

# Supervised Block-Aware Factorization Machine for Multi-Block Quality-Relevant Monitoring

Qinqin Zhu

*Department of Chemical Engineering, University of Waterloo, ON  
N2L 3G1 Canada (e-mail: qinqin.zhu@uwaterloo.ca)*

---

**Abstract:** Multi-block multivariate statistical methods have been developed to extract useful information from process and quality data in the era of big data, where process variables are partitioned into several meaningful blocks. However, most of these methods did not consider cross-correlations among divided blocks, which leads to inferior monitoring performance. In this article, a block-aware factorization machine (BAFM) algorithm is proposed to exploit information from process and quality data. In BAFM, quality data are first classified into normal and abnormal labels with principal component analysis based quality monitoring framework. Afterwards, a block number is attached to each process variable, and the interactions among different variables (both within and cross blocks) are learned through latent variables, which is supervised by the classified quality labels. Apart from the variable relation within the same block, BAFM also incorporates the block information; thus, both inner and cross correlations are constructed. The monitoring framework based on BAFM is developed, and its effectiveness and superiority are demonstrated through the Tennessee Eastman process.

*Keywords:* Quality-relevant monitoring, block-aware factorization machine, supervised learning, multi-block processes

---

## 1. INTRODUCTION

Process monitoring based on multivariate statistical methods is widely studied in both academia and industry to detect anomalies in process and quality data. Among them, principal component analysis (PCA), partial least squares (PLS) and canonical correlation analysis (CCA) are three popular algorithms (Qin (2012); Wise and Gallagher (1996)), where PCA can be applied for pure process monitoring or quality monitoring, while PLS and CCA are utilized for quality-relevant monitoring (Liu et al. (2018)). These algorithms and their variants have achieved varying degrees of success in medicine, economics, computer science, materials science, and other engineering fields.

With the advent of Industry 4.0, when collected data volumes increase explosively, it poses a large challenge on existing methods in terms of scalability and computation efficiency. Traditional PCA, PLS and CCA are not scalable for voluminous data, since the whole dataset needs to be fed into the models to extract latent structures. Therefore, their distributed or multi-block counterparts are developed to handle issues brought by big data (Ge (2017); Jiang et al. (2019)). In multi-block models, process variables are decomposed into different blocks based on prior knowledge or algorithms such as mutual information (Jiang and Yan (2014)). Then PCA, PLS and CCA are performed on each block to capture inner-correlations among variables inside the same blocks. Multi-block models work well to alleviate the burden of large volumes of data. However, cross-correlations remain untouched in these models.

In this article, a block-aware factorization machine (BAFM) algorithm is proposed to extract both inner and cross correlation from a new perspective, which is inspired by the idea of field-aware factorization machine (FFM) (Juan et al. (2016)). FFM plays an important role to predict click-through rate in advertising industry, where the concept of field is attached to features from advertisement and user sides to denote sources of these features. However, FFM cannot be applied directly in chemical processes, since in these processes (1) both process and quality variables are sampled continuously and with various sampling frequencies; (2) the number of blocks is usually larger than two; and (3) the samples are not sparse (Ranzato et al. (2008)). Instead, these issues will be addressed in BAFM.

In BAFM, quality variables are transformed to indicators or labels ( $\{0,1\}$ ) by performing PCA-based quality monitoring on them, where 0 stands for normal samples while 1 is for abnormal ones. Simultaneously, each process variable is attached with a block number based on prior knowledge. Then in BAFM, relations between process variables and classified quality labels are built through latent variables for each block. The cross-block information is also extracted by modeling the interactions of latent variables with different block numbers. The monitoring scheme is also established for BAFM, which is verified through the Tennessee Eastman process (TEP). It is noted that the application of BAFM is not limited to multi-block processes; it can be applied wherever sources of variables can provide extra information into the system modeling.

The remaining sections of the article are organized as follows. Section 2 reviews quality monitoring scheme based on PCA and describes how it is employed to convert quality variables into labels. In Section 3, details of BAFM algorithm are presented, including its motivation, modeling objective, solution derivation, optimization strategies, and monitoring scheme. TEP is employed to demonstrate the effectiveness of the proposed algorithm and its corresponding monitoring framework in Section 4. Finally, conclusions are drawn in the last section.

## 2. QUALITY MONITORING WITH PCA

Zhu and Qin (2019) classified monitoring schemes into three scenarios: inferential monitoring, quality-irrelevant process monitoring, and quality monitoring. Inferential monitoring, also referred to as quality-relevant monitoring, infers quality faults from process variables. Quality-irrelevant process monitoring detects anomalies in process variables that have no impact on quality data. Quality monitoring performs latent variable regression methods on quality data to monitor variations in quality spaces, which is an unsupervised monitoring scheme.

In BAFM, quality data need to be transformed into labels for further processing, where labels are defined as discrete integers attached to each sample. In this article, we use label 1 to denote abnormal samples, while using label 0 for normal scenarios, and quality monitoring with PCA is employed to categorize original quality data into labels.

### 2.1 Principal Component Analysis

Assuming that a data matrix  $\mathbf{Y} \in \mathbb{R}^{n \times p}$  is composed by  $n$  samples and each sample has  $p$  variables, PCA can be applied on  $\mathbf{Y}$  to extract its latent structure. Its objective is to maximize the variance of the extracted variables, which is represented as

$$\begin{aligned} \max_{\mathbf{p}} \quad & \mathbf{p}^\top \mathbf{Y}^\top \mathbf{Y} \mathbf{p} \\ \text{s.t.} \quad & \|\mathbf{p}\| = 1 \end{aligned} \quad (1)$$

where  $\mathbf{p} \in \mathbb{R}^p$  is the loading vector. Lagrange multipliers or nonlinear iterative partial least squares (NIPALS) algorithm can be used to derive the solution of Eq. (1) (Wold et al. (1987)).

After performing PCA,  $\mathbf{Y}$  is projected onto a lower-dimensional space defined by a small number of latent variables ( $\mathbf{t}_1, \dots, \mathbf{t}_l$ ), where  $l$  is the number of principal components. The mean-centered  $\mathbf{Y}$  can be formulated as

$$\mathbf{Y} = \mathbf{T}\mathbf{P}^\top + \mathbf{E} \quad (2)$$

where  $\mathbf{T} = [\mathbf{t}_1, \dots, \mathbf{t}_l]$  and  $\mathbf{P} = [\mathbf{p}_1, \dots, \mathbf{p}_l]$  are score matrix and loading matrix respectively, and  $\mathbf{E}$  is the residual.  $\mathbf{T}\mathbf{P}^\top$  comprises principal component subspace, which retains most information in the original data, while  $\mathbf{E}$  mainly contains noises.

For a new sample  $\mathbf{y}$ , the predicted and residual parts can be calculated by

$$\begin{aligned} \mathbf{t} &= \mathbf{P}^\top \mathbf{y} \\ \hat{\mathbf{y}} &= \mathbf{P}\mathbf{t} = \mathbf{P}\mathbf{P}^\top \mathbf{y} \\ \tilde{\mathbf{y}} &= (\mathbf{I} - \mathbf{P}\mathbf{P}^\top) \mathbf{y} \end{aligned} \quad (3)$$

Table 1. Monitoring statistics and control limits for PCA

Statistics	Calculation	Control limit
$T^2$	$\mathbf{t}^\top \mathbf{\Lambda}^{-1} \mathbf{t}$	$\frac{l(n^2-1)}{n(n-l)} F_{l, n-l; \alpha}$
$Q$	$\tilde{\mathbf{y}}^\top \tilde{\mathbf{y}}$	$g\chi_{h; \alpha}^2$

$\hat{\mathbf{y}}$  and  $\tilde{\mathbf{y}}$  reflect the variations in principal component and residual subspaces respectively.

### 2.2 Classification of Quality Data

In the literature, Hotelling's  $T^2$  and  $Q$  statistics (Jackson (2005); Joe Qin (2003); Nomikos and MacGregor (1995)) are widely used to monitor variations in principal component and residual subspaces in PCA, which can be formulated as

$$\begin{aligned} T^2 &= \mathbf{t}^\top \mathbf{\Lambda}^{-1} \mathbf{t} \\ Q &= \tilde{\mathbf{y}}^\top \tilde{\mathbf{y}} \end{aligned} \quad (4)$$

where  $\mathbf{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_l)$  contains nonzero eigenvalues of  $\mathbf{Y}^\top \mathbf{Y}$ , and  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_l > 0$ .

Assuming that the data are sampled from processes with a multivariate normal distribution or the collected data volumes are large enough, control limit of  $T^2$  can be obtained by an  $F$ -distribution, and control limit for  $Q$  is gained by a  $\chi^2$ -distribution (Nomikos and MacGregor (1995)). The monitoring indices and their corresponding control limits are summarized in Table 1. It is noted that  $\alpha$  is the confidence interval, and the calculations of  $g$  and  $h$  can be found in MacGregor et al. (1994).

In BAFM, the aforementioned PCA-based quality monitoring scheme is utilized to classify quality data as follows.

1. If  $T^2$  exceeds the control limit, then an anomaly is detected in the principal component subspace, which affects  $\mathbf{Y}$  and is classified as a fault. In this scenario, the sample is labeled as  $y = 1$ .
2. If  $Q$  exceeds the control limit, a potentially quality fault is detected with  $(1-\alpha) \times 100\%$ . Since the original relation among quality variables may be broken, the sample is also labeled as  $y = 1$  in this scenario.
3. If both  $T^2$  and  $Q$  are within their control limits, then the sample is regarded as normal, and its label  $y$  is 0.

The quality labels are denoted as  $\mathbf{y}_l \in \mathbb{R}^n$ , and they are used to supervise the training of BAFM described in the following section.

## 3. BLOCK-AWARE FACTORIZATION MACHINE

Logistic regression (LR) was employed in Jin et al. (2007) for quality prediction and control, and the optimization problem is formulated as

$$\min_{\mathbf{w}, b} - \sum_{i=1}^n \{y_i \log \sigma(g(\mathbf{x}_i)) + (1 - y_i) \log [1 - \sigma(g(\mathbf{x}_i))]\} \quad (5)$$

where  $y_i$  and  $\mathbf{x}_i$  are  $i^{\text{th}}$  classified quality label and process variables in  $\mathbf{y}_l$  and  $\mathbf{X} \in \mathbb{R}^{n \times m}$  respectively.  $n$  is total number of samples,  $m$  is number of process variables, and  $g(\mathbf{x})$  is a linear model for LR, which is defined as

$$g_{\text{LR}}(\mathbf{x}) = \mathbf{w}^\top \mathbf{x} + b \quad (6)$$

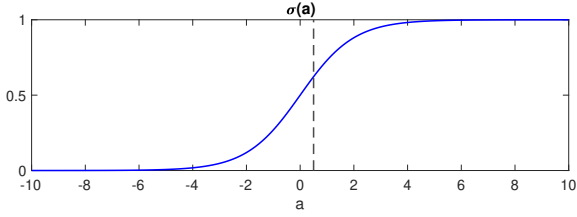


Fig. 1. Sigmoid function

where  $\mathbf{w} \in \mathbb{R}^m$  and  $b \in \mathbb{R}$  are weights and bias for logistic regression respectively.  $\sigma$  is the sigmoid function,

$$\sigma(a) = \frac{1}{1 + e^{-a}} \quad (7)$$

which is depicted in Fig. 1. It is observed that the output ranges of sigmoid function are  $[0, 1]$ .

In Eq. (5), LR can be applied in large-scale industrial processes, since the samples can be trained sequentially. The weight vector  $\mathbf{w}$  is learned for each process variable in Eq. (6). However, the sources of variables are not incorporated into Eq. (6), which might lead to sub-optimal performance. The same issue occurs in some other classification models such as factorization machines (Rendle (2010)). In order to extract valuable information from sources of process variables, block-aware factorization machine (BAFM) method is proposed in this section, which is adapted from FFM in advertising industry (Juan et al. (2016)).

### 3.1 BAFM Objective

Different from Eq. (6), BAFM assigns a block number to each process variable, and instead of a weight for each process variable, the weight is trained for each (block, process variable) pair. BAFM model can be expressed as

$$g_{\text{BAFM}}(\mathbf{x}) = \sum_{i_1=1}^m \sum_{i_2=i_1+1}^m (\mathbf{w}_{i_1, f_2} \cdot \mathbf{w}_{i_2, f_1}) x_{i_1, f_1} x_{i_2, f_2} + b \quad (8)$$

where  $m$  is number of process variables,  $\mathbf{w}_{i, f} \in \mathbb{R}^k$  is the weight or latent variable for  $i^{\text{th}}$  variable in field  $f$ ,  $b$  is the bias term,  $k$  is the length of latent variables, and  $x_{i, f}$  is the value of  $i^{\text{th}}$  variable in  $\mathbf{x}$ , which belongs to block  $f$ . In Eq. (8), cross-block correlation is incorporated by learning the interaction weights ( $\mathbf{w}_{i_1, f_2} \cdot \mathbf{w}_{i_2, f_1}$ ) when process variables are from different blocks ( $f_1 \neq f_2$ ), and the inner-block correlation is extracted when  $x_{i_1, f_1}$  and  $x_{i_2, f_2}$  are from the same block.

The objective function of BAFM can be derived by substituting Eq. (8) into Eq. (5). Additionally, to prevent ill-conditioned solutions, a regularization term is also considered. Therefore, the optimization problem of BAFM can be represented as

$$\min_{\mathbf{w}, b} \mathcal{E} = - \sum_{i=1}^n \{y_i \log \sigma(g_{\text{BAFM}}(\mathbf{x}_i)) + (1 - y_i) \log [1 - \sigma(g_{\text{BAFM}}(\mathbf{x}_i))]\} + \lambda \|\mathbf{w}\|^2 \quad (9)$$

where  $\lambda$  is regularized parameter, and it can also mitigate overfitting issues. Eq. (9) is also referred to as loss or cost function of BAFM.

### 3.2 Derivation of BAFM

Similar to LR, gradient descent can be employed to derive the solution of BAFM (Menard (2002)), where  $\mathbf{w}$  and  $b$  are calculated iteratively. Taking derivative with respect to  $\mathbf{w}_{i_1, f_2}$ ,  $\mathbf{w}_{i_2, f_1}$  and  $b$  in Eq. (9) leads to

$$\mathbf{g}_{w_{i_1, f_2}} = \sum_{j=1}^n (\sigma_j - y_j) \mathbf{w}_{i_2, f_1} x_{i_1, f_1} x_{i_2, f_2} + \lambda \mathbf{w}_{i_1, f_2} \quad (10)$$

$$\mathbf{g}_{w_{i_2, f_1}} = \sum_{j=1}^n (\sigma_j - y_j) \mathbf{w}_{i_1, f_2} x_{i_1, f_1} x_{i_2, f_2} + \lambda \mathbf{w}_{i_2, f_1} \quad (11)$$

$$g_b = \sum_{j=1}^n (\sigma_j - y_j) \quad (12)$$

where  $\mathbf{g}_{w_{i_1, f_2}} \equiv \frac{\partial \mathcal{E}}{\partial \mathbf{w}_{i_1, f_2}}$ ,  $\mathbf{g}_{w_{i_2, f_1}} \equiv \frac{\partial \mathcal{E}}{\partial \mathbf{w}_{i_2, f_1}}$ ,  $g_b \equiv \frac{\partial \mathcal{E}}{\partial b}$ , and  $\sigma_j$  stands for  $\sigma(g_{\text{BAFM}}(\mathbf{x}_j))$ . Therefore,  $\mathbf{w}_{i_1, f_2}$ ,  $\mathbf{w}_{i_2, f_1}$  and  $b$  can be updated by following the negative gradient to minimize the loss function in Eq. (9).

$$\mathbf{w}_{i_1, f_2}^{(t+1)} \leftarrow \mathbf{w}_{i_1, f_2}^{(t)} - \eta_w \mathbf{g}_{w_{i_1, f_2}} \quad (13)$$

$$\mathbf{w}_{i_2, f_1}^{(t+1)} \leftarrow \mathbf{w}_{i_2, f_1}^{(t)} - \eta_w \mathbf{g}_{w_{i_2, f_1}} \quad (14)$$

$$b^{(t+1)} \leftarrow b^{(t)} - \eta_b g_b \quad (15)$$

where  $\eta_w$  and  $\eta_b$  are learning rates for weights and bias, respectively.

### 3.3 Adaptive Gradient Algorithm

It is noted that in Eqs. (10) - (15), the gradients of weights  $\mathbf{w}_{i_1, f_2}$  and  $\mathbf{w}_{i_2, f_1}$  and bias  $b$  need to be calculated over the whole dataset for a single round, which is quite inefficient for large volumes. Therefore, stochastic gradient methods (Bottou (2010)) are preferable instead for voluminous samples, where only a sample is employed to update  $\mathbf{w}_{i_1, f_2}$ ,  $\mathbf{w}_{i_2, f_1}$  and  $b$  each time in Eqs. (10) - (12) as follows.

$$\mathbf{g}_{w_{i_1, f_2}} = (\sigma_j - y_j) \mathbf{w}_{i_2, f_1} x_{i_1, f_1} x_{i_2, f_2} + \lambda \mathbf{w}_{i_1, f_2} \quad (16)$$

$$\mathbf{g}_{w_{i_2, f_1}} = (\sigma_j - y_j) \mathbf{w}_{i_1, f_2} x_{i_1, f_1} x_{i_2, f_2} + \lambda \mathbf{w}_{i_2, f_1} \quad (17)$$

$$g_b = \sigma_j - y_j \quad (18)$$

Additionally, selection of learning rates is important: if  $\eta_w$  and  $\eta_b$  are too small, it will take longer time to converge; if their values are too large, the algorithms tend to diverge. We adopt the adaptive gradient method *AdaGrad* (Duchi et al. (2011)) in BAFM to mitigate the sensitivity of the algorithm to various learning rates. In *AdaGrad*, the values of learning rates decrease with the number of considered samples, where the squared of gradients are accumulated as follows.

$$G_{w_{i_1, f_2}}^{(t+1)} \leftarrow G_{w_{i_1, f_2}}^{(t)} + \mathbf{g}_{w_{i_1, f_2}}^{(t)} \cdot \mathbf{g}_{w_{i_1, f_2}}^{(t)} \quad (19)$$

$$G_{w_{i_2, f_1}}^{(t+1)} \leftarrow G_{w_{i_2, f_1}}^{(t)} + \mathbf{g}_{w_{i_2, f_1}}^{(t)} \cdot \mathbf{g}_{w_{i_2, f_1}}^{(t)} \quad (20)$$

$$B^{(t+1)} \leftarrow B^{(t)} + g_b^{(t)} \cdot g_b^{(t)} \quad (21)$$

Then the weights and bias are updated by

$$\mathbf{w}_{i_1, f_2}^{(t+1)} \leftarrow \mathbf{w}_{i_1, f_2}^{(t)} - \frac{\eta_w 0}{\sqrt{G_{w_{i_1, f_2}}^{(t+1)}}} \mathbf{g}_{w_{i_1, f_2}} \quad (22)$$

$$\mathbf{w}_{i_2, f_1}^{(t+1)} \leftarrow \mathbf{w}_{i_2, f_1}^{(t)} - \frac{\eta_w 0}{\sqrt{G_{w_{i_2, f_1}}^{(t+1)}}} \mathbf{g}_{w_{i_2, f_1}} \quad (23)$$

---

**Algorithm 1** Block-Aware Factorization Machine

---

1. Scale  $\mathbf{X}$  and  $\mathbf{Y}$  to zero mean and unit variance.
  2. Transform  $\mathbf{Y}$  to quality labels  $\mathbf{y}_l$  with PCA-based quality monitoring.
  2. Choose the parameters: initial learning rates  $\eta_{w0}$  and  $\eta_{b0}$ , regularized parameter  $\lambda$ , dimension of latent variables  $k$ , and iterations  $I$ .
  3. Initialize
    - $G^{(0)} \in \mathbb{R}^{m \times f}$  with each element equal to 1;
    - $B^{(0)} = 1$ ;
    - $\mathbf{w}^{(0)} \in \mathbb{R}^{m \times f \times k}$ , where each element follows the uniform distribution  $\mathbf{U}(0, \frac{1}{\sqrt{k}})$ ;
    - $b^{(0)}$  follows uniform distribution  $\mathbf{U}(0, 1)$ .
 where  $m$  is total number of process variables, and  $f$  is total number of blocks.
  4. For each pair of variables  $(x_{i_1, f_1}, x_{i_2, f_2})$  in sample  $\mathbf{x}$  and its corresponding quality label  $y$ ,
    - Calculate gradients  $\mathbf{g}_{w_{i_1, f_2}}$ ,  $\mathbf{g}_{w_{i_2, f_1}}$  and  $g_b$  with Eqs. (16) - (18);
    - Accumulate sum of squared gradients  $G_{w_{i_1, f_2}}^{(t+1)}$ ,  $G_{w_{i_2, f_1}}^{(t+1)}$  and  $B^{(t+1)}$  with Eqs. (19) - (21);
    - Update weights and bias with Eqs. (22) - (24).
 Repeat the above steps for each sample in collected process data  $\mathbf{X}$  and classified quality labels  $\mathbf{y}_l$ .
  5. Repeat Step 4 for  $I$  iterations or until both  $\mathbf{w}$  and  $b$  converge.
- 

$$b^{(t+1)} \leftarrow b^{(t)} - \frac{\eta_{b0}}{\sqrt{B^{(t+1)}}} g_b \quad (24)$$

where  $\eta_{w0}$  and  $\eta_{b0}$  are initial learning rates. Since  $G_{w_{i_1, f_2}}^{(t+1)}$ ,  $G_{w_{i_2, f_1}}^{(t+1)}$  and  $B^{(t+1)}$  are monotonically increasing, the actual learning rates  $\frac{\eta_{w0}}{\sqrt{G_{i_1, f_2}^{(t+1)}}}$ ,  $\frac{\eta_{w0}}{\sqrt{G_{i_2, f_1}^{(t+1)}}}$  and  $\frac{\eta_{b0}}{\sqrt{B^{(t+1)}}}$  are

decreasing with the number of processed samples. Thus, in the later training stage, the learning rates are close to zero, which is beneficial for the convergence of the algorithm.

### 3.4 BAFM Algorithm

BAFM algorithm is summarized in Algorithm 1. The following are several notes regarding to the parameter initialization in Algorithm 1.

1. The initial values of elements in  $G^{(0)}$  and  $B^{(0)}$  are set to 1 to improve the robustness of BAFM when  $\sum_t \mathbf{g}_{w_{i, f}}^{(t)} \cdot \mathbf{g}_{w_{i, f}}^{(t)}$  or  $\sum_t g_b^{(t)} \cdot g_b^{(t)}$  is close to zero.
2. The parameters  $\eta_{w0}$ ,  $\eta_{b0}$  and  $\lambda$  is selected through cross validation (Zhu et al. (2017)), and the dimension of latent variables  $k$  is determined by the data volume.

As shown in Algorithm 1, BAFM can capture both inner-block and cross-block correlations: when  $x_{i_1, f_1}$  and  $x_{i_2, f_2}$  come from the same block, where  $f_1 = f_2$ , then Eqs. (22) - (23) update latent variables belonging to the same block. When  $x_{i_1, f_1}$  and  $x_{i_2, f_2}$  are from different blocks ( $f_1 \neq f_2$ ),  $\mathbf{w}_{i_1, f_2}$  and  $\mathbf{w}_{i_2, f_1}$  are updated respectively. Additionally, since the sample  $(\mathbf{x}, y)$  is trained sequentially, BAFM is capable to process large volumes of datasets.

### 3.5 Monitoring Scheme

In Algorithm 1, latent variables  $\mathbf{w}$  and  $b$  are learned for BAFM. Given a new sample  $\mathbf{x}$ , the predicted quality data

Table 2. Blocks of process variables

Blocks	XMEAS	XMV
1	1, 2, 3, 4	1, 2, 3, 4
2	5, 6, 7, 8, 9, 21	10, 11
3	10, 11, 12, 13, 14, 20, 22	5, 6, 7
4	15, 16, 17, 18, 19	8, 9

is obtained as follows.

$$\hat{y} \equiv \sigma(g_{\text{BAFM}}(\mathbf{x})) = \sigma \left[ \sum_{i_1=1}^m \sum_{i_2=i_1+1}^m (\mathbf{w}_{i_1, f_2} \cdot \mathbf{w}_{i_2, f_1}) x_{i_1, f_1} x_{i_2, f_2} + b \right] \quad (25)$$

where  $\hat{y} \in [0, 1]$  as shown in Fig. 1.

For quality-relevant monitoring scheme based on BAFM, a threshold 0.5 is set: when  $\hat{y}$  is larger than or equal to 0.5, then the sample  $\mathbf{x}$  is classified as a quality-relevant fault; otherwise, the process is regarded as normal.

## 4. TENNESSEE EASTMAN PROCESS

Tennessee Eastman process (TEP) is a benchmark process in process systems engineering to demonstrate the effectiveness of control schemes and multivariate statistical analysis algorithms (Downs and Vogel (1993)). TEP produces  $G$  and  $H$  as main products, and  $F$  as byproduct, from reactants  $A$ ,  $C$ ,  $D$  and  $E$ . There are five main components in the process (i.e., reactor, condenser, stripper, compressor and separator), and two sets of variables are collected: manipulated variables (XMV(1-12)) and process measurements (XMEAS(1-41)). The detailed description of these variables can be found in Downs and Vogel (1993).

In this case study, XMEAS(1-22) and XMV(1-11) are selected as process variables, and XMEAS(35-36) are for quality variables. Prior knowledge is used to partition the process variables into sub-blocks in this article, which is summarized in Table 2. It is noted that variables in Block 1 are from input feed streams, Block 2 is for reactor and condenser, Block 3 is for separator and compressor, and Block 4 is for Stripper.

Downs and Vogel (1993) simulates one normal scenario (IDV(0)) and 15 known disturbances (IDV(1-15)), and each scenario has training and test datasets. In order to realize full potentials of BAFM, it is better to have balanced normal and abnormal samples in the training dataset. Therefore, the following training strategy is proposed.

- i. Pre-process TEP data to remove samples without corresponding quality values due to inconsistent sampling rates (Zhu et al. (2016)).
- ii. Perform PCA on normal quality data from IDV(0) to construct quality monitoring framework, including PCA model structure and control limits.
- iii. Employ the monitoring scheme to classify quality data from all training samples in IDV(0-15).
- iv. Conduct BAFM on the classified quality labels and their corresponding process variables from IDV(0-15).

In this section, multi-block PLS (MBPLS) (Choi and Lee (2005)), LR and BAFM based monitoring schemes are performed on the selected process and quality variables to compare their monitoring performance. Their parameters

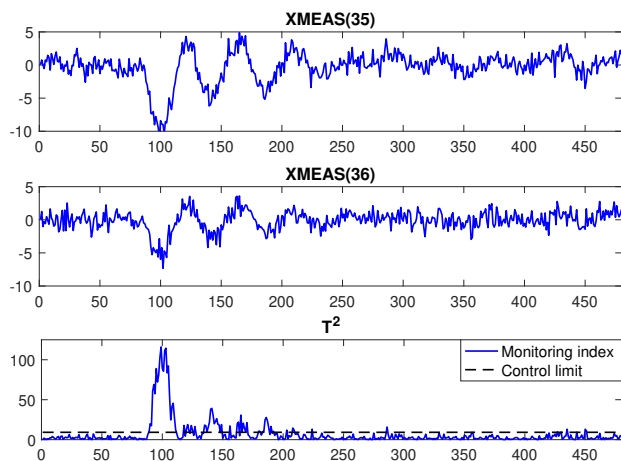


Fig. 2. Quality variations and quality monitoring results

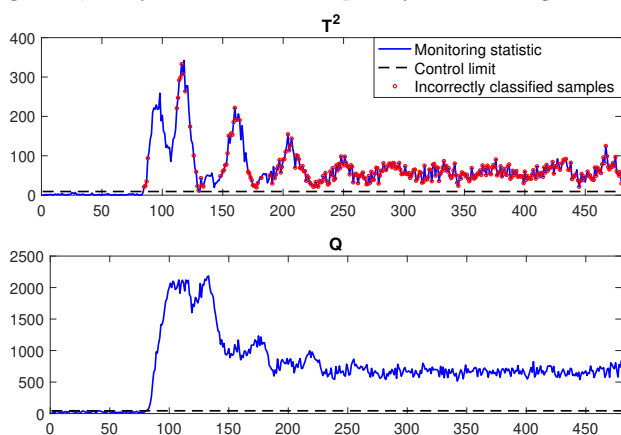


Fig. 3. Monitoring results with MBPLS in IDV(1)

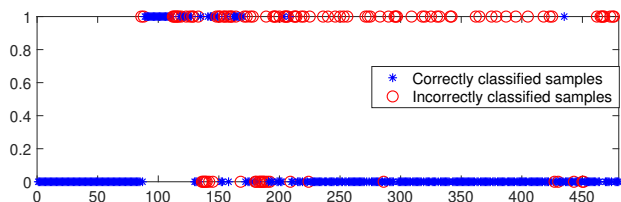


Fig. 4. Monitoring results with LR in IDV(1)

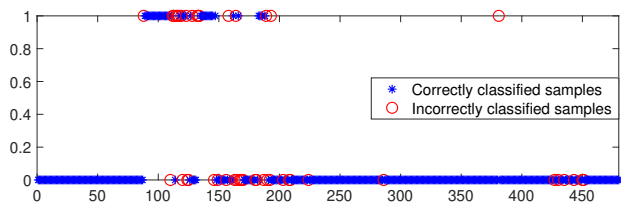


Fig. 5. Monitoring results with BAFM in IDV(1)

are selected with cross-validation: In MBPLS, the number of principal components  $l = 2$ ; In LR,  $\eta = 0.39$  and  $\lambda = 0.0001$ ; In BAFM,  $\eta_{w0} = 0.44$ ,  $\eta_{b0} = 0.07$ ,  $\lambda = 0.0001$ ,  $f = 4$  (total number of blocks) and  $k = 4$ . In LR and BAFM, the number of iteration is set to 100. IDV(1) and IDV(14) are utilized to illustrate their performance.

*IDV(1) - a step change in A/C feed ratio:* In IDV(1), a step change occurs in A/C feed ratio while B composition is constant. The variations of quality variables and quality

Table 3. FDR and FAR of MBPLS, LR and BAFM (%)

	IDV(1)		IDV(14)	
	FDR	FAR	FDR	FAR
MBPLS	100.00	79.64	71.43	69.98
LR	64.86	20.44	14.29	6.98
BAFM	80.81	4.18	0.00	0.21

monitoring results are presented in Fig. 2, which indicates that the disturbance in IDV(1) affects quality variables transiently. Figs. 3 - 5 show the monitoring performance of MBPLS, LR and BAFM, respectively, and Table 3 summarizes their fault detection rates (FDR) and false alarm rates (FAR). In Figs. 3 and 4, both MBPLS and LR fail to present the transient disturbance and keep raising the false alarms after Sample 200, which can also be observed from their high false alarm rates in Table 3. In contrast, BAFM successfully captures the trend: alarms are raised only for Samples 90 - 200, while others are classified as normal. In BAFM, the blocks of each variable provides extra information about the process, and thus it achieves the best performance with high fault detection rates and the lowest false alarm rates.

*IDV(14) - sticking in reactor cooling water valve:* The reactor cooling water valve is sticking in IDV(14), and as shown in Fig. 6, the disturbance does not affect quality variables. However, in Fig. 7, the  $T^2$  monitoring index of MBPLS exceeds its control limit incorrectly with a high false alarm rate 69.98%. In contrast, both LR and BAFM have lower false alarm rates with BAFM performs the best as shown in Figs. 8 - 9 and Table 3. It is noted that in this scenario the fault detection rates of LR and BAFM are very low, which is expected due to the quality-irrelevant nature of the disturbance.

## 5. CONCLUSIONS

In this article, a block-aware factorization machine based monitoring scheme is proposed for supervised learning and quality-relevant monitoring. In BAFM, continuous quality variables are classified into two categories (normal and abnormal ones) with PCA-based quality monitoring method. Process variables are attached with a block number to indicate their sources. Then interactions between classified quality labels and process variables are extracted, including both inner-block and cross-block relations. The case study with Tennessee Eastman process demonstrates the superiority of BAFM-based monitoring method over multi-block partial least squares and logistic regression algorithms.

## ACKNOWLEDGEMENTS

The author would like to acknowledge the financial support provided by the Chemical Engineering Department at the University of Waterloo.

## REFERENCES

Bottou, L. (2010). Large-scale machine learning with stochastic gradient descent. In *Proceedings of COMPSTAT'2010*, 177-186. Springer.

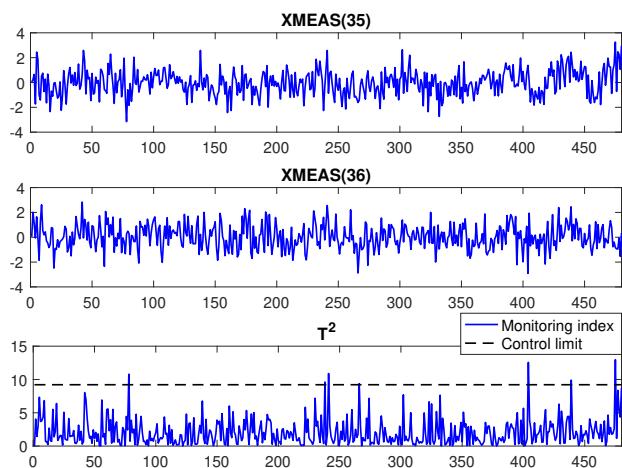


Fig. 6. Quality variations and quality monitoring results

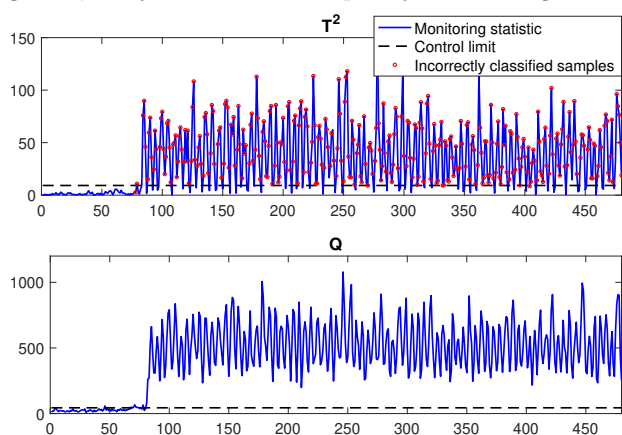


Fig. 7. Monitoring results with MBPLS in IDV(14)

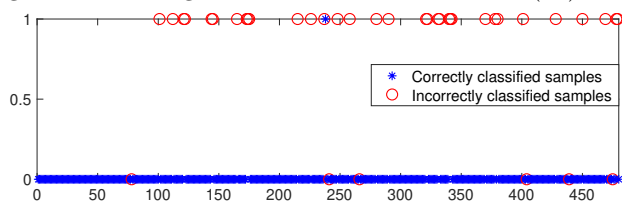


Fig. 8. Monitoring results with LR in IDV(14)

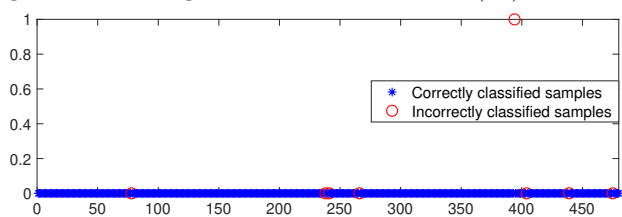


Fig. 9. Monitoring results with BAFM in IDV(14)

Choi, S.W. and Lee, I.B. (2005). Multiblock pls-based localized process diagnosis. *Journal of Process Control*, 15(3), 295–306.

Downs, J.J. and Vogel, E.F. (1993). A plant-wide industrial process control problem. *Computers & chemical engineering*, 17(3), 245–255.

Duchi, J., Hazan, E., and Singer, Y. (2011). Adaptive subgradient methods for online learning and stochastic optimization. *Journal of Machine Learning Research*,

12(Jul), 2121–2159.

Ge, Z. (2017). Review on data-driven modeling and monitoring for plant-wide industrial processes. *Chemometrics and Intelligent Laboratory Systems*, 171, 16–25.

Jackson, J.E. (2005). *A user's guide to principal components*, volume 587. John Wiley & Sons.

Jiang, Q. and Yan, X. (2014). Plant-wide process monitoring based on mutual information–multiblock principal component analysis. *ISA transactions*, 53(5), 1516–1527.

Jiang, Q., Yan, X., and Huang, B. (2019). Review and perspectives of data-driven distributed monitoring for industrial plant-wide processes. *Industrial & Engineering Chemistry Research*, 58(29), 12899–12912.

Jin, R., Li, J., and Shi, J. (2007). Quality prediction and control in rolling processes using logistic regression. *Transactions of NAMRI/SME*, 35, 113–120.

Joe Qin, S. (2003). Statistical process monitoring: basics and beyond. *Journal of Chemometrics: A Journal of the Chemometrics Society*, 17(8-9), 480–502.

Juan, Y., Zhuang, Y., Chin, W.S., and Lin, C.J. (2016). Field-aware factorization machines for CTR prediction. In *Proceedings of the 10th ACM Conference on Recommender Systems*, 43–50. ACM.

Liu, Q., Zhu, Q., Qin, S.J., and Chai, T. (2018). Dynamic concurrent kernel CCA for strip-thickness relevant fault diagnosis of continuous annealing processes. *Journal of Process Control*, 67, 12–22.

MacGregor, J.F., Jaeckle, C., Kiparissides, C., and Koutoudi, M. (1994). Process monitoring and diagnosis by multiblock PLS methods. *AIChE Journal*, 40(5), 826–838.

Menard, S. (2002). *Applied logistic regression analysis*, volume 106. Sage.

Nomikos, P. and MacGregor, J.F. (1995). Multivariate SPC charts for monitoring batch processes. *Technometrics*, 37(1), 41–59.

Qin, S.J. (2012). Survey on data-driven industrial process monitoring and diagnosis. *Annual reviews in control*, 36(2), 220–234.

Ranzato, M., Boureau, Y.L., and Cun, Y.L. (2008). Sparse feature learning for deep belief networks. In *Advances in neural information processing systems*, 1185–1192.

Rendle, S. (2010). Factorization machines. In *2010 IEEE International Conference on Data Mining*, 995–1000. IEEE.

Wise, B.M. and Gallagher, N.B. (1996). The process chemometrics approach to process monitoring and fault detection. *Journal of Process Control*, 6(6), 329–348.

Wold, S., Esbensen, K., and Geladi, P. (1987). Principal component analysis. *Chemometrics and intelligent laboratory systems*, 2(1-3), 37–52.

Zhu, Q., Liu, Q., and Qin, S.J. (2016). Concurrent canonical correlation analysis modeling for quality-relevant monitoring. *IFAC-PapersOnLine*, 49(7), 1044–1049.

Zhu, Q., Liu, Q., and Qin, S.J. (2017). Concurrent quality and process monitoring with canonical correlation analysis. *Journal of Process Control*, 60, 95–103.

Zhu, Q. and Qin, S.J. (2019). Supervised diagnosis of quality and process faults with canonical correlation analysis. *Industrial & Engineering Chemistry Research*.