Root cause diagnosis of plant-wide oscillations using Granger causality

Tao Yuan* and S. Joe Qin**

* Ming Hsieh Department of Electrical Engineering, University of Southern California, Los Angeles, CA 90089 USA (e-mail: tyuan@usc.edu).
** Mork Family Department of Chemical Engineering and Materials Science, and Ming Hsieh Department of Electrical Engineering, University of Southern California, Los Angeles, CA 90089 USA (e-mail: sqin@usc.edu)

Abstract: Oscillations travel along propagation paths and may impact the control performance of the whole plant. This paper presents a data-driven method for diagnosing the root cause of the plant-wide oscillation. The major contribution is the application of combining Granger causality, a statistical method based on linear prediction theory and principal component feature selection to provide a reliable diagnosis of oscillation propagation. Two case studies are used to demonstrate our proposed method. The work presented here may have significant implication for diagnosing of other kind of disturbance propagation.

Keywords: oscillation; diagnosis; propagation; granger causality; spectral analysis.

1. INTRODUCTION

Oscillation is periodic phenomena with well defined amplitude and frequency. Oscillating control loops are very common in industrial processes, which could be due to aggressive controller tuning, valve stiction, external disturbances, etc. Once an oscillation is generated somewhere, it may propagate to the whole plant through mass and heat transfer. This kind of plant-wide oscillation will cause poor control performance, low product quality and excessive energy consumption. It is difficult for process engineers to know where the original source of the plant-wide oscillation is and do further maintenance or adjustment. Therefore, strategies for detecting and diagnosing of plant-wide oscillations are essential to solve this problem.

Several methods have been proposed to deal with root cause diagnosis of plant-wide oscillations. Thornhill et al. [2003] proposed a non-linearity index along with process understanding to find the root cause due to valve stiction. Choudhury et al. [2007] employ changes in controller gain to distinguish an internally generated oscillation from an external oscillatory disturbance. Xia et al. [2005] proposed multi-resolution spectral ICA to detect and isolate the sources of multiple oscillations, a significance index is utilized to indicate possible root cause. Recently, Jiang et al. [2009] incorporate process knowledge into a graph theoretical approach to help diagnosis of root cause. Bauer et al. [2007] proposed a data-driven method to find the propagation direction of disturbance using the idea of transfer entropy.

This paper describes a strategy through the application of Granger causality [Granger, 1969] to analyze the cause and effect relationships in plant-wide oscillation data. The Granger causality is a statistical hypothesis method to determining whether one time series is helpful to forecast another. Recently Granger causality has gained great attention in many areas [Seth, 2005, Hong et al., 2009, Dhamala et al., 2008] to extract useful dynamic information and study the inner causal relationships.

The main idea of this paper is to combine feature selection using a latent variable method with Granger causality both in time-domain and frequency domain to find the root cause of plant-wide oscillations. It is a data-based method and could be implemented during normal operation time without any plant interruption. The rest of this paper is organized as follows. Section 2 introduces the main algorithm of our diagnosing strategy which consists of three parts: time-domain Granger causality, spectral Granger causality and feature selection. In section 3, through a simulated example and an industrial case study, we show how the presented method could successfully lead to the correct root cause. Section 4 summarizes the work and gives conclusions.

2. METHODS AND IMPLEMENTATION

2.1 Time-domain Granger Causality

Granger causality(G-causality) is a measure of causal effect from one time series to another and is based on linear predictions of time series. According to G-causality, we call $X_2$ causes $X_1$ if the inclusion of past observations of $X_2$ reduces the prediction error of $X_1$ in a linear regression model of both $X_1$ and $X_2$, as compared to a model which includes only $X_1$ information.

Time Series Modeling For two time series $X_1(t)$ and $X_2(t)$ from two stationary stochastic processes. We could construct bivariate autoregressive(AR) model:
\[ X_1(t) = \sum_{j=1}^{k} A_{11;j} X_1(t - j) + \sum_{j=1}^{k} A_{12;j} X_2(t - j) + \xi_{12}(t) \]  
(1)
\[ X_2(t) = \sum_{j=1}^{k} A_{21;j} X_1(t - j) + \sum_{j=1}^{k} A_{22;j} X_2(t - j) + \xi_{21}(t) \]  
(2)

Here \( A \) are the AR coefficients, \( \xi \)'s are the prediction errors (residuals) of the model. \( k \) is the model order which defines how many time lag terms will be included in the regression model. Equation (1) and (2) are called unrestricted model or full model. Also we could exclude the cross correlation terms of \( X_1 \) and \( X_2 \) to perform univariate AR modeling on each time series and obtain:
\[ X_1(t) = \sum_{j=1}^{k} B_{11;j} \xi_1(t - j) + \xi_{11}(t) \]  
(3)
\[ X_2(t) = \sum_{j=1}^{k} B_{21;j} \xi_2(t - j) + \xi_{21}(t) \]  
(4)

If variance of \( \xi_{12} \) is smaller than \( \xi_1 \) in some suitable statistical sense, which means prediction of time series \( X_1(t) \) become more accurate when incorporating the past values of \( X_2(t) \). Then there is a causal influence from \( X_2(t) \) to \( X_1(t) \), and it can be quantified in time domain by:
\[ F_{j\rightarrow i} = \ln \frac{\text{var}(\xi_i)}{\text{var}(\xi_{ij})} \]  
(5)
where \( \xi_i \) is derived from regression using \( X_i(t) \) only and \( \xi_{ij} \) is derived when full model is utilized.

It is easy and intuitive to generalize bivariate Granger causality to multivariate case. For a system of \( n \) variables \((1,2, \cdots, n)\), \( X_j \) causes \( X_i \) if knowing \( X_j \) helps to predict \( X_i \) when all other variables are also included in the regression model. The residual covariance matrix of the full model could be written as:
\[ \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} & \cdots & \Sigma_{1n} \\ \Sigma_{21} & \Sigma_{22} & \cdots & \Sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma_{n1} & \Sigma_{n2} & \cdots & \Sigma_{nn} \end{bmatrix} \]  
(6)

Start from the full model, if we exclude a predictor variable \( X_j \) at a time, then a corresponding restricted model could be obtained. Its residual matrix is:
\[ \rho = \begin{bmatrix} \rho_{11} & \cdots & \rho_{1(j-1)} & \rho_{1(j+1)} & \cdots & \rho_{1n} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \rho_{(j-1)1} & \cdots & \rho_{(j-1)(j-1)} & \rho_{(j-1)(j+1)} & \cdots & \rho_{(j-1)n} \\ \rho_{(j+1)1} & \cdots & \rho_{(j+1)(j-1)} & \rho_{(j+1)(j+1)} & \cdots & \rho_{(j+1)n} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{n1} & \cdots & \rho_{n(j-1)} & \rho_{n(j+1)} & \cdots & \rho_{nn} \end{bmatrix} \]  
(7)
with size \((n-1)\) by \((n-1)\)

Therefore, the Granger causality from variable \( j \) to variable \( i \), conditioned on other variables, is measured by:
\[ F_{j\rightarrow i|n} = \ln \frac{\rho_{ii}}{\Sigma_{ii}} \]  
(8)

**Significance Level**

Given G-causality index, statistical significance needs to be established before making an inference on causality relationship. Consider two models, 1 and 2. Model 1 is the restricted model which has \( p_1 \) parameters, and model 2 is the full model which has \( p_2 \) parameters, where \( p_2 > p_1 \). The full model with more parameters will always be able to fit the data at least as well as the restricted model with fewer parameters. Here we want to determine whether model 2 gives a significantly better fit to the data. \( F \) test is applied:
\[ \frac{RSS_1 - RSS_2}{RSS_2/n-p_2} \sim F_{p_2-p_1,n-p_2} \]  
(9)
where \( RSS_i \) is the sum of squares of residual in model \( i \), and \( n \) is total number of observations. When the \( p \)-value is less than the significance level \( \alpha \) (typically 0.01 or 0.05), the null hypothesis that there is no causality could be rejected.

### 2.2 Spectral Granger Causality

Since we want to deal with oscillation phenomenon, it is natural to link it with spectral analysis. The spectral decomposition of Granger’s time-domain causality was proposed by Geweke in 1982 [Geweke, 1982]. Geweke showed that, for the data from one time-series, the power at a specific frequency can be decomposed into an intrinsic part and a part predicted by the data from another time-series. To derive the frequency-domain Granger causality, we start with (1) and (2), after performing Fourier transform yields:
\[ \begin{pmatrix} A_{11}(f) & A_{12}(f) \\ A_{21}(f) & A_{22}(f) \end{pmatrix} \begin{pmatrix} X_1(f) \\ X_2(f) \end{pmatrix} = \begin{pmatrix} E_1(f) \\ E_2(f) \end{pmatrix} \]  
(10)
where the components of the coefficient matrix \([A_{ij}(f)]\) are \( A_{lm}(f) = \delta_{lm} - \sum_{j=1}^{k} A_{jm}(j)e^{-2\pi f j} \). Denote transfer function matrix \( H(f) = [A_{ij}(f)]^{-1} \),
\[ \begin{pmatrix} X_1(f) \\ X_2(f) \end{pmatrix} = \begin{pmatrix} H_{11}(f) & H_{12}(f) \\ H_{21}(f) & H_{22}(f) \end{pmatrix} \begin{pmatrix} E_1(f) \\ E_2(f) \end{pmatrix} \]  
(11)
Thus, the spectral density matrix \( S(f) \) is derived as:
\[ S(f) = \langle X(f)X^* (f) \rangle = H(f)SH^*(f) \]  
(12)
For \( X_1 \) process, left-multiply (11) on both sides with
\[ \begin{pmatrix} 1 \\ -\Sigma_{12}/\Sigma_{11} \end{pmatrix} \]  
(13)
and obtain:
\[ \begin{pmatrix} X_1(f) \\ X_2(f) \end{pmatrix} = \begin{pmatrix} \tilde{H}_{11}(f) & \tilde{H}_{12}(f) \\ \tilde{H}_{21}(f) & \tilde{H}_{22}(f) \end{pmatrix} \begin{pmatrix} E_1(f) \\ E_2(f) \end{pmatrix} \]  
(14)
where \( \tilde{E}_2(f) = E_2(f) - \frac{\Sigma_{12}}{\Sigma_{11}} E_1(f) \), the new transfer function \( \tilde{H}(f) \) become \( \tilde{H}_{11}(f) = H_{11}(f) + \frac{\Sigma_{12}}{\Sigma_{11}} H_{12}(f) \),
\[ \tilde{H}_{12}(f) = H_{12}(f), \quad \tilde{H}_{21}(f) = H_{21}(f) + \frac{\Sigma_{12}}{\Sigma_{11}} H_{11}(f), \quad \text{and} \quad \tilde{H}_{22}(f) = H_{22}(f) \]. The spectrum of \( X_1(t) \) is represented into two parts:
\[ S_{11}(f) = \tilde{H}_{11}(f)\Sigma_{11}\tilde{H}_{11}^*(f) + \tilde{H}_{12}(f)\Sigma_{22}\tilde{H}_{21}^*(f) \]  
(15)
where \( \Sigma_{22} = \Sigma_{22} - \frac{\Sigma_{12}}{\Sigma_{11}} \tilde{H}_{11}(f) \). Therefore, the spectral Granger causality from \( X_2 \) to \( X_1 \) at frequency \( f \) is:
\[ I_{2\rightarrow 1}(f) = \ln \frac{S_{11}(f)}{S_{11}(f) - \left( \Sigma_{22} - \frac{\Sigma_{12}}{\Sigma_{11}} \right) |H_{12}(f)|^2} \]  
(16)
which could be interpreted in this way: at a given frequency, how many power of \( X_1 \) could be explained or predicted by history of \( X_2 \).

It is easy to find the plant-wide oscillation frequency through spectrum plot of time-series data. This is will define a frequency range which we most cared about. Then, we could choose a subset of variables which are candidate root cause, and perform pair-wise spectral Granger causality. If some variable has strong causal effect on other loops and receive negligible contribution from others, then we could infer it to be the root cause of the whole plant oscillation.

2.3 Feature Selection

Feature selection, also known as variable selection, is the technique of selecting a subset of relevant features for building models. In our work, we would like to locate the plant-wide oscillation source by analyzing the causality network between different variables. In a chemical plant, there are many process variables and controller output, some of them are irrelevant to the problem we want to diagnose and can not well represent the oscillation causality path. We would like to pick up a subset of variables that together have good predictive power of the oscillation. Building such a feature representation is an opportunity to incorporate process understanding and knowledge into the data and can be specific to a particular application. Feature selection can reduce dimensionality effectively, improve result comprehensibility and provide a better understanding of the underlying process that generates the data. Therefore, we make feature selection be a preprocessing step to the causality analysis for oscillation.

There are a number of feature selection methods, including clustering, linear transformation, wavelet or convolutions of kernels, etc. In our case, we want to eliminate those variables that are irrelevant to the plant-wide oscillation.

We have made such an assumption: if a variable does not show significant power during oscillation frequency range that we are interested to diagnose, then it can not be the source of the corresponding plant-wide oscillation.

Principal Component Analysis (PCA) has been widely used for treating multivariate data in various areas such as process monitoring, quality control and fault diagnosis. In this section, the general principle of using PCA as a feature selection method will be introduced. Consider a data matrix \( X \in \mathbb{R}^{n \times m} \) consisting of \( n \) observation rows and \( m \) variable columns. The matrix \( X \) can be decomposed into a score matrix \( T \) and a loading matrix \( P \) as follows,

\[
X = T P^T + E
\]

where \( E \) is the residue matrix. Each column of \( T \) is termed principal component (PC). Each PC can be decomposed into a linear combination of a set of original variables, i.e.,

\[
l_i = x_1 p_{i1} + x_2 p_{i2} + \cdots + x_j p_{ij} + \cdots + x_n p_{in}
\]

Thus we can use \( p_{i2}^2 \) to measure the contribution from any variable to a specified PC. With dimension reduced, we will get several principal components (PC) to represent the raw data sets. Some PCs will carry the plant-wide oscillation pattern and we choose those variables which make strong contribution to those oscillating PCs as features. The following oscillation significance index (\( OSI \)) is computed to rank the contribution:

\[
OSI_i = \sum_{j \in \text{oscpc}} \lambda_j p_{ij}^2
\]

where \( \lambda_j \) is the eigenvalue of corresponding PC which shows whole plant oscillation. Based on empirical experience, variables with \( OSI \) index larger than 1/3 of the \( \max(OSI) \) are selected to be features and will be included in the causality regression model to do further analysis.

3. CASE STUDIES

3.1 Simulated 4 \times 4 System

Consider the four-input, four-output system with a common white noise disturbance. The open-loop process transfer function matrix \( G(q^{-1}) \) and disturbance model are given as follows

\[
G(q^{-1}) =
\begin{bmatrix}
0.05 q^{-3} & 0 & 0.7 q^{-3} & 0 \\
1 - 0.95 q^{-1} & 0.02966 q^{-3} & 0 & 0 \\
1 - 1.62 q^{-1} + 0.76 q^{-2} & 1 - 0.934 q^{-1} & 0.35 q^{-3} & 0.5 q^{-2} \\
0 & 0.5 q^{-3} - 0.4875 q^{-6} & 1 - 0.5 q^{-1} + 0.2q^{-2} & 0.2 q^{-5} \\
0 & 0 & 1 - 0.8 q^{-1} & 1 - 0.8 q^{-1}
\end{bmatrix}
\]

\[
N(q^{-1}) =
\begin{bmatrix}
1 - 0.1875 q^{-1} & 0 & 0 & 0 \\
1 - 0.9875 q^{-1} & 0 & 0 & 0 \\
0 & 1 - 0.1875 q^{-1} & 0 & 0 \\
0 & 0 & 1 - 0.9875 q^{-1} & 0 \\
0 & 0 & 0 & 1 - 0.1875 q^{-1} \\
0 & 0 & 0 & 1 - 0.9875 q^{-1}
\end{bmatrix}
\]

\[
K_c = [0.816 \ 0.625 \ 0.184 \ 0.37]
\]

In this system, four discrete PI controllers are utilized with from \( K_c \{1 + (\frac{\Delta T}{T_c})|\frac{\Delta T}{T_c}\} \), where \( K_c \) is the proportional gain, \( \Delta T \) is the controller sampling time and \( T_c \) is the integral time. When we set \( K_c = [0.816 \ 0.625 \ 0.184 \ 0.37] \) and \( T_c = [20 16 286 5] \), the four loops are well controlled and there is no obvious oscillation, which implies no aggressive tuning oscillation problem. Then, a valve stiction model is added into the loop 3 in order to generate the oscillation phenomena. The time series plot of the four controlled variables (CV) together with corresponding autocorrelation plot is shown in Fig. 1(a) and Fig. 1(b).

It is easy to see that there exists a common oscillation pattern in several loops. Since the number of variables is not large, feature selection process is omitted. Time-domain Granger causality method is applied on the total four controlled variables. Fig. 2 shows the time-domain causality map of the 4 variables. In this map, each node represents one variable and if there is a significant large causal relationship from one variable to another, then a directed edge with arrow will represent it. Possible root cause suggested by this map is CV3, meanwhile, an oscillation propagation path could be recovered which starts from CV3 to CV1 then affect the CV4. CV2 is not in this path which is consistent with the information we could tell from Fig. 1(b) the autocorrelation plot, where CV2
Given such a real data set, feature selection method is implemented first to select the most relevant variables be a subset for further causality analysis. There is an assumption here, if a variable does not show significant power at the common oscillation frequency, it is not likely to be a root cause of oscillation. There are also significant causal relationships between CV1, CV2 and CV4, which form a causal feedback structure. Both the time-domain and frequency domain causality method lead to the same root cause which is same as our experimental design.

3.2 Industrial application

Fig. 4 shows a process schematic diagram from Eastman Chemical Company, USA. The advanced control technology group had identified a need to diagnose a common oscillation of about 2 hours (320 samples/cycle). Circled marked variables are all affected by this plant-wide oscillation and show similar pattern. FC, LC, PC, TC represent flow, level, pressure and temperature, which are controlled variables. 15 controlled variables along with controller output (OP) records are available. The sampling interval is 20 seconds.

Fig. 4. Process schematic of a plant from Eastman Chemical Company

has dominant prediction power to others at this frequency and is likely to be the root cause of oscillation. There are also significant causal relationships between CV1, CV2 and CV4, which form a causal feedback structure.

Both the time-domain and frequency domain causality method lead to the same root cause which is same as our experimental design.

3.2 Industrial application

Fig. 4 shows a process schematic diagram from Eastman Chemical Company, USA. The advanced control technology group had identified a need to diagnose a common oscillation of about 2 hours (320 samples/cycle). Circled marked variables are all affected by this plant-wide oscillation and show similar pattern. FC, LC, PC, TC represent flow, level, pressure and temperature, which are controlled variables. 15 controlled variables along with controller output (OP) records are available. The sampling interval is 20 seconds.

Given such a real data set, feature selection method is implemented first to select the most relevant variables be a subset for further causality analysis. There is an assumption here, if a variable does not show significant power at the common oscillation frequency, it is not likely

Fig. 4. Process schematic of a plant from Eastman Chemical Company

has dominant prediction power to others at this frequency and is likely to be the root cause of oscillation. There are also significant causal relationships between CV1, CV2 and CV4, which form a causal feedback structure. Both the time-domain and frequency domain causality method lead to the same root cause which is same as our experimental design.

3.2 Industrial application

Fig. 4 shows a process schematic diagram from Eastman Chemical Company, USA. The advanced control technology group had identified a need to diagnose a common oscillation of about 2 hours (320 samples/cycle). Circled marked variables are all affected by this plant-wide oscillation and show similar pattern. FC, LC, PC, TC represent flow, level, pressure and temperature, which are controlled variables. 15 controlled variables along with controller output (OP) records are available. The sampling interval is 20 seconds.

Given such a real data set, feature selection method is implemented first to select the most relevant variables be a subset for further causality analysis. There is an assumption here, if a variable does not show significant power at the common oscillation frequency, it is not likely

Fig. 4. Process schematic of a plant from Eastman Chemical Company

has dominant prediction power to others at this frequency and is likely to be the root cause of oscillation. There are also significant causal relationships between CV1, CV2 and CV4, which form a causal feedback structure. Both the time-domain and frequency domain causality method lead to the same root cause which is same as our experimental design.

3.2 Industrial application

Fig. 4 shows a process schematic diagram from Eastman Chemical Company, USA. The advanced control technology group had identified a need to diagnose a common oscillation of about 2 hours (320 samples/cycle). Circled marked variables are all affected by this plant-wide oscillation and show similar pattern. FC, LC, PC, TC represent flow, level, pressure and temperature, which are controlled variables. 15 controlled variables along with controller output (OP) records are available. The sampling interval is 20 seconds.

Given such a real data set, feature selection method is implemented first to select the most relevant variables be a subset for further causality analysis. There is an assumption here, if a variable does not show significant power at the common oscillation frequency, it is not likely

Fig. 4. Process schematic of a plant from Eastman Chemical Company

has dominant prediction power to others at this frequency and is likely to be the root cause of oscillation. There are also significant causal relationships between CV1, CV2 and CV4, which form a causal feedback structure. Both the time-domain and frequency domain causality method lead to the same root cause which is same as our experimental design.

3.2 Industrial application

Fig. 4 shows a process schematic diagram from Eastman Chemical Company, USA. The advanced control technology group had identified a need to diagnose a common oscillation of about 2 hours (320 samples/cycle). Circled marked variables are all affected by this plant-wide oscillation and show similar pattern. FC, LC, PC, TC represent flow, level, pressure and temperature, which are controlled variables. 15 controlled variables along with controller output (OP) records are available. The sampling interval is 20 seconds.

Given such a real data set, feature selection method is implemented first to select the most relevant variables be a subset for further causality analysis. There is an assumption here, if a variable does not show significant power at the common oscillation frequency, it is not likely

Fig. 4. Process schematic of a plant from Eastman Chemical Company

has dominant prediction power to others at this frequency and is likely to be the root cause of oscillation. There are also significant causal relationships between CV1, CV2 and CV4, which form a causal feedback structure. Both the time-domain and frequency domain causality method lead to the same root cause which is same as our experimental design.

3.2 Industrial application

Fig. 4 shows a process schematic diagram from Eastman Chemical Company, USA. The advanced control technology group had identified a need to diagnose a common oscillation of about 2 hours (320 samples/cycle). Circled marked variables are all affected by this plant-wide oscillation and show similar pattern. FC, LC, PC, TC represent flow, level, pressure and temperature, which are controlled variables. 15 controlled variables along with controller output (OP) records are available. The sampling interval is 20 seconds.

Given such a real data set, feature selection method is implemented first to select the most relevant variables be a subset for further causality analysis. There is an assumption here, if a variable does not show significant power at the common oscillation frequency, it is not likely
to be the source. 15 controlled variables along with their controller output are constructed in a whole data matrix with 8640 samples. After performing PCA, five principal components (PC) are shown in Fig. 5. PC1 and PC3 carry the 2 hours oscillation pattern, thus we calculate the oscillation significance index from each variable to these two oscillating PC.

Fig. 6 is a bar plot showing the OSI of 30 variables. The horizontal red line marks the threshold which is chosen to be the 1/3 of the maximum OSI in this group. The 15 variables with OSI larger than the threshold are selected to be features to do further causality analysis.

Time-domain causality method is applied to measure the cause-effect relationship between each feature. Model order of the vector AR model is set to be 4 suggested by Bayesian Information criterion. Fig. 7 is the causality network graph, the redline connecting two nodes without an arrow indicates there is a causality from each other, which is called causality feedback. From this map, there are several important information we could dig out. First, variable 10 and 11 have largest positive value of flow, thus they are regarded as sources. 8 variables in the total 15 features are chosen to apply pairwise spectral Granger causality method. They are variable with tag 2, 3, 4, 6, 8, 11, 13 and 14, representing all the important oscillating control loops. Fig. 9 is the similar spectral causality graph we mentioned early in this section. The subgraph with title "11→others" has many strong peaks at frequency around 0.003 which means variable 11 has dominant prediction power and causal effect to other variables at this frequency. The red dashed line in other subgraphs measures the causality from other variables to variable 11, which are all very small and can be negligible. Furthermore, in the subgraph "11→others", three black dotted line each forms a strong peak implies another oscillation propagation path from variable 11 to variable 2, 3, 4 (LC1 and FC1 cascade feedback structure), which has not been recovered in Fig. 7 using time-domain causality method. Therefore, variable 11 is inferred to be the root cause of oscillation with 0.003 normalized frequency.
Oscillations in industrial plant often propagate from one unit to other units. It is of interest to extract the information flow from the process operational data and diagnose the root cause by analyzing the routing operation data. In this paper, the concept of Granger causality has been reviewed, which reveals the cause and effect relationship in terms of prediction. A novel data-driven method for diagnosis of plant-wide oscillations is presented. This method combines time-domain Granger causality, spectral Granger causality along with PCA based feature selection to locate the source of oscillation propagation and explore related dynamics. Time-domain causality is better at recovering the propagation path while spectral causality helps to analyze pair-wise causality at a specified frequency range. Feature selection serves as a bridge to connect and explain Granger causality and the oscillation cause-effect relationship. This method is data-based and does not require process models, also there will be no perturbation to the original plant when implementing the method. Two case studies have shown the effectiveness and accuracy of the new strategy.

5. ACKNOWLEDGMENTS

The authors are grateful for the financial support from China Scholarship Council (CSC) and the Center for Interactive Smart Oilfield Technologies (CISOFT), and thanks to Eastman Chemical Company for providing the industrial data.

REFERENCES


