-inverse of a matrix.

\[ I_n \] and \( 0_n \) denote the identity and zero matrix with appropriate dimension, respectively, and \( \text{diag}(\cdot, \cdot, \cdot) \) a
diagonal matrix. $\mathcal{E}(\cdot), \text{cov}(y,x)$ represent mean value and covariance of $y$ and $x$. $x \sim \mathcal{N}(0, \Sigma)$ means that $x$ is normal distributed with zero mean and covariance (matrix) $\Sigma$. $\chi^2(l)$ and $\mathcal{F}(l, N)$ stand for $\chi^2$-distributed with $l$ degrees of freedom and $\mathcal{F}$-distributed with $l, N$ degrees of freedom, respectively. $\| \cdot \|$ represents the Euclidean norm, and

$$\| \xi(k) \|_{2,[0,N]} = \sqrt{\sum_{k=0}^{N} \xi^T(k) \xi(k)}$$

for a given vector $\xi(k)$ the $l_{2,[0,N]}$-norm.

2. BASIC PM-FD SCHEMES

In this section, we review the basic PM-FD schemes and present them in the form of the off-line design and on-line computation algorithms. Their application will be studied in the subsequent sections.

2.1 Principal component analysis

Consider a process with $m$ measurement signals which are denoted by a (column) observation vector. The off-line design procedure of a standard PCA approach (Jackson and Mudholkar [1979], Chiang et al. [2001]) used for the purpose of fault detection consists of

Step 1: $N$ samples for each measurement are first collected and normalized to zero mean, and often in addition scaled to unit variance and finally written into $Z^T = [z_1 \cdots z_N] \in \mathbb{R}^{m \times N}$ with $z_i \in \mathbb{R}^m, i = 1, \ldots, N$, denoting the $i$-th (normalized) observation vector.

Step 2: Form covariance matrix $\Sigma \approx \frac{1}{N-1} Z^T Z = P \Lambda P^T, \Lambda = \text{diag}(\sigma_1^2, \cdots, \sigma_m^2)$

where $\sigma_1 \geq \cdots \geq \sigma_m > 0$ are the singular values and $P$ is the unitary matrix consisting of the eigenvectors.

Step 3: Determine the number of the principal components $l$ (Valle et al. [1999]) and, associated with it, divide $P, \Lambda$ into two parts

$$\Lambda = \begin{bmatrix} \Lambda_{pc} & 0 \\ 0 & \Lambda_{res} \end{bmatrix}, \Lambda_{pc} = \text{diag}(\sigma_1^2, \cdots, \sigma_l^2)$$

$$\Lambda_{res} = \text{diag}(\sigma_{l+1}^2, \cdots, \sigma_m^2) \in \mathbb{R}^{(m-l) \times (m-l)}$$

$$P = [P_{pc} \ P_{res}] \in \mathbb{R}^{m \times m}, P_{pc} \in \mathbb{R}^{m \times l}$$

Step 4: Setting thresholds for the SPE (squared prediction error) and $T^2$ statistic for a (given) significance level $\alpha$ (Chen et al. [2004]):

$$\text{SPE : } J_{th,SPE} = \theta_1 \left( c_{\alpha} \sqrt{\frac{2 \chi^2_{l,0}}{l}} + 1 + \frac{\theta_2 h_0 (h_0 - 1)}{\theta_1^2} \right)^{\frac{1}{h_0 - 1}}$$

$$T^2 : J_{th,T^2} = \frac{1}{N(N-l)} \chi^2(l, N-l)$$

where $c_{\alpha}$ is the normal deviate corresponding to the upper $1-\alpha$ percentile,

$$\theta_1 = \sum_{j=l+1}^{m} (\sigma_j^2)^i, i = 1, 2, 3, h_0 = 1 - \frac{2\theta_1 \theta_2}{3\theta_3^2}$$

The on-line computation for PM-FD consists of

Step 1: Normalization of the new measurement sample

$$z = (I - P_{pc} P_{pc}^T) z^T$$

Step 2: On-line computation of SPE and $T^2$ statistic

$$SPE = \| (I - P_{pc} P_{pc}^T) z \|^2 = z^T (I - P_{pc} P_{pc}^T) z$$

$$T^2 = z^T P_{pc} \Lambda_{pc}^{-1} P_{pc}^T z$$

Step 3: Fault detection logic is

$$SPE \leq J_{th,SPE} \quad \text{and} \quad T^2 \leq J_{th,T^2} \quad \Longrightarrow \quad \text{fault-free, otherwise faulty}$$

2.2 Dynamic PCA (DPCA)

DPCA (Chiang et al. [2001]) is a natural extension of the PCA to deal with those dynamic processes, which can be roughly expressed in terms of the serial correlations between the actual observation vector with the previous e.g. $h$ observations. Assume that for the training purpose observations in the time interval $[k-N, k]$ are available. Let $z_k^T = [z_k^T(i) \cdots z_k^T(i-h)]$, the data are formed into

$$Z_k(h) = \begin{bmatrix} z_k^T(h) \\
\vdots \\
z_k^T_{(N-h)}(h) \end{bmatrix} \in \mathbb{R}^{(N-h+1) \times (m(h+1))}$$

The remaining off-line and on-line steps are then identical with the ones given in the standard PCA. Note that for the on-line computation of $T^2$ and $SPE$ statistic at the time instant $k$, the (extended) observer vector $z_k(h)$ is needed.

2.3 Partial least squares

Suppose that the process under consideration has a measurement vector $y \in \mathbb{R}^m$ and the process variables under monitoring build vector $\theta \in \mathbb{R}^n, 1 \leq \alpha < m$. The off-line training/modelling of the PLS method consists of (Höskuldsson [1988], Chiang et al. [2001])

Step 1: $N$ samples of $y$ and $\theta$ are first collected, normalized to zero mean and scaled to unit variance, and finally written into

$$Y = \begin{bmatrix} y_1^T \\
\vdots \\
y_N^T \end{bmatrix} \in \mathbb{R}^{N \times m}, \Theta = \begin{bmatrix} \theta_1^T \\
\vdots \\
\theta_N^T \end{bmatrix} \in \mathbb{R}^{N \times \alpha}$$

Step 2: Recursive computation: for $i = 1, \cdots, \gamma$

$$(w_i^*, q_i^*) = \arg \max_{\|w_i\| = 1, \|q_i\| = 1} w_i^T Y_i^T \Theta q_i, Y_i = Y - t_i w_i^T \Xi$$

$$t_i = Y_i w_i^*, p_i = Y_i^T t_i / \| t_i \|, Y_{i+1} = Y_i - t_i p_i^T$$

where $\gamma$ is determined by applying a known stopping criteria.

Step 3: Form $P, Q, R$ and $T$:

$$P = [p_1 \cdots p_\gamma] \in \mathbb{R}^{m \times \gamma}, T = [t_1 \cdots t_\gamma] \in \mathbb{R}^{N \times \gamma}$$

$$Q = [q_1^\gamma \cdots q_\gamma^\gamma] \in \mathbb{R}^{\alpha \times \gamma}, R = [r_1 \cdots r_\gamma] \in \mathbb{R}^{\alpha \times \gamma}$$

$$r_i = w_i, r_i = \prod_{j=1}^{i-1} (I - w_j p_j^T) w_i, i = 2, \cdots, \gamma$$

$$T = Y R, P^T R = R^T P = I \in \mathbb{R}^{\gamma \times \gamma}$$
and in terms of them matrices $Y, \Theta$ are written into
\[ Y = TP^T + EY, \Theta = TQT + F\Theta \] (10)
Step 4: Setting thresholds for the SPE and $T^2$ statistic under a (given) significance level $\alpha$
\[ SPE : J_{th,SPE} = g\chi^2_{h,\alpha} \] (11)
\[ T^2 : J_{th,T} = \gamma \left( N^2 - 1 \right) N(\gamma, N - \gamma) \] (12)
where $g\chi^2_{h,\alpha}$ is the $\chi^2$ distribution with scaling factor $g = \frac{s}{\mu_{SPE}}$ and $h = \frac{2\mu_{SPE}}{s}$ degree of freedom, $\mu_{SPE}, S$ are respectively the mean and variance of SPE (Nomikos and Macgregor [1995]).

The on-line computation for PM-FD consists of
Step 1: Normalization of the new measurement sample
Step 2: On-line computation of $T^2$ and SPE statistic, and prediction of $\theta$:
\[ T^2 = y^T R^{-1} \left( \frac{T^T T}{N-1} \right) R^T y \] (13)
\[ SPE = \sqrt{ \left( I - PR^T \right) y} \] (14)
Step 3: Detection of faults in $\theta$: fault detection logic is
\[ T^2 > J_{th,T^2} \text{ and } SPE \leq J_{th,SPE} \Rightarrow \text{faulty in } \theta, \text{ otherwise fault-free} \]

2.4 Kalman filter scheme
KF is a standard technique (Hassibi et al. [1999]) to deal with PM-FD problems in stochastic processes. Assume that the process/plant is described by
\[ x(k+1) = A(k)x(k) + B(k)u(k) + w(k) \] (15)
\[ y(k) = C(k)x(k) + D(k)u(k) + v(k) \] (16)
and the above model is available, where system matrices $A(k), B(k), C(k), D(k)$ are known, $x(k) \in \mathcal{R}^n, u(k) \in \mathcal{R}^l, y(k) \in \mathcal{R}^m$ denote the vectors of the state variable, process inputs and outputs, respectively. $w(k) \in \mathcal{R}^r$, $v(k) \in \mathcal{R}^s$ are assumed to be normal distributed satisfying
\[ \mathcal{E} \left( w(k) \right) = 0, \mathcal{E} \left( v(k) \right) = 0 \] (17)
\[ \mathcal{E} \left[ \begin{bmatrix} w(i) \\ v(i) \\ x(0) \end{bmatrix} \right] = \begin{bmatrix} Q(i) & S(i) \\ S^T(i) & R(i) \end{bmatrix} \begin{bmatrix} 1, i = j \\ 0, i \neq j \end{bmatrix} \] (18)
The on-line realization of a KF used for the PM-FD purpose consists of
\[ \hat{x}(k+1) = A(k)\hat{x}(k) + B(k)u(k) + L(k)r(k) \] (19)
\[ r(k) = y(k) - \hat{y}(k), \hat{y}(k) = C(k)\hat{x}(k) + D(k)u(k) \] (20)
\[ L(k) = \left( A(k)X(k)C^T(k) + S(k) \right) W^{-1}(k) \] (21)
\[ X(k+1) = A(k)X(k)A^T(k) + Q(k) - L(k)W(k)L^T(k) \] (22)
with $\hat{x}(0) = 0$, where $r(k)$ is the so-called residual vector, which is independent of $u(k)$, and $cov(r(k), r(k)) = W(k)$.

2.5 Observer-based schemes
Assume that the process model described by
\[ x(k+1) = Ax(k) + Bu(k), y(k) = Cx(k) + Du(k) \] (23)
is available, where system matrices $A, B, C, D$ are constant and known. The on-line computation of the observer-based schemes, including FDF and DO (Ding [2008]), is dedicated to generating the residual vector $r(k)$
\[ \text{FDF: } \begin{cases} \hat{x}(k+1) = A\hat{x}(k) + Bu(k) + L(y(k) - \hat{y}(k)) \\ r(k) = V(y(k) - \hat{y}(k)), \hat{y}(k) = C\hat{x}(k) + Du(k) \end{cases} \] (24)
\[ \text{DO: } \begin{cases} z(k+1) = A_zz(k) + B_zu(k) + L_zy(k) \in \mathcal{R}^s \\ r(k) = V(y(k) - C_zz(k) - D_zu(k) \in \mathcal{R}^s \] (25)
where $s$ is the order of the DO.

The off-line design of an FDF consists in the computation of the observer gain matrix $L$ and post-filter $V$, while for the DO solving the so-called Luenberger equations with
\[ TA - A_LT = L_zC, C_LT = VC, B_z = TB - L_zD \] (26)
\[ D_z = VD, A_z \in \mathcal{R}^{r \times s}, T \in \mathcal{R}^{s \times m} \] (27)
for $A_z, T, L_z, V$ and $C_z, B_z, D_z$ is involved (Patton et al. [2000], Ding [2008]).

2.6 Parity space scheme
On the assumption of the process model (23), the off-line design of the PS-FD system consists in finding the so-called parity vector $\alpha_s$ by solving
\[ \alpha_s \Gamma_s = 0, \Gamma_s^T = \begin{bmatrix} C^T & A^T & C \end{bmatrix} \] (27)
with $(s+1)m > n$ (Ding [2008]). The on-line computation is dedicated to the residual generation
\[ r(k) = \alpha_s(y_s(k) - H_{s,u}u_s(k)), y_s(k) = \begin{bmatrix} y(k-s) \\ \vdots \\ y(k) \end{bmatrix} \] (28)
\[ u_s(k) = \begin{bmatrix} u(k-s) \\ \vdots \\ u(k) \end{bmatrix}, H_{s,u} = \begin{bmatrix} D & 0 & \cdots & 0 \\ CB & D & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \end{bmatrix} \] (29)

2.7 A practical FD scheme based on the GLR technique
On the assumption that the process is described by
\[ r = \varepsilon + f \in \mathcal{R}^m, f = \begin{cases} 0: \text{fault-free} \\ \neq 0: \text{fault} \end{cases} \] (29)
where $\varepsilon \sim \mathcal{N}(0, \Sigma)$, $\Sigma$ is constant and known, and $N$ observations, $r_1, \ldots, r_N$, are available, the GLR technique is a powerful tool to detect the change in $f$ (Basseville and Nikiforov [1993]). Given the significance level $\alpha$, the off-line computation deals with determining $X_\alpha$ using the table of $\chi^2$-distribution with $m$ degree of freedom so that $\text{prob}(\chi > \chi_\alpha) = \alpha$, where $\text{prob}(\chi > \chi_\alpha)$ denotes the probability of $\chi > \chi_\alpha$, and setting the threshold
\[ J_{th} = \chi^2 \] (30)
The on-line computation for PM-FD includes

12382
Step 1: computation of the testing statistic
\[ J = N \bar{r}^T \Sigma^{-1} \bar{r}, \bar{r} = \sum_{i=1}^{N} r_i \quad (31) \]

Step 2: detection logic:
\[ J \leq J_{th} \Rightarrow \text{ fault-free, otherwise a fault is detected} \]

3. ON THE APPLICATION UNDER REQUIRED OPERATION CONDITIONS

In this section, we shall study the application of the PM-FD methods introduced in the last section on the assumption that they are applied under the required operating conditions.

3.1 On the PCA/DPCA and its relation to the GLR

A successful application of the PCA-based PM-FD is based on the assumption that the process observation is normal distributed with constant mean \( \mu \) and covariance matrix \( \Sigma \), which can also be considered as the process model. In this sense, the first step in the off-line design procedure serves modelling by identifying \( \mu, \Sigma \) using the process data.

The core of the PCA technique is the decomposition of the (normalized) observation space into the so-called principal component subspace and residual subspace, respectively spanned by \( P_{pc} \) and \( P_{res} \). It is well-known that the original idea behind the PCA is to reduce the dimension of a data set, while retaining as much as possible the variation present in the data set (Jolliffe [1986]). Differently, in the PCA-based PM-FD technique, the both projections of \( z \) onto \( P_{pc} \) and \( P_{res} \), i.e.

\[ \tilde{z} = P_{pc} \bar{z}, \tilde{z} = P_{res} \bar{z} \Rightarrow z = \tilde{z} + \tilde{z} \quad (32) \]

should be applied for the FD purpose, since a fault can appear in one of these two subspaces. Note that both \( T^2 \) and SPE indices are of a quadratic form respectively associated with \( \tilde{z}_{pc} = P_{pc}^{T} \tilde{z} \sim \mathcal{N}(0, \Sigma_{pc}) \) and \( \tilde{z}_{res} = P_{res}^{T} \tilde{z} \sim \mathcal{N}(0, \Sigma_{res}) \). As well-known, for a fixed \( N \), the GLR technique with the test statistic (31) and the associated threshold leads to an optimal FD performance. It is evident that the \( T^2 \) index is exactly the GLR test statistic for \( N = 1 \) and thus delivers an optimal FD in the principal component subspace. Similarly, the so-called Hawkins’ statistic with \( T^2_{res} = z^T P_{res} \Sigma_{res}^{-1} P_{res} z \) can also be used for FD in the residual subspace, which is however barely used in practice due to the drawback with the possible ill-conditioning \( \Sigma_{res} \) when the singular values of \( \sigma_{i+1}, \ldots, \sigma_m \) are very close to zero. In (Ding et al. [2010]), a modified statistic and the associated threshold are proposed

\[ T^2_{res} = z^T P_{res} \Sigma_{res}^{-1} P_{res} z, \tilde{z} = \text{diag} \left( \frac{\sigma_1^2}{\sigma_{i+1}^2}, \ldots, \frac{\sigma_m^2}{\sigma_{m-1}^2}, 1 \right) \]
\[ J_{th, res} = \sigma_n^2 X^2_n(m - l), \tilde{z} \sim \mathcal{N}(0, \Sigma_l) \]

which deliver an optimal FD in the residual subspace like the Hawkins’ \( T^2_{res} \) index but without numerical problem.

It is worth pointing out that the statistic associated with the projection \( \tilde{z} \) often provides a (significantly) better FD performance than the one corresponding to the principal subspace. This fact can be easily seen by noting (30) and (31), which can be interpreted as: under the same false alarm rate (expressed in terms of \( \alpha \)) the the statistic built by the projection with a smaller covariance is more sensitive to the faults.

Qin [2003, 2009] has reported the use of the so-called combined indices, which are a simultaneous use of the \( T^2, \text{SPE} (T^2_{SPE}) \) indices and will ensure a high fault detectability. Considering that \( P^T z \sim \mathcal{N}(0, \Lambda) \), \( z^T P A^{-1} P^T z \) is \( \chi^2 \)-distributed with \( m \) degrees of freedom, it is reasonable to introduce the statistic

\[ T^2_{comb} = z^T P \Sigma P^T z, \tilde{z} = \text{diag} \left( \frac{\sigma_1^2}{\sigma_{i+1}^2}, \ldots, \frac{\sigma_m^2}{\sigma_{m-1}^2}, 1 \right) \quad (33) \]

which is a combined index \( T^2_{comb} = \sigma_n^2 \left( T^2 + T^2_{res} \right) \), and the corresponding threshold is (Ding et al. [2010])

\[ J_{th, comb} = \sigma_n^2 X^2_n(m) \quad (34) \]

Following the above discussion, it is clear that \( T^2_{comb} \), \( J_{th, comb} \) is exactly the statistic and the associated threshold delivered by the GLR method and thus offer the optimal FD performance.

The statistic \( T^2_{comb} \) can also be interpreted as a weighted quadratic form of the observation project \( P^T z \). It is interesting to notice that the detection coupled with a stronger variance is less weighted, while the direction with a weaker variance has a stronger weighting.

In the extreme case that the covariance matrix \( \Sigma \) is singular, an SVD yields

\[ \Sigma \approx \frac{1}{N-1} Z^T Z = \left[ \begin{array}{cc} P & P_{\perp} \end{array} \right] \left[ \begin{array}{c} \Lambda \ 0 \\ 0 \ 0 \end{array} \right] \left[ \begin{array}{c} P^T \\ P_{\perp}^T \end{array} \right] \quad (35) \]

As a result, the FD can be achieved using e.g. the statistic (33) with the associated threshold (34) and, in addition, the parity check

\[ P_{\perp}^T z = 0 \quad (36) \]

The corresponding detection logic is

\[ T^2_{comb} \leq J_{th, comb} \quad (36) \text{ and (36) is true } \Rightarrow \text{ fault-free, otherwise a fault is detected} \]

The contribution plots are a standard PCA-based fault identification technique widely used in practice (Chiang et al. [2001]). Recently, Qin [2009], Alcala and Qin [2009], Li et al. [2010a] proposed to estimate/identify the faults by means of reconstruction. The basic idea of reconstruction-based fault estimation is a least square (optimal) estimation of the fault on the assumption that \( z \sim \mathcal{N}(\Psi f, \Sigma) \), where \( f \) denotes the (constant) fault vector and \( \Psi \) a known matrix modelling the distribution of \( f \) in the observation subspace. On the same assumption, we are also able to use the GLR method for estimating \( \Psi f \) and further for \( f \) with

\[ \Psi f \approx \frac{1}{M} \sum_{i=1}^{M} z_i \Rightarrow \hat{f} = \Psi \frac{M}{M} \sum_{i=1}^{M} z_i \quad (37) \]

where \( z_i, i = 1, \ldots, M \), are the \( M \) observations of \( z \) (Ding et al. [2010]). It is well-known that in the GLR framework we can also estimate the change in the covariance matrix \( \Sigma \) caused by a fault (Basseville and Nikiforov [1993]).

3.2 On the PLS and a modified form

The core of the PLS-based PM-FD is the correlation models given in (10). They describe a linear mapping/regression of the (normalized) observation \( y \) to the
process/product quality variable vector $\theta$ and further the subspace in the observation $y$ which is correlated with $\theta$. The original idea behind the PLS is to identify the correlation models based on the covariance matrix $\text{cov}(y, \theta)$ and on the assumption that $y \sim \mathcal{N}(0, \Sigma_y)$, $\theta \sim \mathcal{N}(0, \Sigma_0)$, and apply them to predict $\theta$ using the (on-line) observation $y$ and to detect the faults in $\theta$ subspace using the statistic based on $\hat{y}_0 = PR^T y$ and its residual.

Recently, Zhou et al. [2010] reveal that $\hat{y}_0$ may contain components being uncorrelated with $\theta$, while $\hat{y}_{0,\text{res}} = (I - PR^T) y$ may be partly correlated with $\theta$ but not used for the prediction of $\theta$. Also, $\hat{y}_{0,\text{res}}$ could become large, which is in turn not suitable for the use as a residual signal. From the mathematical viewpoint, it means $\hat{y}_0$ and $\hat{y}_{0,\text{res}}$ are not orthogonal (Li et al. [2010b]). As a result, the PM-FD performance of the PLS in this form may become poor. To solve this problem, they have proposed the so-called total projection technique which results in a further decomposition of $\hat{y}_0$ and $\hat{y}_{0,\text{res}}$. Below is an alternative solution (Yin et al. [2011]). Suppose that the regression

$$\theta = \Psi y + \varepsilon_\theta \in \mathbb{R}^\alpha$$

(38)

holds with $\text{rank}(\Psi) = \alpha$, where $\varepsilon_\theta$ is the part in $\theta$ which is uncorrelated with $y$, i.e. $\mathcal{E}(\theta y^T) = 0$. On the assumption of a large $N$, we have

$$\Theta_T Y = \frac{\Psi y^T Y + E_\theta y^T}{N - 1} \approx \frac{\Psi y^T Y}{N - 1}$$

$(39)$

$$\psi = \Theta_T Y (Y^T) + y$$

(40)

Based on an SVD of $\Psi^T \psi$, $\Psi^T \psi = \left[ U_{\psi} \hat{U}_{\psi} \right] \left[ \Lambda_{\psi} 0 \right] \left[ 0 \right]$ and $y$ can be decomposed into two orthogonal subspaces

$$y = \hat{y} + \tilde{y}, \hat{y} = U_{\psi} \hat{U}_{\psi} y, \tilde{y} = \hat{U}_{\psi} \hat{U}_{\psi} y$$

(41)

It is then reasonable, following the GLR method, to define

$$T_{y}^2 = (N - 1) y^T U_{\psi} (U_{\psi}^T Y Y^T U_{\psi})^{-1} U_{\psi} y$$

(42)

$$T_{\tilde{y}}^2 = (N - 1) \tilde{y}^T \hat{U}_{\psi} \left( \hat{U}_{\psi}^T Y Y^T \hat{U}_{\psi} \right)^{-1} \hat{U}_{\psi} \tilde{y}$$

(43)

$$J_{th,y} = \alpha \left( \frac{N^2 - 1}{N - \alpha} \right) F_{\alpha}(\alpha, N - \alpha)$$

(44)

$$J_{th,y} = \left( \frac{m - \alpha}{N - m + \alpha} \right) \frac{N^2 - 1}{N - \alpha} F_{\alpha}(m - \alpha, N - m + \alpha)$$

(45)

which can then be used to detect the change in the $\hat{y}$-subspace. This alternative solution does not only deliver a better detection and prediction performance but also require less computation in comparison with the standard PLS scheme as given in Subsection 2.3.

It is worth mentioning that the PLS method can also serve as a tool to solve fault isolation, estimation and prediction problems. To this end, data collected in the faulty cases are needed and then used to identify the correlations between the different faults and the observations.

3.3 On the integration of Kalman filter and GLR schemes

A KF delivers an estimate $\hat{x}(k)$ which contains information useful for process monitoring. It is well known that the residual vector $r(k)$ generated by KF (19)-(22) is white and satisfies, in the fault-free case, $r(k) \sim \mathcal{N}(0, W(k))$. For the FD purpose, it is helpful to transform $r(k)$ into

$$\hat{r}(k) = W^{-1/2}(k)r(k) \sim \mathcal{N}(0, I)$$

(46)

On the assumption that the faults to be detected are modelled by two additive terms $E(k), f(k)$ in (15)-(16) with known matrices $E(k), F(k)$ and the fault vector $f(k) \in \mathbb{R}^M$, the dynamics of $\hat{r}(k)$ is governed by

$$\hat{r}_s(k) = \hat{r}_{s,0}(k) + f_s(k), f_s(k) = H_{f,s}(k)f_s(k)$$

(53)

where $H_{f,s}(k)$ is the Hankel matrix similarly structured like $H_{f,u}$ introduced in the PS residual generation, $\hat{r}_{s,0}(k) \sim \mathcal{N}(0, I)$ represents $\hat{r}_s(k)$ for the fault-free case and it is assumed that the system is fault-free before the time instant $k - s$. In order to detect the faults, we can apply the GLR method, which results in

$$J = \hat{r}_s(k)^T \hat{r}_s(k) = \sum_{i=0}^{\infty} \hat{r}(k - i)^T \hat{r}(k - i)$$

(47)

$$J_{th} = \chi^2_n(s + 1)m$$

(48)

The GLR estimate for $f_s(k)$ is given by

$$\hat{f}_s(k) = (H_{f,s}(k))^T \hat{r}_s(k)$$

In case that the system is time invariant and $f$ is constant, the fault detection can be achieved in the steady state by using threshold setting $J_{th} = \chi^2_n(m)$ and the test statistic

$$J = (s + 1) \left( \sum_{i=0}^{s} \hat{r}(k - i) \right)^T \left( \sum_{i=0}^{s} \hat{r}(k - i) \right)$$

The GLR estimate for $f$ is given by

$$\hat{f}(k) = \left( F + C(I - A + LC)^{-1}(E - LF) \right)^T \left( \sum_{i=0}^{s} \hat{r}(k - i) \right)$$

3.4 On the observer-based and parity space schemes

The major difference between the KF and FDF schemes is that the FDF is generally applied to the processes with (deterministic) disturbances. Replacing $w(k), v(k)$ in system model (15)-(16) by $E_d(k) d(k), F_d(k) d(k)$ to model the influence of unknown but $J_{2,[0,N]}$-bounded disturbance vector $d(k) \in \mathbb{R}^k$, Ding et al. [2009a] propose to use the following (time varying) FDF for process monitoring and residual generation,

$$\hat{x}(k + 1) = A(k) \hat{x}(k) + B(k) u(k) + L(k) r(k)$$

(49)

$$r(k) = V^{-1/2}(k) (g(k) - \hat{y}(k))$$

(50)

$$L(k) = (A(k) X(k) C^T(k) + E_d(k) F_d^T(k)) V^{-1/2}(k)$$

(51)

$$X(k + 1) = A(k) X(k) A^T(k) + Q(k) - L(k) L^T(k)$$

(52)

$$X(0) = I, V(k) = C(k) X(k) C^T(k) + F_d(k) F_d^T(k)$$

(53)

It is further demonstrated that for the evaluation functions

$$J_{2,[0,N]} = \sum_{k=0}^{N} r^T(k) r(k) = \| r(k) \|^2 \sum_{[0,N]}$$

(54)

$$J_{MH}(j) = \sum_{k=j}^{j+M-1} r^T(k) r(k), [j, j + M - 1] \in [0, N]$$

(54)
where \( J_{\text{MH}}(j) \), \( j = 0, 1, \ldots, N - M + 1 \), denotes a moving horizon residual evaluation, and a unified threshold \[ J_{\text{th},2,[0,N]} = J_{\text{th,MH}} = \delta^2_{e} \] (55)

\[
\sup \|d(k)\|_{2,[0,N]} \leq \delta^2_{e}, \quad \sup \|x(0)\| \leq \delta^2_{e}
\]

the FDF (49)-(52) delivers an optimal FD performance. The major advantages of this scheme are (a) its similarity with the KF scheme (b) the simple threshold setting for which no engineering effort is needed e.g. for the \( H_{\infty} \)-norm computation (Ding [2008]).

The most popular application of the DO scheme is to generate a residual signal decoupled from a unknown input vector represented e.g. by an additive term \( E_d(k) \) in (23). To this end, the condition, \( T E_d = 0 \), should be added to the Luemberger equations given in (26). It is worth to mention that this scheme is also often used for the fault isolation purpose, in which the faults are decomposed into different groups. A bank of residual signals will then be generated, and each of them is decoupled from some groups of the faults but sensitive to the others.

The difficulty met in practice by applying the DO scheme is solving the Luemberger equations together with \( T E_d = 0 \). In (Ding [2008]), a practical algorithm is provided, which only deals with straightforward solutions of linear equations and can be used both for the DO and parity space schemes. Consider (23) with additive disturbance terms \( d \) and \( \alpha \) solve \( \alpha \), \( \Gamma \), \( H_{s,d} \), \( H_{s,u} \). Generating \( r(k) \) according to (28), we receive a (parity space) residual signal decoupled from \( d(k) \). If we further set

\[
A_z = \begin{bmatrix} 0 & \cdots & 0 & g_1 \\ 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 1 & g_s \end{bmatrix}, \quad L_z = - \begin{bmatrix} \alpha_{s,0} \\ \vdots \\ \alpha_{s,s} \end{bmatrix} - \begin{bmatrix} g_1 \\ \vdots \\ g_s \end{bmatrix}, \quad \alpha_{s,s}
\]

\[
B_z = \begin{bmatrix} \alpha_{s,0} + g_1 \alpha_{s,s} \\ \alpha_{s,1} + g_2 \alpha_{s,s} \\ \vdots \\ \alpha_{s,s-1} + g_s \alpha_{s,s} \end{bmatrix}, \quad D = \Gamma \Gamma^T
\]

then the DO given in (25) delivers \( r(k) \) decoupled from \( d(k) \). In the above setting,

\[
g^T = [g_1 \cdots g_s]
\]

is the design parameter which can be e.g. set equal to zero.

The DO or parity space schemes can only be applied for a successful FD if the residual evaluation and threshold setting are implemented in an integrated way. We would like to recommend an integrated use of the Kalman filter and the GLR for the stochastic processes and the unified solution proposed by Ding et al. [2000], Zhang and Ding [2008] for the deterministic systems. The latter is a time invariant version of the FD scheme given in (49)-(55).

Suppose that a DO is designed using the algorithm given in (56) and the generated residual is decoupled from some disturbances. Consider the error dynamics of the DO (25),

\[
e(k+1) = A_z e(k) + \bar{w}(k), \quad \bar{w}(k) = T w(k) - L_z v(k) \] (58)

\[
r(k) = C_z e(k) + V v(k), \quad e(k) = T x(k) - z(k)
\] (59)

for the stochastic system, and

\[
e(k+1) = A_z e(k) + \bar{E}_d(k), \quad \bar{E}_d = T(E_d - L_z F_d) \] (60)

\[
r(k) = C_z e(k) + V F_d(k), \quad e(k) = T x(k) - z(k)
\] (61)

for the deterministic system. Re-construct the DO as

\[
z(k+1) = A_z z(k) + B_z u(k) + L_z y(k) + L_r(k)
\]

\[
r(k) = V y(k) - C_z z(k) - D_z u(k)
\]

Now, designing the (new) observer gain matrix \( L \), which is equivalently given in (57) for the design approach given in (56), using the Kalman filter scheme (the time invariant version) or the unified solution together with the associated evaluation function/test statistic and the threshold, we are able to construct an optimal FD system.

4. ON THE APPLICATION UNDER INDUSTRIAL OPERATION CONDITIONS

In the industrial applications, we often meet situations, in which the operation conditions are significantly different from those conditions assumed for the application of any basic PM-FD method. In this section, we shall focus on addressing process dynamics and uncertainty issues. From the system theoretic viewpoint, the most efficient way to deal with the process dynamics is the use of a system model. Among those known techniques dealing with uncertainties in the industrial environment, the adaptive technique is an efficient and widely used tool, which allows us to update the process/plant model on-line to cope with the (uncertain) changes in the process/plant. Below, we shall study the PM-FD methods in this context and on the assumption that only process data are available.

4.1 On the dynamic issues

Fig.1 sketches a typical industrial process schematically. It is evident that the data-driven methods PCA (DPCA) and PLS cannot be applied to such processes for an efficient and reliable PM-FD. There are two major reasons: (a) the mean values of the process variables are dynamic process (b) an explicit model for the internal relations between the process/measurement variables and input signals which drive the process operation is missing. For a batch process shown in Fig.1, the multiple-model PCA method by Zhao et al. [2004] provides a possible solution if only those stages with steady behavior are taken into account. Based on certain equivalence between the PCA and PS schemes, Gertler et al. [1999] proposed a PCA-based FD method for dynamic processes. Although model-based PM-FD methods are powerful tools to deal with such dynamic processes, their application is often unrealistic due to the missing models or the sophisticated modelling procedure and the needed system theoretical knowledge.

Recently, Ding et al. [2009c] proposed an approach, based on the work by Qiu and Li [2001], Wang and Qiu [2002], which combines the ideas from PCA/DPCA and observer-based schemes and can be efficiently applied for PM-FD in dynamic processes. Below, we present the major ideas of a PCA-like version of this approach (Li and Qiu [2001]).

Assume that for the training/modelling input and output data \( y(k), u(k) \in [i, i + s + N] \) are available, where \( s \) is
Fig. 1. Schematic description of an industrial batch process

an integer selected to be larger than the possible system order. Similar to the DPCA, let data matrix

\[
Z = \begin{bmatrix} Y \end{bmatrix}, Y = \begin{bmatrix} Y(i) \\ Y(i+s) \end{bmatrix}, U = \begin{bmatrix} U(i) \\ U(i+s) \end{bmatrix}
\]

\[Y(j) = [y(j) \cdots y(j+N)], U(j) = [u(j) \cdots u(j+N)]\]

On the assumption that the dynamics of the process under monitoring is governed by (15)-(16), we have

\[
\begin{bmatrix} Y \\ Y \end{bmatrix} = \begin{bmatrix} \Gamma_s & H_{s,u} \\ 0 & I \end{bmatrix} \begin{bmatrix} X(i) \\ U \end{bmatrix} + \begin{bmatrix} \Phi \\ 0 \end{bmatrix}, \Phi = H_{s,w}W + V
\]

\[X(j) = [x(i) \cdots x(i+N)], H_{s,w} = \begin{bmatrix} W(i) \\ \vdots \\ W(i+s) \end{bmatrix}, W(j) = [w(j) \cdots w(j+N)]
\]

\[W = \begin{bmatrix} V(i) \\ \vdots \\ V(i+s) \end{bmatrix}, V(j) = [v(j) \cdots v(j+N)]\]

For a large \( N \), it turns out

\[
\frac{ZZ^T}{N} \approx \begin{bmatrix} \Gamma_s & H_{s,u} \\ 0 & I \end{bmatrix} \begin{bmatrix} X(i) \\ U \end{bmatrix} Z^T + \begin{bmatrix} \Phi \\ 0 \end{bmatrix} \begin{bmatrix} \Phi \\ 0 \end{bmatrix}^T\]

(63)

Note that \( \Phi \sim N(0, \Sigma_{\Phi}) \). Moreover, the first term in (63) describes the system dynamics and the response to the input signal. Thus, if it is possible to remove it from \( \frac{ZZ^T}{N} \), we are able to build a covariance matrix related to the noises, similar to the PCA, and then use it for the PM-FD purpose. To this end, we do an SVD on \( \frac{ZZ^T}{N} \), which leads to

\[
\frac{ZZ^T}{N} = U_z \begin{bmatrix} \Lambda_{X,U} & 0 \\ 0 & \Lambda_{\Phi} \end{bmatrix} U_z^T, Z = [U_{z,X,U} U_{z,\text{res}}] \]

(64)

\[
U_{z,\text{res}} = \begin{bmatrix} U_{z,\text{res},1} \\ U_{z,\text{res},2} \end{bmatrix}, \Gamma_s H_{s,u} = 0 \implies \begin{bmatrix} \Gamma_s \\ 0 \end{bmatrix} = \begin{bmatrix} \Phi \\ 0 \end{bmatrix} \begin{bmatrix} \Phi \\ 0 \end{bmatrix}^T \]

(65)

\[
U_{z,\text{res}} \left( \frac{ZZ^T}{N} \right) U_{z,\text{res}} = \Lambda_{\Phi} = \begin{bmatrix} \Lambda_{\Phi} \end{bmatrix} \begin{bmatrix} \Lambda_{\Phi} \\ \end{bmatrix}^T \]

(66)

where \( \Lambda_{X,U} \) includes those singular values corresponding to the influence of input data set \( U \) on the process variables and hence being (significantly) larger than the singular values included in \( \Lambda_{\Phi} \) (Qin and Li [2001], Wang and Qin [2002]). The matrix \( U_{z,\text{res}} \left( \frac{ZZ^T}{N} \right) U_{z,\text{res}} \) can be considered as the normalized covariance matrix, as used in the PCA, and thus it is reasonable to use the PCA-algorithm or, as described in Subsection 3.1, the statistic

\[
T_z^2 = z^T(k)U_{z,\text{res}}\Lambda_{\Phi}^{-1}U_{z,\text{res}}^Tz(k) \]

(67)

\[
z(k) = [y^T(k) \cdots y^T(k+s) w^T(k) \cdots w^T(k+s)]
\]

and the threshold setting \( J_{th} = \chi^2_{\alpha}(s+1)(m+l) \) for PM-FD. Note that the statistic (67) contains (considerably) redundant information and is too conservative to perform an effective FD. To solve this problem, an observer-based solution can be applied (Ding et al. [2009c]). It has been proven that \( U_{z,\text{res},1} \) spans the parity space and \( U_{z,\text{res},1}^T H_{s,u} = -U_{z,\text{res},2}^T \) (Wang and Qin [2002]). Select the last \( m \) vectors from \( U_{z,\text{res}} \) and denote them by

\[\begin{bmatrix} \alpha_1 \alpha_2 \cdots \alpha_i \cdots \alpha_m \end{bmatrix} \in U_{z,\text{res}}, \alpha_1 \in U_{z,\text{res},1}, \beta_1 \in U_{z,\text{res},2}, i = 1, \ldots, m \]

Remember that each of them satisfies (Ding et al. [2009c])

\[
\alpha_i \Gamma_s = 0, \begin{bmatrix} B_z \\ D_z \end{bmatrix} = \begin{bmatrix} \alpha_i H_{s,u,0} \\ \vdots \\ \alpha_i H_{s,u,m} \end{bmatrix} = \begin{bmatrix} \beta_i,0 \\ \vdots \\ \beta_i,m \end{bmatrix}
\]

\[H_{s,u} = [H_{s,u,0} \cdots H_{s,u,m}], \begin{bmatrix} \beta_1,0 \cdots \beta_i,m \end{bmatrix}\]

A result, we are able to construct \( m \) DOs, as given in (25) and using the solution (56), which deliver \( m \) residual signals \( r_1(k), \ldots, r_m(k) \). Note that

\[
r(k) = \begin{bmatrix} r_1(k) \\ \vdots \\ r_m(k) \end{bmatrix} \sim N \left( \begin{bmatrix} \sigma_{11}^2 \\ \vdots \\ \sigma_{mm}^2 \end{bmatrix} \right)
\]

where \( \sigma_{11}, \ldots, \sigma_{mm} \) denote the last \( m \) (smallest) singular values in \( \Lambda_{\Phi} \). Thus, for the PM-FD purpose, for instance the following test statistic and threshold can be used

\[
T^2 = \sum_{i=1}^{m} \sigma_i^2 r_i^2(k), J_{th} = \chi^2_{\alpha}(m)
\]

which is less conservative than (67). This method is comparable with the standard PCA, in which the normalization step is replaced by the SVD (64) that serves removing the deterministic part in the system dynamics. For a dynamic system of the \( n \)-th order with a large \( n \), the dimension of (parity) vector \( \alpha_i \) being equal to \( m(s+1) > mn \), may become very high. Therefore, it it advising to use a recursive computation given in the form of a DO, as described above.

4.2 Adaptive and recursive computation issues

In the PCA, PLS and the above PCA-like observer-based PM-FD schemes, a basic assumption is that the covariance matrix is constant. In the industrial environment, the processes and plants often involve frequent changes in operating regimes, which lead to significant variations in the mean and (co-)variance of the process variables. A powerful tool to deal with such problems is the adaptive technique. Among the existing adaptive schemes, the recursive technique has been well developed and widely applied to the standard data-driven methods thanks to its reduced on-line computation. Since the early work by Helland et al. [1992], Qin [1998], numerous recursive schemes have been reported. While the recursive updates of the mean, variance and the associated normalization of the new observation are trivial (Li et al. [2000]), the research focus is mainly on the recursive computation of SVD which
is used both in the PCA and PLS. Li et al. [2000] proposed to apply the rank-one modification technique for a recursive updating of the (normalized) covariance matrix and its SVD used in the PCA. In order to reduce the needed on-line computations, recently Elshenawy et al. [2010] proposed two alternative recursive PCA (RPCA) schemes, one is based on the first-order perturbation (FOP) theory, which is a rank-one update of the eigenpairs of the data covariance matrix, and another one on the data-projection method (DPM) which serves as a simple approach for recursively updating the singular values and associated eigenvectors of interests. It is worth to mention that in Elshenawy et al. [2010] the standard PCA and different RPCA algorithms have been compared regarding to their online computation complexity. It is demonstrated that the DPM is the simplest algorithm if only a number of singular values and associated eigenvectors are needed for constructing the testing statistic.

Recently, Wang et al. [2005] pointed out numerous deficiencies in the existing RPCA algorithms due to ever-growing data set which leads to a reduction in the speed of adaptation. As a solution, they proposed the moving window PCA (MWPCA) scheme. He and Yang [2008] also proposed an MWPCA scheme with an improved RPCA algorithm. In both studies, significant improvement of the adaptation performance is demonstrated.

Recall our discussion in Subsections 3.1 and 3.2, we would like to recommend the use of the FOP method for the RPCA and RPLS. One important reason is that all singular values and the associated eigenvectors of the normalized covariance matrix are needed for an efficient and reliable FD. Note that the modified PLS proposed in Subsection 3.2 mainly deals with two SVDs. A combination of the MWPCA or MWPLS with the FOP-based RPCA or RPLS may provide us with promising PM-FD results.

Based on the similar idea behind the RPCA, Naik et al. [2010] proposed two alternative schemes for recursively updating parity vectors or PCA-like diagnostic observer respectively using the FOP and DPM algorithms for the needed SVD. They have reported satisfactory PM-FD results by applying the algorithms to a dynamic (benchmark) plant.

It is worth to mention that in the existing RPCA/RPLS schemes no attention has been paid to the convergence or stability issues. In particular, the latter plays a central role when a dynamic process is under consideration and a PCA-like diagnostic observer is applied. The well-established adaptive control and observer theory (Aström and Wittenmark [1995], Zhang [2002]) offers an alternative and powerful solution. Using this theory, Ding et al. [2009b] developed an adaptive PCA-like diagnostic observer whose stability and convergence speed can be proven. We are convincing that the combination of the data-driven methods and the adaptive control technique will lead to new successful solutions.

5. CONCLUSIONS

In this paper, we have first reviewed the basic data-driven and model-based PM-FD methods, their standard forms and some recent studies on them and their interconnections. Based on this work, we have proposed (a) modifications on the PCA and PLS (b) the integration of the Kalman filter and GLR schemes (c) the combined use of the observer-based and PS schemes. These modifications and the integrated schemes are useful or necessary for an efficient and reliable PM-FD, also for an application in the process satisfying all those ideal assumptions/conditions. Our survey study has been further dedicated to the application of these PM-FD methods in the real industrial environment. In order to solve the problems with process dynamics by using the data-driven methods, a PCA-like PS or DO solution is investigated, whose idea and the basic computation are similar to the PCA method. We have finally briefly surveyed the application of the adaptive/recursive techniques to the data-driven methods aiming at enhancing the system robustness against uncertainties in the industrial environment.

This survey report is based on our work in a number of industrial and research projects. All these basic PM-FD methods have been tested in their standard forms and on numerous industrial and academic benchmark processes and plants. Also the proposed modification schemes are implemented and in test. Due to the space limitation, these results cannot be summarized in this paper. We expect to include them in the final version of this report.

Acknowledgement: The authors would like to thank S. Yin, A. Haghani, A. Naik, Z. Hu, P. Deng, P. Tao, D. Schoch, T. Stargala, G. Gruss and H. Neumann for their valuable technical contributions in the relevant projects. The authors are also very grateful to the reviewers for their valuable comments.

REFERENCES


