Preprints of the 21st IFAC World Congress (Virtual)
Berlin, Germany, July 12-17, 2020

Dynamic Weighted Canonical Correlation Analysis for Auto-Regressive Modeling

Qinqin Zhu * **Qiang Liu** ** **S. Joe Qin** ***

* *Department of Chemical Engineering, University of Waterloo, ON N2L 3G1 Canada (e-mail: qinqin.zhu@uwaterloo.ca)*
** *State Key Laboratory of Synthetical Automation for Process Industries, Northeastern University, Liaoning 110819, China*
*** *School of Data Science, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong*

Abstract: Canonical correlation analysis (CCA) is widely used as a supervised learning method to extract correlations between process and quality datasets. When used to extract relations between current data and historical data, CCA can also be regarded as an auto-regressive modeling method to capture dynamics. Various dynamic CCA algorithms were developed in the literature. However, these algorithms do not consider strong dependence existing in adjacent samples, which may lead to unnecessarily large time lags and inaccurate estimation of current values from historical data. In this paper, a dynamic weighted CCA (DWCCA) algorithm is proposed to address this issue with a series of polynomial basis functions. DWCCA extracts dynamic relations by maximizing correlations between current data and a weighted representation of past data, and the weights rely only on a limited number of polynomial functions, which removes the negative effect caused by strongly collinear neighboring samples. After all the dynamics are exploited, static principal component analysis is then employed to further explore the cross-correlations in the dataset. The Tennessee Eastman process is utilized to demonstrate the effectiveness of the proposed DWCCA method in terms of prediction efficiency and collinearity handling.

Keywords: dynamic weighted canonical correlation analysis, auto-regressive modeling, principal component analysis

1. INTRODUCTION

Due to complicated mechanisms and large number of control loops involved in modern industrial processes, the collected data are usually highly auto-correlated and cross-correlated, where auto-correlations indicate the dynamic relations between samples and their historical data points, while cross-correlations reflect the interaction between different variables. Static multivariate statistical methods are widely used to exploit cross-correlations from process and quality variables. Among them, principal component analysis (PCA) is preferred to extract static relations between process or quality data, while supervised learning methods, such as partial least squares (PLS) and canonical correlation analysis (CCA), build latent models with the supervision of quality data (Jackson (2005); MacGregor and Kourti (1995); Qin (2012); Yin et al. (2014)). Improved versions of these algorithms were also widely studied, such as nonlinear ones (Liu et al. (2017); Zhu et al. (2017b)) and robust ones (Ge et al. (2008)).

However, these algorithms cannot capture dynamic characteristics of industrial processes, where sampled data points have dependence on past data. Thus, their dynamic counterparts also receive high attentions in both academia and industry. A straightforward method was designed by Ku et al. (1995) by aggregating historical data in a single matrix and performing static PCA on this matrix. In their work, the extracted latent variables are a mixture of static and dynamic variations, making it difficult to develop subsequent monitoring and diagnosis frameworks. Furthermore, it may require more latent variables than original process variables. Recently, Li et al. (2014) and Dong and Qin (2018a) proposed advanced dynamic modeling algorithms, where dynamic structures are constructed by maximizing the relations between current data and a weighted combination of historical data. Compared with Li et al. (2014)'s method, dynamic-inner models developed by Dong and Qin (2018a) obtain explicit dynamic inner and outer modeling objectives, which have demonstrated their superiority in both numerical simulations and industrial processes. It is noted that these proposed dynamic techniques can be applied to other multivariate statistical methods as well.

In industrial processes, due to high sampling frequencies or relatively slow continuous process changes, collected data usually contain strong dependence in successive samples. This requires extra lagged variables to represent historical effects, and the extracted correlations might be artificially large with redundant information. Thus, addressing the collinear issue in auto-regressive models is important for effective dynamic modeling. However, the existing dynamic algorithms pay no attention to dealing with this issue, leading to sub-optimal performance.

In this article, a dynamic weighted CCA (DWCCA) method is proposed to capture latent dynamic relations by maximizing correlations between present and past data with a weighted representation. To overcome the collinearity issue in sampled variables, a series of basis functions, polynomial functions, are designed. In DWCCA, each weight of historical sample is represented by a linear combination of these basis functions, which effectively reduces the required number of lagged variables. After exploiting the dynamic variations, a static PCA is performed on the residual dataset to separate static variations from noise. Tennessee Eastman process (TEP) is employed to demonstrate the advantages of DWCCA over dynamic-inner CCA (DiCCA) proposed recently by Dong and Qin (2018a), which has shown its superiority over other algorithms. It is noted that the techniques developed for DWCCA are also applicable to other algorithms to alleviate the negative effects brought by collinear adjacent samples.

The remainder of the article is organized as follows. In Section 2, CCA algorithm is reviewed, in terms of both static supervised learning method and dynamic time series method. Section 3 proposes the DWCCA algorithm, with details on selection of basis functions, formulation of objective, derivation of solution, and comprehensive procedure of DWCCA. The case study on TEP is studied in Section 4 to compare the prediction performance of DWCCA and DiCCA. Conclusions and future works are summarized in the final section.

2. CANONICAL CORRELATION ANALYSIS

2.1 Static Supervised Learning Method

CCA was proposed by Hotelling (1936) to exploit correlations between two datasets $\mathbf{X} \in \mathbb{R}^{n\times m}$ and $\mathbf{Y} \in \mathbb{R}^{n\times p}$, where n is number of collected sample, and m and p are number of process variables and quality variables, respectively. The mathematical formulation of CCA is

$$\begin{array}{c} \max \; J = \mathbf{w}^\top \mathbf{X}^\top \mathbf{Y}\mathbf{c} \\ \text{s.t. } ||\mathbf{X}\mathbf{w}|| = 1, ||\mathbf{Y}\mathbf{c}|| = 1 \end{array} \tag{1}$$

where $\mathbf{w} \in \mathbb{R}^m$ and $\mathbf{c} \in \mathbb{R}^p$ are weighting vectors for $\mathbf{X}$ and $\mathbf{Y}$, respectively. The latent variables can be extracted iteratively, and the detailed description of CCA can be found in Hotelling (1936). To improve its robustness to noise in strong collinear scenarios, Zhu et al. (2016) added two regularization terms in both constraints in Eq. (1).

After performing CCA on $\mathbf{X}$ and $\mathbf{Y}$, the original datasets can be decomposed as

$$\begin{cases} \mathbf{X} = \mathbf{T}\mathbf{P}^\top + \mathbf{E} \\ \mathbf{Y} = \mathbf{T}\mathbf{Q}^\top + \mathbf{F} \end{cases} \tag{2}$$

where $\mathbf{T} \in \mathbb{R}^{n\times l}$ is score matrix, $\mathbf{P} \in \mathbb{R}^{m\times l}$ and $\mathbf{Q} \in \mathbb{R}^{p\times l}$ are loading matrices, and $\mathbf{E} \in \mathbb{R}^{n\times m}$ and $\mathbf{F} \in \mathbb{R}^{n\times p}$ are residuals for $\mathbf{X}$ and $\mathbf{Y}$ respectively. l is number of latent variables, which can be determined with cross validation.

For a new sample $\mathbf{x}$, the quality data $\mathbf{y}$ can be predicted by $\mathbf{x}$ directly, which is represented as

$$\hat{\mathbf{y}} = \mathbf{Q}\mathbf{t} = \mathbf{Q}\mathbf{R}^\top \mathbf{x} \tag{3}$$

where $\mathbf{R} \in \mathbb{R}^{m\times l} = \mathbf{W}(\mathbf{P}^\top \mathbf{W})^{-1}$, and $\mathbf{W} \in \mathbb{R}^{m\times l}$.

2.2 Dynamic Time Series Method

It is noted that in Eq. (3), CCA is used as a supervised learning method to extract static relations between process and quality variables with the supervision of quality data. In addition, CCA can also be utilized as a time series method to build dynamic relations within one dataset.

For ease of representation, we re-denote the dimension of process data as $\mathbf{X} \in \mathbb{R}^{(N+s+1)\times m} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{s+N+1}]^\top$, where $n = N + s + 1$ is total number of samples, and s is the distributed lag to represent the dynamic order of the system. In dynamic processes with one latent factor, we assume that the current latent variable $t_k = \mathbf{w}^\top \mathbf{x}_k$ is a weighted sum of past scores $[t_{k-1}, t_{k-2}, \ldots, t_{k-s}]$ by

$$t_k = \beta_1 t_{k-1} + \beta_2 t_{k-2} + \ldots + \beta_s t_{k-s} + \tilde{t}_k \tag{4}$$

where $\boldsymbol{\beta} \in \mathbb{R}^s = [\beta_1, \beta_2, \ldots, \beta_s]^\top$ is a vector of weighted parameters, and $\tilde{t}_k$ is the modeling error containing static variations and noise.

The following matrices are formulated in order to extract dynamic relations within $\mathbf{X}$.

$$\begin{array}{l} \mathbf{X}_i = [\mathbf{x}_i, \mathbf{x}_{i+1}, \ldots, \mathbf{x}_{i+N}]^\top, i \in \{1, 2, \ldots, s+1\} \\ \mathbf{Z}_s = [\mathbf{X}_s, \mathbf{X}_{s-1}, \ldots, \mathbf{X}_1] \end{array} \tag{5}$$

Then the objective of dynamic CCA is

$$\begin{array}{c} \max \; J = \mathbf{w}^\top \mathbf{X}_{s+1}^\top \mathbf{Z}_s (\boldsymbol{\beta} \otimes \mathbf{w}) \\ \text{s.t. } \mathbf{w}^\top \mathbf{X}_{s+1}^\top \mathbf{X}_{s+1} \mathbf{w} = 1, \; (\boldsymbol{\beta} \otimes \mathbf{w})^\top \mathbf{Z}_s^\top \mathbf{Z}_s (\boldsymbol{\beta} \otimes \mathbf{w}) = 1 \end{array} \tag{6}$$

where the symbols have the same meanings as in Eq. (1). The algorithm in Eq. (6) was recently proposed by Dong and Qin (2018a), which is referred to as dynamic-inner CCA (DiCCA). Though other dynamic CCA algorithms are available (Liu et al. (2018)), DiCCA has shown superiority with explicit inner and outer modeling objectives.

In DiCCA, the dynamic relation between current sample and previous s samples is extracted, and sample $\mathbf{x}_k$ can be decomposed as

$$\mathbf{x}_k = \hat{\mathbf{x}}_k + \tilde{\mathbf{x}}_k = \mathbf{P}\mathbf{t}_k + \tilde{\mathbf{x}}_k \tag{7}$$

where the score vector $\mathbf{t}_k = \mathbf{R}^\top \mathbf{x}_k$. DiCCA works well to model the dynamics in the system, which can be employed to extract dynamic latent variables and locate root causes of oscillations.

3. DYNAMIC WEIGHTED CANONICAL CORRELATION ANALYSIS

In the era of big data, due to the advancement of sensory technologies, data sampling frequency in modern industrial processes can be very high; it is important for information collection, but has led to strong collinearity in successive observations. In this case, the number of relevant historical data s in DiCCA might be unnecessarily large due to redundant information in adjacent samples $\mathbf{x}_{k-i} (i \in \{1, 2, \ldots, s\})$, and the modeling performance of DiCCA will degrade.

In order to overcome the aforementioned issue, a dynamic weighted CCA (DWCCA) is proposed in this section. In DWCCA, we assume that each individual weight parameter β_i is a weighted combination of a few parameters or functions only, which can be calculated by

$$\beta_i = \sum_{j=0}^{q+1} \phi_j(i) b_j, \; i \in \{1, 2, \ldots, s\} \tag{8}$$

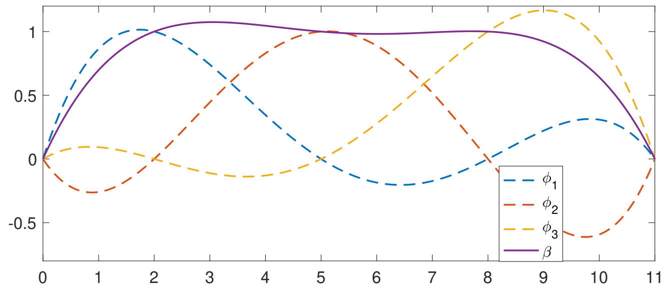

Fig. 1. Example of polynomial series and variations of $\boldsymbol{\beta}$

where $\boldsymbol{\phi}_j$ ($j \in \{0,1,\ldots,q+1\}, q \le s$) are known basis functions, and $\mathbf{b} = [b_0, b_1, \ldots, b_{q+1}]^\top$ is the new weight vector. In Eq. (8), similar adjacent samples will be represented with the same weight vector $\mathbf{b}$, thus eliminating redundant information in the modeling factors.

3.1 Selection of Function $\boldsymbol{\phi}_j$

It is noted that the design of basis functions $\boldsymbol{\phi}_j$ in Eq. (8) is very important for the success of DWCCA algorithm, and a variety of functions can be selected as $\boldsymbol{\phi}_j$, including Gaussian, Fourier and polynomial basis. In this article, due to the page limit, the following polynomial series with $q+1$ degrees (Almon (1965)) is considered as the basis in DWCCA; comparison with other basis functions will be conducted in a future work.

$$\begin{aligned}
\boldsymbol{\phi}_0(i) &= \tfrac{(i-i_1)(i-i_2)\ldots(i-i_{q+1})}{(i_0-i_1)(i_0-i_2)\ldots(i_0-i_{q+1})} \\
\boldsymbol{\phi}_1(i) &= \tfrac{(i-i_0)(i-i_2)\ldots(i-i_{q+1})}{(i_1-i_0)(i_1-i_2)\ldots(i_1-i_{q+1})} \\
&\vdots \\
\boldsymbol{\phi}_{q+1}(i) &= \tfrac{(i-i_0)(i-i_1)\ldots(i-i_q)}{(i_{q+1}-i_0)(i_{q+1}-i_1)\ldots(i_{q+1}-i_q)}
\end{aligned} \tag{9}$$

where i varies in the range of $\{1, 2, \ldots, s\}$. The number of parameter points q is determined by the number of available samples to some extent, and $\{i_0, i_1, \ldots, i_{q+1}\}$ are parameter points where $0 \le i_j \le s+1$. It is necessary to estimate the data distribution with different q, and to select the best one.

The assumption of polynomial functions is that the effect of a historical data will rise for a time period, then decline. It has no impact on samples before time 0 or after time $s+1$. To satisfy this requirement, i_0 and i_{q+1} are set as 0 and $s+1$ respectively, and then b_0 and b_{q+1} in Eq. (8) reduce to zero. Thus, Eq. (8) can be rewritten as

$$\beta_i = \sum_{j=1}^{q} \phi_j(i) b_j = \boldsymbol{\phi}(i)^\top \mathbf{b},\ i \in \{1, 2, \ldots, s\} \tag{10}$$

where $\boldsymbol{\phi}(i) = [\phi_1(i), \ldots \phi_q(i)]^\top$, and $\mathbf{b} = [b_1, \ldots, b_q]^\top$.

For better understanding, the following instance is used to illustrate the generation of $\boldsymbol{\beta}$ with a polynomial series. Assume that $s = 10$, $q = 3$, and parameter points are $\{0, 2, 5, 8, 11\}$. Then with Eq. (9), the variations of each $\boldsymbol{\phi}_j$ and the corresponding $\boldsymbol{\beta}$ when $b_1 = b_2 = b_3$ are presented in Fig. 1. It is observed that in Fig. 1, the effect of a sample is in presence for a period, and then diminishes after $s+1$, which satisfies the pre-set requirement for the basis functions.

Remark 1. The performance of DWCCA depends on the selection of q, while the locations of parameter points $\{i_0, i_1, \ldots, i_{q+1}\}$ make no difference (Almon (1965)). Additionally, values of the parameter points are not necessary to be integers.

3.2 DWCCA Objective

With Eq. (10) for each β_i, $\boldsymbol{\beta}$ can then be calculated by

$$\boldsymbol{\beta} = \boldsymbol{\Phi}^\top \mathbf{b} \tag{11}$$

where $\boldsymbol{\Phi} \in \mathbb{R}^{q\times s} = [\boldsymbol{\phi}(1), \boldsymbol{\phi}(2), \ldots, \boldsymbol{\phi}(s)]$.

Substituting Eq. (11) into Eq. (4) results in the relation of latent scores in DWCCA as follows

$$\begin{aligned}
t_k &= [\mathbf{x}_{k-1}^\top, \mathbf{x}_{k-2}^\top, \ldots, \mathbf{x}_{k-s}^\top](\boldsymbol{\beta} \otimes \mathbf{w}) + \tilde{t}_k \\
&= \underbrace{[\mathbf{x}_{k-1}^\top, \mathbf{x}_{k-2}^\top, \ldots, \mathbf{x}_{k-s}^\top](\boldsymbol{\Phi}^\top \mathbf{b} \otimes \mathbf{w})}_{\hat{t}_k} + \tilde{t}_k
\end{aligned} \tag{12}$$

where $(\boldsymbol{\Phi}^\top \mathbf{b} \otimes \mathbf{w})$ is the Kronecker product between $\boldsymbol{\Phi}^\top \mathbf{b}$ and $\mathbf{w}$.

In DWCCA, latent variables are extracted by maximizing the correlation between t_k and predicted score $\hat{t}_k$ in Eq. (12), which is mathematically expressed as

$$\max\ J = \frac{\sum_{k=s+1}^{s+N+1} t_k \hat{t}_k}{\sqrt{\sum_{k=s+1}^{s+N+1} t_k^2}\sqrt{\sum_{k=s+1}^{s+N+1} \hat{t}_k^2}}$$

With the aid of the denotations in Eq. (5), the above objective function can be represented as

$$\begin{aligned}
\max\ J =& \mathbf{w}^\top \mathbf{X}_{s+1}^\top \mathbf{Z}_s (\boldsymbol{\Phi}^\top \mathbf{b} \otimes \mathbf{w}) \\
\text{s.t.}\quad & \mathbf{w}^\top \mathbf{X}_{s+1}^\top \mathbf{X}_{s+1} \mathbf{w} = 1 \\
& (\boldsymbol{\Phi}^\top \mathbf{b} \otimes \mathbf{w})^\top \mathbf{Z}_s^\top \mathbf{Z}_s (\boldsymbol{\Phi}^\top \mathbf{b} \otimes \mathbf{w}) = 1
\end{aligned} \tag{13}$$

Remark 2. It is noted that only when $q = s$ and $\boldsymbol{\Phi}$ is an $s \times s$ identity matrix, DWCCA reduces to DiCCA. Generally, DWCCA requires fewer lagged variables to represent dynamic relations between current data and historical data since $q \le s$.

3.3 DWCCA Solution

To derive the solution of DWCCA, Lagrange multipliers are employed to Eq. (13) as follows.

$$\begin{aligned}
\mathcal{L} = \mathbf{w}^\top \mathbf{X}_{s+1}^\top \mathbf{Z}_s (\boldsymbol{\Phi}^\top \mathbf{b} \otimes \mathbf{w}) + \tfrac{\lambda_w}{2}(1 - \mathbf{w}^\top \mathbf{X}_{s+1}^\top \mathbf{X}_{s+1} \mathbf{w}) \\
+ \tfrac{\lambda_b}{2}\left[1 - (\boldsymbol{\Phi}^\top \mathbf{b} \otimes \mathbf{w})^\top \mathbf{Z}_s^\top \mathbf{Z}_s (\boldsymbol{\Phi}^\top \mathbf{b} \otimes \mathbf{w})\right]
\end{aligned}$$

Differentiating $\mathcal{L}$ with respect to $\mathbf{w}$ and $\mathbf{b}$ leads to

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial \mathbf{w}} =& \mathbf{X}_{s+1}^\top \mathbf{Z}_s (\boldsymbol{\Phi}^\top \mathbf{b} \otimes \mathbf{w}) + (\boldsymbol{\Phi}^\top \mathbf{b} \otimes \mathbf{I}_m)^\top \mathbf{Z}_s^\top \mathbf{X}_{s+1} \mathbf{w} \\
& - \lambda_w \mathbf{X}_{s+1}^\top \mathbf{X}_{s+1} \mathbf{w} \\
& - \lambda_b (\boldsymbol{\Phi}^\top \mathbf{b} \otimes \mathbf{I}_m)^\top \mathbf{Z}_s^\top \mathbf{Z}_s (\boldsymbol{\Phi}^\top \mathbf{b} \otimes \mathbf{I}_m) \mathbf{w} = 0
\end{aligned} \tag{14}$$

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial \mathbf{b}} =& \boldsymbol{\Phi}(\mathbf{I}_s \otimes \mathbf{w})^\top \mathbf{Z}_s^\top \mathbf{X}_{s+1} \mathbf{w} \\
& - \lambda_b \boldsymbol{\Phi}(\mathbf{I}_s \otimes \mathbf{w})^\top \mathbf{Z}_s^\top \mathbf{Z}_s (\mathbf{I}_s \otimes \mathbf{w}) \boldsymbol{\Phi}^\top \mathbf{b} = 0
\end{aligned} \tag{15}$$

where the relation $\mathbf{a} \otimes \mathbf{b} = (\mathbf{I}_a \otimes \mathbf{b})\mathbf{a} = (\mathbf{a} \otimes \mathbf{I}_b)\mathbf{b}$ is utilized, and $\mathbf{I}_a$ and $\mathbf{I}_b$ are identity matrices with dimension a and b respectively.

Denote

$$\begin{aligned}
\mathbf{T}_s &= [\mathbf{t}_s, \mathbf{t}_{s-1}, \ldots, \mathbf{t}_1] \\
&= [\mathbf{X}_s \mathbf{w}, \mathbf{X}_{s-1} \mathbf{w}, \ldots, \mathbf{X}_1 \mathbf{w}] \equiv \mathbf{Z}_s (\mathbf{I} \otimes \mathbf{w})
\end{aligned}$$

$$\mathbf{X}_b = \sum_{i=1}^{s} \boldsymbol{\phi}(i)^\top \mathbf{b} \mathbf{X}_{s-i+1} \equiv \mathbf{Z}_s(\boldsymbol{\Phi}^\top \mathbf{b} \otimes \mathbf{I})$$

Then Eqs. (14) and (15) can be rearranged as

$$\mathbf{X}_{s+1}^\top \mathbf{X}_b \mathbf{w} + \mathbf{X}_b^\top \mathbf{t}_{s+1} = \lambda_w \mathbf{X}_{s+1}^\top \mathbf{t}_{s+1} + \lambda_b \mathbf{X}_b^\top \mathbf{X}_b \mathbf{w} \quad (16)$$

$$\boldsymbol{\Phi} \mathbf{T}_s^\top \mathbf{t}_{s+1} = \lambda_b \boldsymbol{\Phi} \mathbf{T}_s^\top \mathbf{T}_s \boldsymbol{\Phi}^\top \mathbf{b} \quad (17)$$

where $\mathbf{t}_{s+1} = \mathbf{X}_{s+1}\mathbf{w}$.

Remark 3. It can be proved that $J = \lambda_w = \lambda_b$ by pre-multiplying Eqs. (16) and (17) with $\mathbf{w}^\top$ and $\mathbf{b}^\top$ respectively. Thus, maximizing J is equivalent to find the largest λ_w and λ_b, which can be calculated iteratively.

The first set of latent variable can be obtained by iterating through Eqs. (16) and (17), which corresponds to the largest λ_w and λ_b. For the next set of latent variable, the effect of $\mathbf{t}_1$ should be removed from $\mathbf{X}$, which can be achieved by

$$\mathbf{X} := \mathbf{X} - \mathbf{t}\mathbf{p}^\top \quad (18)$$

where loading vector $\mathbf{p}$ is derived by minimizing the regression error between $\mathbf{X}$ and $\mathbf{t}\mathbf{p}^\top$, leading to $\mathbf{p} = \mathbf{X}^\top \mathbf{t}/\mathbf{t}^\top \mathbf{t}$. The same procedure is repeated until enough dynamic latent variables are extracted for DWCCA.

The detailed algorithm of DWCCA is summarized in Algorithm 1, and there are several notes regarding to Algorithm 1.

1. Moore-Penrose inverse is adopted to calculate $\mathbf{b}$ and $\mathbf{w}$, since both $\boldsymbol{\Phi}\mathbf{T}_s^\top \mathbf{T}_s \boldsymbol{\Phi}^\top$ and $\mathbf{X}_{s+1}^\top \mathbf{X}_{s+1} + \mathbf{X}_b^\top \mathbf{X}_b$ may not be invertible. Alternatively, regularization terms can be incorporated into $\mathbf{Z}_s$ to address the ill-conditioned issues caused by strong collinearity in $\mathbf{X}$.
2. Instead of deflating each $\mathbf{X}_i$ ($i \in \{1, 2, \ldots, s+1\}$) individually with $\mathbf{X}_i := \mathbf{X}_i - \mathbf{t}_i \mathbf{p}_i^\top$, where $\mathbf{p}_i = \mathbf{X}_i^\top \mathbf{t}_i / \mathbf{t}_i^\top \mathbf{t}_i$, Eq. (18) is employed to deflate the whole dataset $\mathbf{X}$. Eq. (18) guarantees that both $\mathbf{X}_{s+1}$ and $\mathbf{Z}_s$ are deflated in the same way, ensuring that the dynamic relations between $\mathbf{t}_{s+1}$ and $\mathbf{T}_s$ can still captured by an auto-regressive model for the remaining latent variables (Dong and Qin (2018b)).

3.4 Hybrid Modeling

After performing DWCCA, the dynamic relations between current data and historical data are extracted, and the deflated $\mathbf{X}_{l+1}$ mainly contains static variations and noise. Static PCA is then conducted on $\mathbf{X}_{l+1}$ to separate useful static information from noise. Then the collected data $\mathbf{X}$ is decomposed into three parts by DWCCA and PCA.

$$\mathbf{X} = \hat{\mathbf{X}}_d + \hat{\mathbf{X}}_r + \mathbf{E}_r \quad (19)$$

where $\mathbf{E}_r$ is purely noise, and $\hat{\mathbf{X}}_d$ denotes the dynamics predicted by DWCCA, and it is represented as

$$\hat{\mathbf{X}}_d = \mathbf{T}\mathbf{P}^\top \quad (20)$$

where $\mathbf{T} = \mathbf{X}\mathbf{R}$, and $\mathbf{R} = \mathbf{W}(\mathbf{P}^\top \mathbf{W})^{-1}$.

$\hat{\mathbf{X}}_r$ is the static information extracted by PCA, which is predicted by

$$\hat{\mathbf{X}}_r = \mathbf{T}_r \mathbf{P}_r^\top \quad (21)$$

where $\mathbf{T}_r = [\mathbf{t}_{r1}, \mathbf{t}_{r2}, \ldots, \mathbf{t}_{rl_r}] = \mathbf{X}_{l+1}\mathbf{P}_r$, $\mathbf{P}_r = [\mathbf{p}_{r1}, \mathbf{p}_{r2}, \ldots, \mathbf{p}_{rl_r}]$, and l_r is number of static latent variables.

Algorithm 1 Dynamic Weighted CCA

1. Scale data $\mathbf{X}$ to zero mean and unit variance.
2. Select the weighted functions $\boldsymbol{\phi}_j$, $j \in \{1, 2, \ldots, q\}$ to form the weighted matrix $\boldsymbol{\Phi}$ in Eq. (11), and determine model parameters l and s with cross validation.
3. Initialize $\mathbf{w}$ with the first row of $\mathbf{X}$ and scale it to unit norm. Then conduct the following relations until convergence is achieved.
   i. $\mathbf{t} = \mathbf{X}\mathbf{w}$, and $\mathbf{t} = \mathbf{t}/||\mathbf{t}||$;
   ii. Derive $\mathbf{t}_i = [t_i, t_{i+1}, \ldots, t_{i+N}]^\top$ ($i \in \{1, 2, \ldots, s+1\}$) from $\mathbf{t}$, and form $\mathbf{T}_s = [\mathbf{t}_s, \mathbf{t}_{s-1}, \ldots, \mathbf{t}_1]$;
   iii. $\mathbf{b} = \left(\boldsymbol{\Phi}\mathbf{T}_s^\top \mathbf{T}_s \boldsymbol{\Phi}^\top\right)^\dagger \boldsymbol{\Phi}\mathbf{T}_s^\top \mathbf{t}_{s+1}$, and normalize $\mathbf{b}$ by $\mathbf{b} = \mathbf{b}/(\mathbf{t}_{s+1}^\top \mathbf{T}_s \boldsymbol{\Phi}^\top \mathbf{b})^{\frac{1}{2}}$;
   iv. $\mathbf{X}_b = \sum_{i=1}^{s} \boldsymbol{\phi}(i)^\top \mathbf{b} \mathbf{X}_{s-i+1}$;
   v. $\mathbf{w} = \left(\mathbf{X}_{s+1}^\top \mathbf{X}_{s+1} + \mathbf{X}_b^\top \mathbf{X}_b\right)^\dagger (\mathbf{X}_{s+1}^\top \mathbf{X}_b \mathbf{w} + \mathbf{X}_b^\top \mathbf{t}_{s+1})$;
   vi. Calculate correlation $J = \mathbf{t}_{s+1}^\top \mathbf{X}_b \mathbf{w}$.
4. Deflate $\mathbf{X}$ by removing the effect of extracted score $\mathbf{t}$,
$$\mathbf{p} = \mathbf{X}^\top \mathbf{t}/\mathbf{t}^\top \mathbf{t}$$
$$\mathbf{X} := \mathbf{X} - \mathbf{t}\mathbf{p}^\top$$
5. Repeat Step 3 and 4 until l dynamic latent variables are extracted.

For a new sample $\mathbf{x}_k$ sampled at time k, the dynamic part $\hat{\mathbf{x}}_{d,k}$ is obtained through predicted score vector $\hat{\mathbf{t}}_k$ as follow.

$$\hat{\mathbf{x}}_{d,k} = \mathbf{P}\hat{\mathbf{t}}_k \quad (22)$$

where $\hat{\mathbf{t}}_k$ is comprised by l dynamic latent variables, i.e., $\hat{\mathbf{t}}_k = [\hat{t}_{k,1}, \hat{t}_{k,2}, \ldots, \hat{t}_{k,l}]$, and each element $\hat{t}_{k,i}$ ($i \in \{1, 2, \ldots, l\}$) is connected with historical scores with an auto-regressive model by

$$\hat{\mathbf{t}}_{k,i} = (\boldsymbol{\phi}(1)^\top \mathbf{b}_i q^{-1} + \boldsymbol{\phi}(2)^\top \mathbf{b}_i q^{-2} + \ldots + \boldsymbol{\phi}(s)^\top \mathbf{b}_i q^{-s})\mathbf{t}_{k,i}$$

where $\mathbf{b}_i$ is the i^{th} extracted weighted vector in DWCCA, and q^{-1} is a backward shift operator.

After removing the dynamic information from $\mathbf{x}_k$, the remaining $\mathbf{x}_{k,l+1}$ includes both static information of large variance and noise. The static part $\hat{\mathbf{x}}_{r,k}$ is predicted by

$$\hat{\mathbf{x}}_{r,k} = \mathbf{P}_r \mathbf{t}_r \quad (23)$$

where $\mathbf{t}_r = \mathbf{P}_r^\top \mathbf{x}_{k,l+1}$.

4. CASE STUDY ON TE PROCESS

The Tennessee Eastman process (TEP) is a typical benchmark process created by Downs and Vogel (1993) to evaluate the effectiveness of the proposed control schemes and data-driven algorithms for system modeling, process monitoring and fault diagnosis. Four reactions occur in the process, where reactants (A, C, D, E) and a catalyst (B) are fed into the reactor to produce main products (G and H) and a byproduct (F).

Table 1. The First Three Weight Vectors $\mathbf{b}_j$

$\mathbf{b}_1$	$\mathbf{b}_2$	$\mathbf{b}_3$
0.3214	0.1091	0.6303
0.3935	0.6011	0.1198

Two sets of variables are available in TEP, which are 12 manipulated variables (XMV(1-12)) and 41 measured

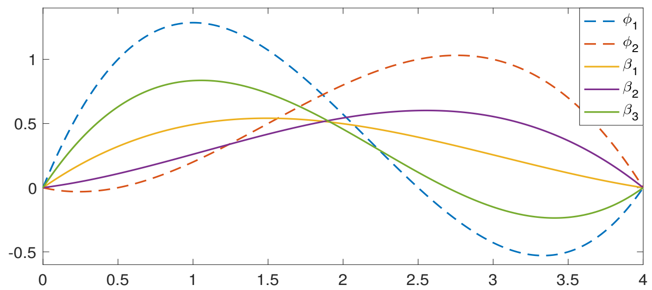

Fig. 2. Polynomial series and weight variations of the first three latent variables in TEP case study

variables (XMEAS(1-41)), and they are sampled at different frequencies: process measurements XMEAS(1-22) and manipulated variables XMV(1-12) are sampled every 3 minutes, component measurements XMEAS(23-36) in Streams 6 and 9 are sampled every 6 minutes, and product analysis XMEAS(37-41) in Stream 11 are with 15-minute sampling frequency.

In this case study, component measurements XMEAS(29-36) in purge gas stream (Stream 9) are selected to compare the effectiveness of DWCCA and DiCCA. In order to compensate the irregular sampling rates in various streams, the preprocessing procedure designed in Zhu et al. (2017a) is adopted here to de-duplicate samples.

Through cross validation, the parameters are selected: for DWCCA, $l = 7$, $s = 3$, $q = 2$, and $l_r = 1$; and for DiCCA, $l = 7$, $s = 12$, and $l_r = 1$. In DWCCA, the parameter points are selected as $[0, 0.5, 2.5, 4]$, and the variations of the polynomial functions ϕ_1 and ϕ_2 are shown in Fig. 2. In Fig. 2, the variations of $\boldsymbol{\beta}_j$ for the first three latent variables are also presented with weight vectors $\mathbf{b}_j$ listed in Table 1.

The auto-correlations and cross-correlations of the process variables, latent variables extracted by DWCCA, and residuals of DWCCA are shown in Figs. 3, 4 and 5, respectively. For ease of presentation, only the correlations of the first five variables are presented. It is evident that both dynamic and static correlations exist in the collected dataset, and DWCCA works well to capture the dynamics in the system. After the exploitation of dynamic variations in the process, only static variations are left in the residuals as shown in Fig. 5.

The extracted auto-correlations by DWCCA and DiCCA are presented in Fig. 6. In Fig. 6, the correlation coefficients captured by DWCCA among the first two dynamic components are comparable with or slightly larger than DiCCA. For the remaining factors, DiCCA has higher auto-correlations than DWCCA; however, this is caused by the strongly collinear successive samples, while after removing the collinearity, the actual auto-correlations between current data and historical data is smaller, as shown in the results of DWCCA.

The mean squared errors (MSE) for each variable of DWCCA and DiCCA are shown in Table 2. As indicated from the table, the prediction performance of DWCCA is better than DiCCA in terms of MSEs with fewer lagged data. The actual variations of each variable and their predictions by DWCCA are shown in Fig. 7. Though

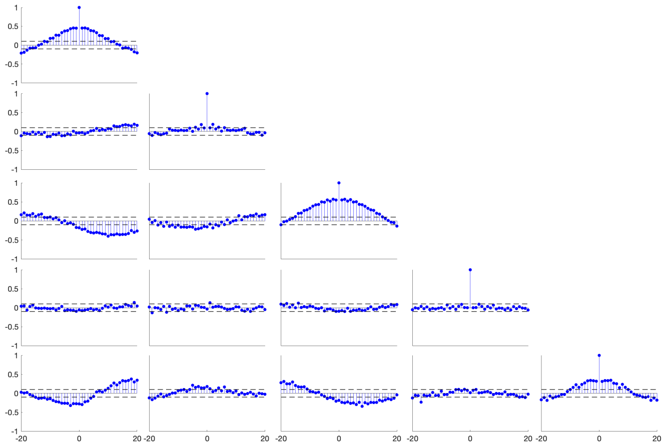

Fig. 3. Auto-correlations and cross-correlations of the first five process variables

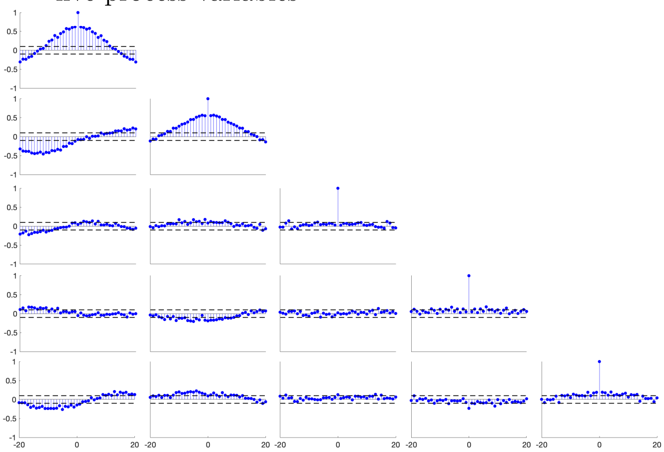

Fig. 4. Auto-correlations and cross-correlations captured by the first five DWCCA dynamic latent variables

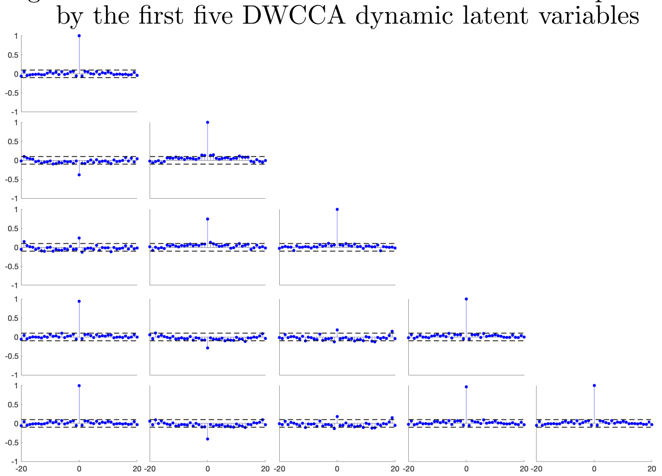

Fig. 5. Auto-correlations and cross-correlations of the first five residuals

DiCCA can model the dynamics in the model, it requires much more number of historical data, since DiCCA ignores strong collinearity in adjacent samples when extracting the dynamic latent variables. Therefore, the superiority of DWCCA over DiCCA has be demonstrated in terms of prediction efficiency and collinearity handling in dynamics extraction.

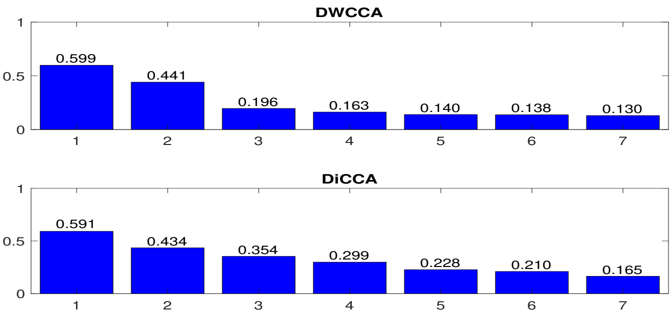

Fig. 6. Auto-correlations extracted by each latent variable

Table 2. Mean Squared Errors of Variables

Variable	DWCCA	DiCCA
XMEAS(29)	0.0059	0.0993
XMEAS(30)	0.0107	0.3889
XMEAS(31)	0.0149	0.1902
XMEAS(32)	0.0179	0.0511
XMEAS(33)	0.0004	0.0451
XMEAS(34)	0.0025	0.0429
XMEAS(35)	0.0004	0.2145
XMEAS(36)	0.3228	0.6327

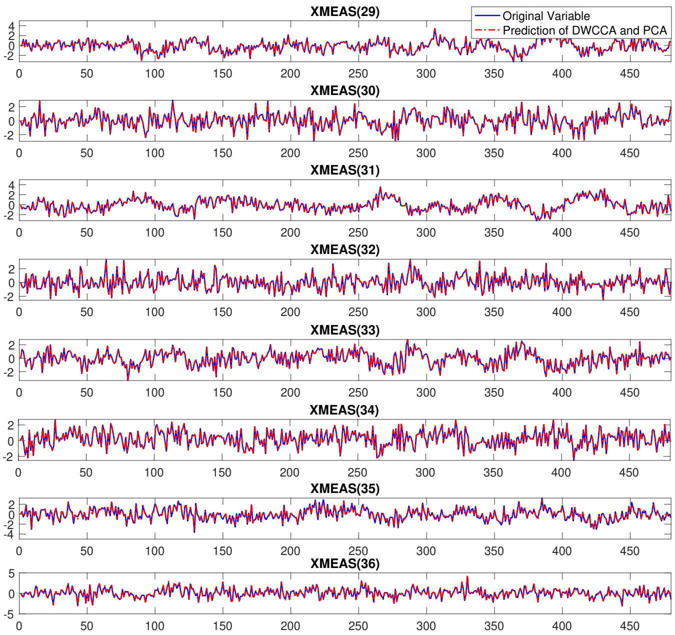

Fig. 7. Actual variations and DWCCA predictions for process variables

5. CONCLUSIONS

In this article, a new dynamic CCA algorithm, referred to as DWCCA, is proposed to capture dynamic relations among measured variables. DWCCA extracts dynamics by maximizing the correlations between current data point and a weighted representation of historical data points. In addition to dynamic modeling capability, DWCCA relies on a set of basis functions to handle the strong collinearity that exists among adjacent samples. Its effectiveness in extraction of dynamic relations and collinearity handling is demonstrated through the TE process. A future work is to study the effect of different basis functions.

ACKNOWLEDGEMENTS

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

REFERENCES

Almon, S. (1965). The distributed lag between capital appropriations and expenditures. *Econometrica: Journal of the Econometric Society*, 178–196.

Dong, Y. and Qin, S.J. (2018a). Dynamic latent variable analytics for process operations and control. *Computers & Chemical Engineering*, 114, 69–80.

Dong, Y. and Qin, S.J. (2018b). A novel dynamic PCA algorithm for dynamic data modeling and process monitoring. *Journal of Process Control*, 67, 1–11.

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

Ge, Z., Yang, C., Song, Z., and Wang, H. (2008). Robust online monitoring for multimode processes based on nonlinear external analysis. *Industrial & Engineering Chemistry Research*, 47(14), 4775–4783.

Hotelling, H. (1936). Relations between two sets of variates. *Biometrica*, 28, 321–377.

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

Ku, W., Storer, R.H., and Georgakis, C. (1995). Disturbance detection and isolation by dynamic principal component analysis. *Chemometrics and intelligent laboratory systems*, 30(1), 179–196.

Li, G., Qin, S.J., and Zhou, D. (2014). A new method of dynamic latent-variable modeling for process monitoring. *IEEE Transactions on Industrial Electronics*, 61(11), 6438–6445.

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.

Liu, Y., Liu, B., Zhao, X., and Xie, M. (2017). A mixture of variational canonical correlation analysis for nonlinear and quality-relevant process monitoring. *IEEE Transactions on Industrial Electronics*, 65(8), 6478–6486.

MacGregor, J.F. and Kourti, T. (1995). Statistical process control of multivariate processes. *Control Engineering Practice*, 3(3), 403–414.

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

Yin, S., Ding, S.X., Xie, X., and Luo, H. (2014). A review on basic data-driven approaches for industrial process monitoring. *IEEE Transactions on Industrial Electronics*, 61(11), 6418–6428.

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. (2017a). Concurrent quality and process monitoring with canonical correlation analysis. *Journal of Process Control*, 60, 95–103.

Zhu, Q., Liu, Q., and Qin, S.J. (2017b). Quality-relevant fault detection of nonlinear processes based on kernel concurrent canonical correlation analysis. In *2017 American Control Conference (ACC)*, 5404–5409. IEEE.