

# Dynamic-Inner Canonical Correlation and Causality Analysis for High Dimensional Time Series Data

Yining Dong<sup>\*,\*\*</sup> S. Joe Qin<sup>\*\*,\*\*\*</sup>

<sup>\*</sup> Ming Hsieh Department of Electrical Engineering, University of  
Southern California, Los Angeles, CA 90089, USA

<sup>\*\*</sup> The Chinese University of Hong Kong, Shenzhen, 2001 Longxiang  
Road, Longgang District, Shenzhen, Guangdong, China

<sup>\*\*\*</sup> Mork Family Department of Chemical Engineering and Material  
Science, University of Southern California, Los Angeles, CA 90089,  
USA (e-mail: sqin@usc.edu).

---

**Abstract:** In this paper, a novel dynamic-inner canonical correlation analysis (DiCCA) algorithm is proposed to extract dynamic components from high dimensional dynamic data. DiCCA extracts latent variables with descending dynamics, which are referred to as *principal time series*. Since DiCCA enables the principal time series to have maximal predictability, the most important dynamic features in the data are guaranteed to be extracted first. Therefore, usually a lower dimensional principal time series are able to provide good representation of the dynamic features, leading to the ease of interpretation and visualization. A case study on the Eastman plant-wide oscillating dataset demonstrates the effectiveness of the proposed method. Combined with Granger causality analysis, major oscillatory latent dynamics are analyzed, identified, and localized to equipment malfunctions.

*Keywords:* latent dynamic model, dynamic data modeling, Granger causality analysis, root cause diagnosis

---

## 1. INTRODUCTION

Data collected from industrial processes are often high dimensional, highly cross-correlated, and highly auto-correlated. Such data can be represented with a lower dimensional latent model, which extracts these major cross-correlations and/or auto-correlations, leaving only insignificant noise in the residuals. Principal component analysis (PCA) is a latent variable modeling method that has been widely used for static data modeling (Jackson (2005); Joe Qin (2003)). It projects the original data onto a lower dimensional subspace such that the projections, which are also called latent variables, capture maximal variance in the data. In this way, the original high dimensional data can be represented by a set of lower dimensional latent variables, which makes PCA a useful dimension reduction method. However, even though PCA is an effective latent variable modeling and dimension reduction tool for static data modeling, it is not appropriate to apply it on dynamic data. This is because for dynamic data, the future data can be partially predicted by the past data. Directly applying PCA to dynamic data cannot guarantee that this predictability is preserved in the latent variables.

Although consecutive samples are not necessary for static data modeling, they are usually required for dynamic data modeling due to time dependence. Therefore, dynamic data can also be viewed as time series data. To model the dynamic data or time series data, several dynamic PCA algorithms have been proposed. A straightforward

extension was proposed by Ku et al. where a number of lagged measurements are included in the data matrix (Ku et al. (1995)). Static PCA is performed on the augmented data matrix to extract latent variables. The drawback of the proposed method is that static components and dynamic components are mixed together in the extracted latent variables, which fails to guarantee that the latent variables are predictable from their past values. In addition, the number of latent variables can be greater than the number of original variables, which makes it not a dimension reduction tool.

Another structured dynamic PCA algorithm was proposed by Li et al. where the variance of a weighted sum of lagged latent variables is maximized (Li et al. (2014)). Although this method preserves the dimension reduction feature, static relationship can dominate a latent variable. Therefore, such latent variable can fail to capture the dynamic components in the data, which means the latent variable does not have the predictability feature.

Recently, Dong and Qin (2017) proposed a DiPCA algorithm where the latent variables are extracted to have the highest covariance with their predictions from the past. The latent variables extracted by DiPCA preserves both dimension reduction and predictability features. The disadvantage of this method is that by maximizing the covariance between the latent variables and their predictions, there is a trade off between predictability and the variance captured in the latent variables. It does not extract latent

variables in a descending order of predictability or  $R^2$  values (coefficient of determination), which is commonly used to assess the prediction performance.

In this paper, a DiCCA (dynamic-inner canonical correlation analysis) algorithm for dynamic data or high dimensional time series modeling is proposed. It extracts latent variables such that they have maximal correlation with the predictions from their past values. The time series formed by the extracted latent variables are called principal time series. Similar to DiPCA, the dimension reduction feature is preserved in DiCCA algorithm. While DiPCA extracts latent variables by considering both variance and predictability, DiCCA focuses on predictability only. It is guaranteed that the extracted latent variables or principal time series have descending order of predictability, or  $R^2$  values. In addition, the extraction of the latent variables and dynamic modeling of the latent variables are achieved simultaneously by solving the DiCCA objective function. This is an advantage of DiCCA over DiPCA.

The remainder of the paper is organized as follows. Section 2 presents the proposed DiCCA algorithm. Section 3 explores the geometric properties of DiCCA. Section 4 shows DiCCA model relations. The industrial case study on TEP oscillating dataset in section 5 demonstrates the effectiveness of the proposed DiCCA modeling algorithm. Section 6 gives conclusions.

## 2. DYNAMIC-INNER CANONICAL CORRELATION ANALYSIS

### 2.1 DiCCA Objective

Let  $\mathbf{x}_k$  denote a sample vector of  $m$  variables at time  $k$ , and assume the latent variable  $t_k$  is a linear combination of the original variables

$$t_k = \mathbf{x}_k^T \mathbf{w} \quad (1)$$

We wish to represent the dynamics of  $\mathbf{x}_k$  in  $t_k$ . This is done by ensuring that  $t_k$  is correlated to its past values as much as possible. In general, the prediction of  $t_k$  from its past can be represented by an auto-regressive (AR) model as

$$t_k = \beta_1 t_{k-1} + \dots + \beta_s t_{k-s} + r_k$$

where  $s$  is the dynamic order of the AR model. For a more general auto-regressive moving average (ARMA) model relation, it is well known that it can be well approximated by a high order AR model. Therefore, when  $s$  is large enough such that the residual  $r_k$  is essentially white noise, the prediction of  $t_k$  from the AR model is

$$\begin{aligned} \hat{t}_k &= \beta_1 t_{k-1} + \dots + \beta_s t_{k-s} \\ &= \mathbf{x}_{k-1}^T \mathbf{w} \beta_1 + \dots + \mathbf{x}_{k-s}^T \mathbf{w} \beta_s \\ &= [\mathbf{x}_{k-1}^T \ \dots \ \mathbf{x}_{k-s}^T] (\boldsymbol{\beta} \otimes \mathbf{w}) \end{aligned}$$

where  $\boldsymbol{\beta} = [\beta_1 \ \beta_2 \ \dots \ \beta_s]^T$  and  $\boldsymbol{\beta} \otimes \mathbf{w}$  is the Kronecker product. Mathematically, we wish to ensure that  $t_k$  is best predicted by  $\hat{t}_k$ . This is done by maximizing the correlation between  $t_k$  and  $\hat{t}_k$ , which is represented as

$$\frac{\sum_{k=s+1}^{s+N} t_k \hat{t}_k}{\sqrt{\sum_{k=s+1}^{s+N} t_k^2} \sqrt{\sum_{k=s+1}^{s+N} \hat{t}_k^2}} \quad (2)$$

It can be shown that when restricting  $\sum_{k=s+1}^{s+N} t_k^2 = 1$  and  $\sum_{k=s+1}^{s+N} \hat{t}_k^2 = 1$ , maximizing (2) is equivalent to minimizing  $\sum_{k=s+1}^{s+N} (t_k - \hat{t}_k)^2$ , the residual sum of squares of the prediction model under these constraints.

Given the data matrix as

$$\mathbf{X} = [\mathbf{x}_1 \ \mathbf{x}_2 \ \dots \ \mathbf{x}_{s+N}]^T$$

and  $\mathbf{t} = \mathbf{X}\mathbf{w}$  as the vector of latent scores, we form the following data matrices from  $\mathbf{X}$ ,

$$\begin{aligned} \mathbf{X}_i &= [\mathbf{x}_i \ \mathbf{x}_{i+1} \ \dots \ \mathbf{x}_{N+i-1}]^T \text{ for } i = 1, 2, \dots, s+1 \\ \mathbf{Z}_s &= [\mathbf{X}_s \ \mathbf{X}_{s-1} \ \dots \ \mathbf{X}_1] \end{aligned}$$

The objective (2) can be rewritten as

$$\begin{aligned} &\frac{\mathbf{w}^T \mathbf{X}_{s+1}^T (\mathbf{X}_s \mathbf{w} \beta_1 + \dots + \mathbf{X}_1 \mathbf{w} \beta_s)}{\|\mathbf{X}_{s+1} \mathbf{w}\| \|\mathbf{X}_s \mathbf{w} \beta_1 + \dots + \mathbf{X}_1 \mathbf{w} \beta_s\|} \\ &= \frac{\mathbf{w}^T \mathbf{X}_{s+1}^T \mathbf{Z}_s (\boldsymbol{\beta} \otimes \mathbf{w})}{\|\mathbf{X}_{s+1} \mathbf{w}\| \|\mathbf{Z}_s (\boldsymbol{\beta} \otimes \mathbf{w})\|} \end{aligned} \quad (3)$$

which can be further simplified as the following objective function,

$$\begin{aligned} \max_{\mathbf{w}, \boldsymbol{\beta}} \quad & J = \mathbf{w}^T \mathbf{X}_{s+1}^T \mathbf{Z}_s (\boldsymbol{\beta} \otimes \mathbf{w}) \\ \text{s.t.} \quad & \|\mathbf{X}_{s+1} \mathbf{w}\| = 1, \|\mathbf{Z}_s (\boldsymbol{\beta} \otimes \mathbf{w})\| = 1 \end{aligned} \quad (4)$$

### 2.2 Extracting One Dynamic Component

Lagrange multipliers are applied to solve the optimization problem in (4). Define

$$\begin{aligned} L &= \mathbf{w}^T \mathbf{X}_{s+1}^T \mathbf{Z}_s (\boldsymbol{\beta} \otimes \mathbf{w}) + \frac{1}{2} \lambda_1 (1 - \mathbf{w}^T \mathbf{X}_{s+1}^T \mathbf{X}_{s+1} \mathbf{w}) \\ &\quad + \frac{1}{2} \lambda_2 (1 - (\boldsymbol{\beta} \otimes \mathbf{w})^T \mathbf{Z}_s^T \mathbf{Z}_s (\boldsymbol{\beta} \otimes \mathbf{w})) \end{aligned}$$

using  $(\boldsymbol{\beta} \otimes \mathbf{w}) = (\mathbf{I} \otimes \mathbf{w}) \boldsymbol{\beta} = (\boldsymbol{\beta} \otimes \mathbf{I}) \mathbf{w}$  and taking derivatives of  $L$  with respect to  $\boldsymbol{\beta}$  and  $\mathbf{w}$  and set them to zero respectively, we have

$$\begin{aligned} \frac{\partial L}{\partial \boldsymbol{\beta}} &= (\mathbf{I} \otimes \mathbf{w})^T \mathbf{Z}_s^T \mathbf{X}_{s+1} \mathbf{w} \\ &\quad - \lambda_2 (\mathbf{I} \otimes \mathbf{w})^T \mathbf{Z}_s^T \mathbf{Z}_s (\mathbf{I} \otimes \mathbf{w}) \boldsymbol{\beta} = 0 \end{aligned} \quad (5)$$

$$\begin{aligned} \frac{\partial L}{\partial \mathbf{w}} &= \mathbf{X}_{s+1}^T \mathbf{Z}_s (\boldsymbol{\beta} \otimes \mathbf{w}) + (\boldsymbol{\beta} \otimes \mathbf{I})^T \mathbf{Z}_s^T \mathbf{X}_{s+1} \mathbf{w} \\ &\quad - \lambda_1 \mathbf{X}_{s+1}^T \mathbf{X}_{s+1} \mathbf{w} - \lambda_2 (\boldsymbol{\beta} \otimes \mathbf{I})^T \mathbf{Z}_s^T \mathbf{Z}_s (\boldsymbol{\beta} \otimes \mathbf{I}) \mathbf{w} = 0 \end{aligned} \quad (6)$$

Pre-multiplying equation (5) by  $\boldsymbol{\beta}^T$  and refer to the constraint, we have  $J = \lambda_2$ . Pre-multiplying equation (6) by  $\mathbf{w}^T$  and refer to the constraint, we have  $2J - \lambda_1 - J = 0$ , leading to  $\lambda_1 = J$ . In addition, define

$$\mathbf{Z}_s (\mathbf{I} \otimes \mathbf{w}) = [\mathbf{X}_s \mathbf{w} \ \mathbf{X}_{s-1} \mathbf{w} \ \dots \ \mathbf{X}_1 \mathbf{w}] = [\mathbf{t}_s \ \dots \ \mathbf{t}_1] \triangleq \mathbf{T}_s$$

$$\mathbf{Z}_s (\boldsymbol{\beta} \otimes \mathbf{I}) = \sum_{i=1}^s \beta_i \mathbf{X}_{s-i+1} \triangleq \mathbf{X}_\beta$$

(6) and (5) can be rewritten as

$$\begin{aligned} \mathbf{T}_s^T \mathbf{X}_{s+1} \mathbf{w} &= J \mathbf{T}_s^T \mathbf{T}_s \boldsymbol{\beta} \\ \mathbf{X}_{s+1}^T \mathbf{X}_\beta \mathbf{w} + \mathbf{X}_{s+1}^T \mathbf{X}_{s+1} \mathbf{w} &= J \mathbf{X}_{s+1}^T \mathbf{X}_{s+1} \mathbf{w} + J \mathbf{X}_\beta^T \mathbf{X}_\beta \mathbf{w} \end{aligned}$$

which can be further simplified as

$$(\mathbf{T}_s^T \mathbf{T}_s)^{-1} \mathbf{T}_s^T \mathbf{X}_{s+1} \mathbf{w} = J \boldsymbol{\beta} \quad (7)$$

$$(\mathbf{X}_{s+1}^T \mathbf{X}_{s+1} + \mathbf{X}_\beta^T \mathbf{X}_\beta)^+ (\mathbf{X}_{s+1}^T \mathbf{X}_\beta + \mathbf{X}_\beta^T \mathbf{X}_{s+1}) \mathbf{w} = J \mathbf{w} \quad (8)$$

where the sign  $()^+$  denotes the Moore-Penrose pseudo-inverse. The equations (8) and (7) imply that  $\mathbf{w}$  is the eigenvector of  $(\mathbf{X}_{s+1}^T \mathbf{X}_{s+1} + \mathbf{X}_\beta^T \mathbf{X}_\beta)^+ (\mathbf{X}_{s+1}^T \mathbf{X}_\beta + \mathbf{X}_\beta^T \mathbf{X}_{s+1})$  corresponding to the largest eigenvalue and  $\boldsymbol{\beta}$  is proportional to the least squares solution of the AR model parameters of time series  $\{t_k\}_{k=1}^{N+s+1}$ . However, since  $\mathbf{X}_\beta$  depends on  $\boldsymbol{\beta}$ ,  $\boldsymbol{\beta}$  and  $\mathbf{w}$  are coupled together and there is no analytical solution to the optimization problem (4). Using  $\mathbf{t}_i = \mathbf{X}_i \mathbf{w}$ ,  $i = 1, 2, \dots, s+1$ , we have

$$\mathbf{X}_\beta \mathbf{w} = \sum_{i=1}^s \beta_i \mathbf{t}_{s-i+1}$$

Eq. (8) can be reformulated as follows

$$J \mathbf{w} = (\mathbf{X}_{s+1}^T \mathbf{X}_{s+1} + \mathbf{X}_\beta^T \mathbf{X}_\beta)^+ (\mathbf{X}_{s+1}^T \sum_{i=1}^s \beta_i \mathbf{t}_{s-i+1} + \mathbf{X}_\beta^T \mathbf{t}_{s+1})$$

The following iterative algorithm is formulated to solve the problem.

- (1) Initialize  $\mathbf{w}$  with unit vector.
- (2) Calculate  $\mathbf{w}, \boldsymbol{\beta}$  by iterating the following relations until convergence.

$\mathbf{t} = \mathbf{X} \mathbf{w}$  and  $\mathbf{t} := \mathbf{t} / \|\mathbf{t}\|$ . Form  $\mathbf{t}_i$  for  $i = 1, \dots, s+1$ ,

and denote  $\mathbf{T}_s = [\mathbf{t}_s \ \dots \ \mathbf{t}_1]$

$$\boldsymbol{\beta} = (\mathbf{T}_s^T \mathbf{T}_s)^{-1} \mathbf{T}_s^T \mathbf{t}_{s+1}$$

Normalize  $\boldsymbol{\beta}$ .  $\boldsymbol{\beta} := \boldsymbol{\beta} / (\mathbf{t}_{s+1}^T \mathbf{T}_s \boldsymbol{\beta})^{0.5}$

$$\mathbf{X}_\beta = \sum_{i=1}^s \beta_i \mathbf{X}_{s-i+1}$$

$$\mathbf{w} = (\mathbf{X}_{s+1}^T \mathbf{X}_{s+1} + \mathbf{X}_\beta^T \mathbf{X}_\beta)^+ (\mathbf{X}_{s+1}^T \sum_{i=1}^s \beta_i \mathbf{t}_{s-i+1} + \mathbf{X}_\beta^T \mathbf{t}_{s+1})$$

- (3) Calculate  $J = \mathbf{t}_{s+1}^T \sum_{i=1}^s \beta_i \mathbf{t}_{s-i+1}$ .

To extract the next dynamic latent variable, the same iteration procedure can be applied to the deflated matrices of  $\mathbf{X}_{s+1}$  and  $\mathbf{Z}_s$ , which are obtained from the deflated  $\mathbf{X}$  as follows,

$$\mathbf{X} := \mathbf{X} - \mathbf{t} \mathbf{p}^T \quad (9)$$

where the loading vector  $\mathbf{p}$  is defined as

$$\mathbf{p} = \mathbf{X}^T \mathbf{t} / \mathbf{t}^T \mathbf{t} \quad (10)$$

To make the score  $\mathbf{t}$  representing the variance captured by the component, it is desirable to rescale  $\mathbf{p}$  to unit norm as follows,

$$\begin{aligned} \mathbf{t} &:= \mathbf{t} / \|\mathbf{p}\| \\ \mathbf{w} &:= \mathbf{w} / \|\mathbf{p}\| \\ \mathbf{p} &:= \mathbf{p} / \|\mathbf{p}\| \end{aligned}$$

### 2.3 Dynamic Inner Modeling

After obtaining the latent variable  $t_k$ , an AR model can be built to describe the dynamics in  $t_k$  as

$$t_k = \beta_1 t_{k-1} + \dots + \beta_s t_{k-s} + r_k \quad (11)$$

where the coefficients  $\boldsymbol{\beta}$  coincidentally is already solved in the iterative algorithm. Therefore, there is no need to fit another AR model.

Compared to other dynamic data modeling algorithms such as DiPLS (Dong and Qin (2015)) and DiPCA (Dong and Qin (2017)), where a re-estimation of  $\boldsymbol{\beta}$  has to be done after the outer model projection, the extraction of the latent variables and dynamic modeling of the latent variables are achieved simultaneously in DiCCA. This is because DiCCA employs consistent outer modeling and inner modeling objectives, which is a unique property of DiCCA and makes it a more efficient dynamic modeling algorithm than the others.

### 2.4 DiCCA Model with $l$ components

DiCCA algorithm extracts latent time series one by one with descending predictability or  $R^2$  values. After  $l$  latent time series are extracted, the next latent time series extracted will have a  $R^2$  value close to 0, which implies that there are little or no dynamics left in the residuals. The orthogonality of the latent scores guarantees that the number of latent time series required to extract all dynamics is fewer than the number of variables, which will be shown later in the paper. Mathematically, by using  $t_{(j),k}$  to denote the  $j^{\text{th}}$  latent score at time  $k$ , and  $\beta_{ji}$  for  $i = 1, 2, \dots, s$  to denote the AR coefficients for the  $j^{\text{th}}$  latent score, we have the prediction model for each score as

$$\begin{aligned} \hat{t}_{(j),k} &= (\beta_{j1} q^{-1} + \beta_{j2} q^{-2} + \dots + \beta_{js} q^{-s}) t_{(j),k} \\ &= G_j(q^{-1}) t_{(j),k} \end{aligned} \quad (12)$$

where  $q^{-1}$  is the backward shift operator. By combining  $l$  prediction models together, we can obtain a prediction model for the latent score vector  $\mathbf{t}_k = [t_{(1),k} \ t_{(2),k} \ \dots \ t_{(l),k}]$  as

$$\begin{aligned} \hat{\mathbf{t}}_k &= \mathbf{G}(q^{-1}) \mathbf{t}_k \\ &= \text{diag}(G_1(q^{-1}), G_2(q^{-1}), \dots, G_l(q^{-1})) \mathbf{t}_k \end{aligned} \quad (13)$$

## 3. DICCA GEOMETRIC PROPERTIES

To explore the geometric properties with  $j$  latent variables (LV) being extracted, we use a subscript to denote the iteration from one LV to another as follows.

$$\mathbf{X}_{(j+1)} = \mathbf{X}_{(j)} - \mathbf{t}_{(j)} \mathbf{p}_j^T \quad \text{with } \mathbf{p}_j = \mathbf{X}_{(j)}^T \mathbf{t}_{(j)} / \mathbf{t}_{(j)}^T \mathbf{t}_{(j)} \quad (14)$$

where the subscript  $j$  denote the quantity of the  $j^{\text{th}}$  LV. The following relationships among the residual matrices and loading vectors can be proven

- (1)  $\mathbf{X}_{(j)} = \mathbf{H} \mathbf{X}_{(i+1)} \quad \forall i < j$
- (2)  $\mathbf{X}_{(j)} = \mathbf{X}_{(i+1)} \mathbf{Z} \quad \forall i < j$
- (3)  $\mathbf{X}_{(j)} \mathbf{w}_i = 0 \quad i < j$
- (4)  $\mathbf{X}_{(j)}^T \mathbf{t}_{(i)} = 0 \quad i < j$
- (5)  $\mathbf{w}_i^T \mathbf{p}_i = 1$
- (6)  $\mathbf{t}_{(i)}^T \mathbf{t}_{(j)} = 0 \quad \forall i \neq j$
- (7)  $\mathbf{w}_i^T \mathbf{p}_j = 0 \quad \forall i < j$
- (8)  $\mathbf{w}_i^T \mathbf{w}_j = 0 \quad \forall i \neq j$

Since the deflation step is similar to the deflation step of PLS, the proof for all of relationships except relationship 8) are the same as the proof for PLS (Helland (1988)).

The proof for relationship 8) is as follows. From previous derivation, we have

$$(\mathbf{X}_{(j)}^T \mathbf{X}_{(j)} + \mathbf{X}_{\beta(j)}^T \mathbf{X}_{\beta(j)})^+ (\mathbf{X}_{(j)}^T \mathbf{X}_{\beta(j)} + \mathbf{X}_{\beta(j)}^T \mathbf{X}_{(j)}) \mathbf{w}_j = J \mathbf{w}_j$$

where

$$\mathbf{X}_{\beta(j)} = \sum_{d=1}^s \beta_{s+1-d,j} \mathbf{X}_{d,(j)}$$

$\mathbf{X}_{d,(j)}$  are sub-matrices of  $\mathbf{X}_{(j)}$ . Due to property 3), we have

$$(\mathbf{X}_{(j)}^T \mathbf{X}_{(j)} + \mathbf{X}_{\beta(j)}^T \mathbf{X}_{\beta(j)}) \mathbf{w}_i = 0$$

Using the property of pseudo-inverse that  $\ker(\mathbf{A})^+ = \ker(\mathbf{A})^T$ ,  $\forall \mathbf{A}$ , we have

$$(\mathbf{X}_{(j)}^T \mathbf{X}_{(j)} + \mathbf{X}_{\beta(j)}^T \mathbf{X}_{\beta(j)})^+ \mathbf{w}_i = 0$$

and

$$\mathbf{w}_i^T (\mathbf{X}_{(j)}^T \mathbf{X}_{(j)} + \mathbf{X}_{\beta(j)}^T \mathbf{X}_{\beta(j)})^+ = 0$$

Therefore,

$$J \mathbf{w}_i^T \mathbf{w}_j = \mathbf{w}_i^T (\mathbf{X}_{(j)}^T \mathbf{X}_{(j)} + \mathbf{X}_{\beta(j)}^T \mathbf{X}_{\beta(j)})^+ (\mathbf{X}_{(j)}^T \mathbf{X}_{\beta(j)} + \mathbf{X}_{\beta(j)}^T \mathbf{X}_{(j)}) \mathbf{w}_j = 0$$

Since  $J \neq 0$ ,  $\mathbf{w}_i^T \mathbf{w}_j = 0, \forall i \neq j$ .

#### 4. DICCA MODEL RELATIONS

For DiCCA, having the same geometric properties as PLS leads to a similar model structure as PLS. Assuming the number of latent variables is chosen as  $l$  in the DiCCA model and defining the following matrices,

$$\begin{aligned} \mathbf{T} &= [\mathbf{t}_{(1)} \quad \mathbf{t}_{(2)} \quad \cdots \quad \mathbf{t}_{(l)}] \\ \mathbf{W} &= [\mathbf{w}_1 \quad \mathbf{w}_2 \quad \cdots \quad \mathbf{w}_l] \\ \mathbf{P} &= [\mathbf{p}_1 \quad \mathbf{p}_2 \quad \cdots \quad \mathbf{p}_l] \end{aligned}$$

By iterating (14), we have the following relations

$$\mathbf{X}_{(l+1)} = \mathbf{X} - \sum_{i=1}^l \mathbf{t}_{(i)} \mathbf{p}_i^T = \mathbf{X} - \mathbf{TP}^T$$

or equivalently

$$\mathbf{X} = \mathbf{X}_{(l+1)} + \mathbf{TP}^T \quad (15)$$

Post-multiplying (15) by  $\mathbf{W}$  and applying geometric property 3), we have

$$\mathbf{XW} = \mathbf{X}_{(l+1)} \mathbf{W} + \mathbf{TP}^T \mathbf{W} = \mathbf{TP}^T \mathbf{W}$$

which leads to

$$\mathbf{T} = \mathbf{XR} \quad (16)$$

where

$$\mathbf{R} = \mathbf{W}(\mathbf{P}^T \mathbf{W})^{-1} \quad (17)$$

Therefore, for a given vector  $\mathbf{x}_k$ , the score vector can be calculated as

$$\mathbf{t}_k = \mathbf{R}^T \mathbf{x}_k$$

and  $\mathbf{x}_k$  can be decomposed as

$$\mathbf{x}_k = \mathbf{P} \mathbf{t}_k + \tilde{\mathbf{x}}_k \quad (18)$$

The decomposition as (18) gives the partition of the space formed by current data. Furthermore, to explore the relations between the past data and current data, relation (13) can be utilized to derive the following relationships,

$$\mathbf{e}_k = \mathbf{x}_k - \mathbf{P} \hat{\mathbf{t}}_k = \mathbf{x}_k - \mathbf{P} \mathbf{G}(q^{-1}) \mathbf{t}_k \quad (19)$$

where  $\mathbf{e}_k$  is the one step ahead prediction error. Since the number of dynamic latent variables  $l$  is selected to extract all the dynamics in the data, there will be little or no

dynamics remaining in  $\mathbf{e}_k$ . With static PCA modeling of  $\mathbf{e}_k$ , we have the following complete decomposition of  $\mathbf{x}_k$  as

$$\mathbf{x}_k = \mathbf{P} \mathbf{G}(q^{-1}) \mathbf{t}_k + \mathbf{P}_r \mathbf{t}_{k,r} + \mathbf{e}_{k,r} \quad (20)$$

where  $\mathbf{t}_{k,r}$  are a collection of static latent variables.

In addition, since there is little or no dynamics left in  $\mathbf{e}_k$  after the removal of the one step ahead prediction, DiCCA can be interpreted as a dynamic whitening filter which removes all the dynamics in the data. Mathematically, based on (13) and (19), the dynamic whitening filter can be written as

$$\mathbf{e}_k = (\mathbf{I} - \mathbf{P} \mathbf{G}(q^{-1}) \mathbf{R}^T) \mathbf{x}_k$$

The block diagram of DiCCA dynamic whitening filter is shown in Fig.1.

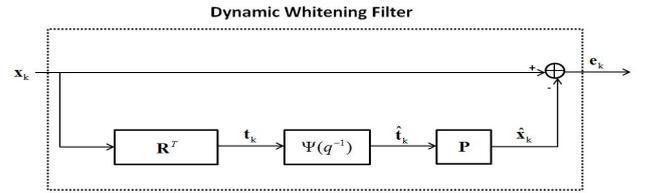


Fig. 1. Dynamic Whitening Filter Structure

#### 5. INDUSTRIAL CASE STUDY TO EXTRACT OSCILLATION COMPONENTS

In this section, DiCCA modeling algorithm is performed on an industrial dataset provided by the Advanced Control Technology Group of Eastman Chemical Company. Fig.2 shows the process diagram. The dataset contains 60

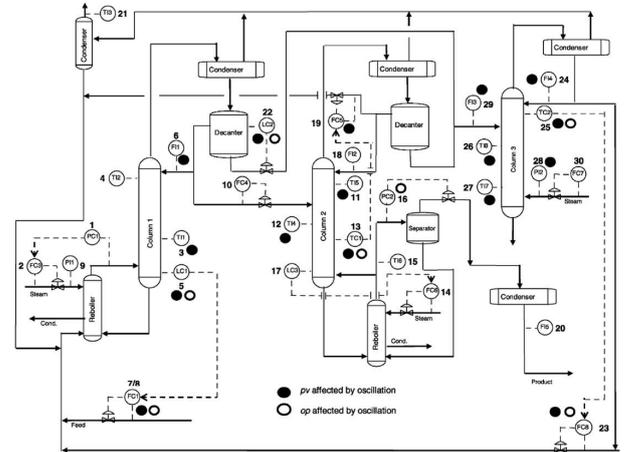


Fig. 2. Process diagram of a plant from Eastman Chemical Company

variables and 8640 samples with a sampling interval of 20 seconds. In addition, there is a plant wide oscillation with a period of nearly 2h in the dataset that has been extensively studied. Several methods have been proposed for oscillation detection (Thornhill and Hägglund (1997); Srinivasan et al. (2007)). To identify the root cause, Thornhill et al. (2003) proposed a non-linearity index and diagnosed the control valve LC2 as the root cause. The plant-wide oscillation disappeared after repairing the valve, which further

confirms the root cause identified by the index. Yuan and Qin (2014) diagnosed the same root cause and propagation path by using Granger causality analysis.

Since oscillating signals are one of the most predictable types of signals, it would be interesting to test DiCCA algorithm on the TEP oscillating dataset. In our analysis 18 variables with the approximate 2h cycle oscillations are selected as the variables for DiCCA modeling, which are LC1.PV, LC1.OP, FC1.SP, FC1.PV, FC2.SP, FC2.PV, TI4.PV, TC1.PV, TC1.OP, FC5.OP, FC5.PV, LC2.PV, LC2.OP, FC8.SP, FC8.PV, TC2.OP, TI8.PV, FI3.PV. The dynamic order  $s$  is chosen to be 28, such that the residuals of the dynamic AR models are essentially white. All 18 latent variables are calculated, the first 3 are selected to shown in Fig.3. From Fig.3, we can see that the first

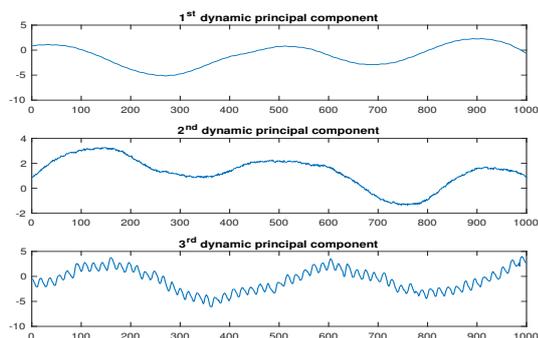


Fig. 3. First 1000 points of the first 3 dynamic principal components

two dynamic principal components contain low frequency oscillations only, while the third dynamic principal component contains both low frequency oscillations and high frequency oscillations. The low-frequency oscillation was studied by Thornhill et al. (2003) and Yuan and Qin (2014), with both reporting that the LC2.SP and LC2.OP as the root cause. Yuan and Qin (2014) conducted Granger causality to clearly identify that the LC2 control loop is causing the low frequency oscillation. However, the high frequency oscillations have not been studied in previously reported work to the best of our knowledge.

To check the predictability of each dynamic latent variable, Fig.4 shows the bar plots of  $R^2$  and the variance captured by each dynamic latent variable.

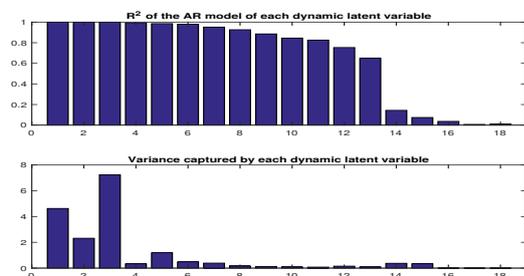


Fig. 4.  $R^2$  and variance captured by each dynamic latent variable

From Fig.4, we can see that the first three principal components have similar  $R^2$  values that are very close

1. In addition, the  $R^2$  decreases monotonically as more latent variables are extracted, which is consistent with the objective of DiCCA. Furthermore,  $R^2$  drops below 0.2 after 13 principal components are extracted, which indicates that no more than 13 dynamic components are required to exhaust the dynamics in all 18 variables.

In addition, the oscillation periods of the low frequency oscillations and high frequency oscillations can be identified based on DiCCA results. To identify the oscillation periods of the low frequency oscillation, a fast Fourier transformation (FFT) is performed on the first dynamic component. Fig.5 shows the result. The period of low frequency oscillation can be identified by the peak location of FFT, which is -5.886 in log scale. This leads to a oscillation period of  $1/e^{-5.886} \approx 357$  sampling intervals. This oscillation period is consistent with the previous analysis in the literature.

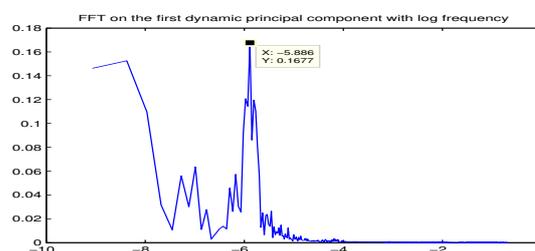


Fig. 5. FFT on the first dynamic component

To identify the period of high frequency oscillation, one option is to apply FFT on the third dynamic component directly. Ideally, we should be able to detect 2 peaks in the FFT plots, one corresponding to the low frequency oscillation and one corresponding to the high frequency oscillation. However, in practice, the large magnitude of low frequency oscillations and noise can cover the peak of the high frequency oscillations, making it hard to detect. To remove the effect of low frequency oscillations, a high pass filter is applied to the third dynamic component first, and FFT is performed on the filtered dynamic component. Fig.6 shows the third dynamic component after filtering, and Fig.7 is the FFT result.

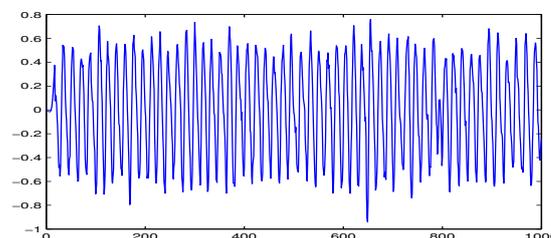


Fig. 6. First 1000 points of the filtered third dynamic component

It can be seen from Fig.6 that low frequency oscillations are successfully removed by the high pass filter. The period of high frequency oscillation can be identified by the peak location in Fig.7. The peak value occurs at 0.05664 in Fig.7, corresponding to a oscillation period of  $1/0.05664 \approx 18$  sampling intervals, which is 6 minutes.

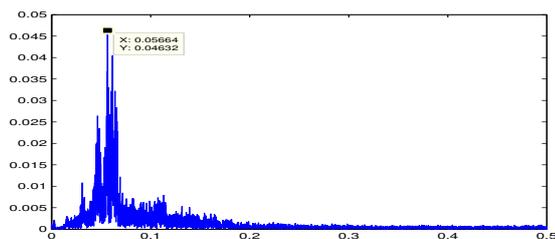


Fig. 7. FFT on the filtered third dynamic component

To further analyze the root cause of the high frequency oscillation, 25 variables containing high frequency oscillations are selected from the dataset, which are TI5.PV, TI4.PV, TC1.PV, TC1.OP, FC6.SP, FC6.PV, FC6.OP, TI6.PV, PC2.PV, PC2.OP, LC3.PV, LC3.OP, FC5.SP, FC5.PV, FC5.OP, FC8.SP, FC8.PV, FC8.OP, FI4.PV, TC2.PV, TC2.OP, TI8.PV, TI7.PV, PI2.PV, FC7.OP. Pairwise spectral Granger causality analysis is performed on these 25 variables in the range of high frequency oscillation frequency, and Fig.8 shows causal relationships found by selecting a threshold (threshold = 1 in this case), where the variables in the same circle have the same causal/effect, while variables not shown in Fig.8 have virtually no intact cause-effect relationship with each other or with these 15 variables. It is obvious from the causality graph in

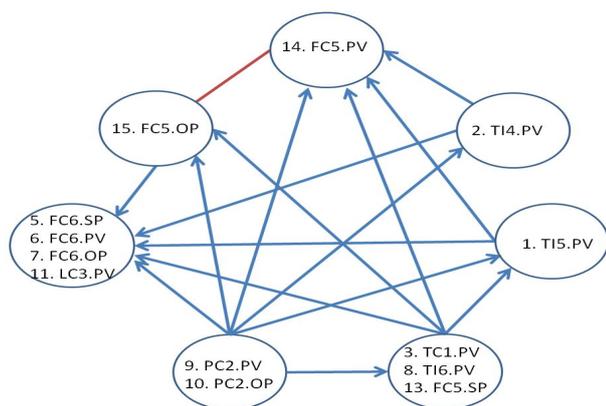


Fig. 8. Causal network by spectral Granger causality analysis

Fig.8 that PC2.PV and PC2.OP are most likely the root cause of high frequency oscillations among variable 1-15. Further, the second group of variables, TC1.PV, TI6.PC and FC5.SP are effect of the PC2 variables as they are physically close to the PC2 control loop, and they further cause other variables to oscillate. Valve stiction or too aggressive tuning in PC2 pressure control loop is likely the physical cause of the high frequency oscillations.

## 6. CONCLUSIONS

In this paper, a dynamic-inner canonical correlation analysis algorithm is proposed and analyzed for dynamic data modeling. DiCCA extracts latent variables in a descending order of predictability, or  $R^2$  values. This guarantees that the most dynamic component is extracted first. In addition, it can be proven that the number of latent variables required to extract all the dynamic components

is smaller than the number of original number of variables. Therefore, DiCCA can be used as an effective tool to extract representative principal time series from high dimensional time series data. After the predictable portion is removed from the data, the prediction error will be essentially white. Therefore, DiCCA can be interpreted as a whitening filter, and traditional PCA can be performed on the prediction error to further model the static components. Case studies on Eastman plant-wide oscillating dataset demonstrates the effectiveness of the proposed method, where the dominating low frequency oscillating components are extracted first. In addition, DiCCA also discovered a high frequency oscillating components that have not been well studied. Granger causality analysis performed on the variables with high frequency oscillations identified the possible root cause effectively.

## ACKNOWLEDGEMENTS

This work was supported in part by the Natural Science Foundation of China (61490704), the Fundamental Research Program of the Shenzhen Committee on Science and Innovations (20160207, 20170155), the Post-doctoral Fellowship Fund of the Chinese University of Hong Kong, Shenzhen, and the Texas-Wisconsin-California Control Consortium.

## REFERENCES

- Dong, Y. and Qin, S.J. (2015). Dynamic-inner partial least squares for dynamic data modeling. *IFAC-PapersOnLine*, 48(8), 117–122.
- Dong, Y. and Qin, S.J. (2017). A novel dynamic pca algorithm for dynamic data modeling and process monitoring. *To appear in Journal of Process Control*.
- Helland, I.S. (1988). On the structure of partial least squares regression. *Communications in statistics-Simulation and Computation*, 17(2), 581–607.
- Jackson, J.E. (2005). *A user's guide to principal components*, volume 587. John Wiley & Sons.
- Joe Qin, S. (2003). Statistical process monitoring: basics and beyond. *Journal of chemometrics*, 17(8-9), 480–502.
- Ku, W., Storer, R.H., and Georgakakis, 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.
- Srinivasan, R., Rengaswamy, R., and Miller, R. (2007). A modified empirical mode decomposition (emd) process for oscillation characterization in control loops. *Control Engineering Practice*, 15(9), 1135–1148.
- Thornhill, N.F., Cox, J.W., and Paulonis, M.A. (2003). Diagnosis of plant-wide oscillation through data-driven analysis and process understanding. *Control Engineering Practice*, 11(12), 1481–1490.
- Thornhill, N.F. and Hägglund, T. (1997). Detection and diagnosis of oscillation in control loops. *Control Engineering Practice*, 5(10), 1343–1354.
- Yuan, T. and Qin, S.J. (2014). Root cause diagnosis of plant-wide oscillations using granger causality. *Journal of Process Control*, 24(2), 450–459.