

Convergent Cross Mapping (CCM) based Approach for Isolating the Source of Plant-wide Disturbances

Muhammad Faisal Aftab¹, Morten Hovd¹ and Selvanathan Sivalingam²

Abstract—Disturbances originating in one control loop of a large industrial plant can propagate far from the source, giving rise to plant-wide oscillations. The underlying interactions among the different control loops make it hard to identify the origin of such large scale disturbances. This paper studies the application of the convergent cross mapping (CCM) based technique to isolate the source of a plant-wide disturbance. The proposed scheme exploits the cause and effect relationships among the affected variables to find the source of disturbance. The states of the causative factors are estimated from the effect variable and the directionality of information flow is established using the correlation between the original and estimated signal. The method is applied to the industrial case study and is shown to be effective in isolating the disturbance origin.

I. INTRODUCTION

In an industrial control system plant-wide disturbances can result in product variability and excessive use of precious resources, with both economically and environmentally detrimental consequences. Thus there is clear motivation to look for the root cause/source of these disturbances, so that targeted remedial action can be taken. Disturbances in a large scale plant may arise due to number of reasons such as a) poor controller tuning b) process degradation, and c) equipment wear and failure, etc. Once started these disturbances travel away from their origin due to mass and energy flows as well as control loop interactions, thus giving rise to plant wide disturbances [7],[20]. In such a scenario, the early and accurate identification of the origin of the disturbance is a key to reduce maintenance and shut down time and thus improving the overall economics.

In order to find the source of plant wide oscillations different approaches, both model based and data driven, have been considered by researchers over the past decade or so. The data driven methods are more popular owing to the fact that model based methods rely on process information/model, expert knowledge and analysis of P&IDs that may not be available or lack desired level of accuracy [8]. whereas data driven methods are based on the historical data alone and are quite flexible to use. The detailed description of both these approaches can be found in [8].

The data driven approaches are mostly based on the analysis of cause and effect relationship, the so called causality analysis. The causality is defined as: *if prediction of time series y is improved by using the knowledge of other time*

series x , then x has a causal influence on y [8]. This definition of causality is taken up by Granger [10] who argued that x "Granger causes (G-Causes)" y if removing x from the universe of all possible causative variables U decreases the predictability of y i.e $\sigma^2(y|U) < \sigma^2(y|(U - x))$. The key assumptions in Granger Causality (GC) paradigm are a) The cause occurs prior to effect and b) The causative information of a cause variable is independently unique to that variable, the so called "separability" assumption i.e if x G-causes y then information about x is uniquely present in x only and can be removed from the system. The later assumption may be difficult to ensure as in dynamic systems the information about x may be redundantly present in effect variable y and impossible to separate [17]. The GC has been applied for diagnosis of plant wide oscillations by Yuan *et. al*[22]. Moreover, though GC is a powerful concept and has both time and frequency domain variants, it is originally applicable to purely linear stochastic systems only, although extensions to nonlinear systems have been proposed [4].

Bauer *et. al* [2] have studied the determination of disturbance propagation path in process systems by employing the "Transfer Entropy" (TE) methodology to measure the extent of information transfer from one variable to other. The method requires tuning of certain parameters like prediction horizon, time interval and embedding dimensions that can significantly effect the results [8]. Moreover, estimation of joint and conditional probability density functions (pdfs) is computationally intensive and needs sufficiently large data sets.

Phase space reconstruction of the attractor manifold based on time delayed embedding is an important and popular concept for analysis of nonlinear dynamical systems. The time series $x(t)$ from a dynamical system is used for reconstruction of the attractor manifold, with embedding dimension m , using the time lagged vectors $\mathbf{x} = [x(t) \ x(t - \tau) \ x(t - 2\tau) \ \dots \ (t - (m - 1)\tau)]$; where τ is the time lag [11], [18], [16] and [12]. Interdependence between two time series $x(t)$ and $y(t)$ using the "nearest neighbors" (NN) from their respective attractor manifolds has been used in number of studies for example [14], [6],and [1]. The similar approach has been adopted by Bauer *et. al* [3] for diagnosis of plant wide disturbances. This method is also sensitive to tuning parameters like embedding dimension, prediction horizon and number of nearest neighbors.

In this paper we use a relatively new method called Convergent Cross Mapping or CCM [17], to locate the source of plant wide disturbances. The method is also based on the nearest neighbor concept but it runs counter to

¹Department of Engineering Cybernetics, Norwegian University of Science and Technology (NTNU), Norway
muhammad.faisal.aftab; morten.hovd@ntnu.no

²Siemens AS, Trondheim, Norway

the popular notion of causality. The method exploits the fact that if $x(t)$ drives $y(t)$ then the effect variable $y(t)$ will contain signatures of $x(t)$, rather than the other way around. The procedure is rather simple and is claimed to work well with the short time series data as well [17]. The extent of the causal influence is determined by estimating the states \hat{x} from near neighbors of y and vice versa. The correlation between estimated and original time series then gives the directionality of information flow. Moreover, the proposed method is augmented with automatic determination of embedding dimension to make it more robust and reliable. The proposed method is applied to an industrial data set suffering from plant wide oscillations and is able to identify the source of disturbance.

The paper is organized as follows. Section II gives the detailed description of the proposed CCM based source detection method and Section III outlines the statistical test to check the significance of the correlation between estimated and original time series. Automatic determination of embedding dimensions is presented in Section IV. Section V describes the steps involved in the proposed method for identifying the root source followed by an industrial case study and conclusions in Sections VI and VII respectively.

II. CONVERGENT CROSS MAPPING (CCM)

Convergent Cross Mapping (CCM) is recently developed by Sugihara *et. al* [17] for the causality analysis of ecological time series. The method make use of the fact that time series from the same dynamical system share a common attractor manifold and hence can be used to estimate the state of each other. In CCM causal relationship $x \rightarrow y$ is established by looking at the signatures of causative factor x in the effect variable y and in case of $x \rightarrow y$, y can give reliable estimates of x . In this way it runs counter to the general notion of causality where the estimation is other way around.

In order to explain the CCM consider two time series $x(t)$ and $y(t)$, each with N samples, from a dynamical system that share a common attractor manifold \mathbf{M} . Making use of the time delayed embedding (Takens [18]) shadow attractor manifolds \mathbf{M}_x and \mathbf{M}_y , with embedding dimension m , can be constructed from the time series $x(t)$ and $y(t)$ respectively. The time delayed embedding for manifold \mathbf{M}_x is represented as $\mathbf{x}_i = [x(i) \quad x(i - \tau) \quad x(i - 2\tau) \dots x(i - (m - 1)\tau)]$ $\forall i = (m - 1)\tau + 1, (m - 1)\tau + 2, \dots N$; Construction of \mathbf{M}_y follows the same procedure.

CCM looks for how well the local neighborhood of \mathbf{M}_x maps to the local neighborhood of \mathbf{M}_y . In order to determine the causative influence of $x(t)$ on $y(t)$, for each vector i in \mathbf{M}_y , $m + 1$ nearest neighbors are identified and their corresponding entries are marked in $x(t)$. The weighted mean of these these $m + 1$ values in $x(t)$ are used to estimate the state of $x_i(t)$ given be $\hat{x}_i(t)|\mathbf{M}_y$; given by

$$\hat{x}_i = \sum_{j=1}^{m+1} w_j x_j \quad (1)$$

The estimate of $\hat{y}|M_x$ can be calculated in an analogous way to infer the causative influence of $y(t)$ on $x(t)$. The root mean square error ($e_{\hat{x}|M_y}$) and correlation ($\rho_{\hat{x}|M_y}$), given in (2), can then be used to determine the extent of causative influence of $x \rightarrow y$.

$$\rho_{\hat{x}|M_y} = \frac{\sum \hat{x}_i x_i - N^* \bar{x}_i \bar{\hat{x}}_i}{\sqrt{(\sum x_i^2 - N^* \bar{x}^2)} \sqrt{(\sum \hat{x}_i^2 - N^* \bar{\hat{x}}^2)}} \quad (2)$$

$$e_{\hat{x}|M_y} = \frac{1}{N^*} \left(\sum_{k=1}^{k=N^*} \hat{x}_k |M_y - x_k \right)^2$$

where N^* is the number of embedded vectors. The greater the correlation $\rho_{\hat{x}|M_y}$ (lower *rms* error), the greater is the causative influence of x on y and vice versa. The extent of causative influence of y on x can be determined similarly. In the case of unidirectional causality from x to y , $x(t)$ contains no information about $y(t)$ and thus cannot reliably estimate the states of $y(t)$ giving lower values of correlation and higher *rms* error. The correlation coefficient, as calculated in (2), is used in this work to determine the direction of information flow. The correlation coefficient is the preferred choice as it can be transformed to normally distributed z -scores using Fisher's z -transformation ([9], [15]) and can thus be checked for significance against some null hypothesis. The significance test will be explained later in section III.

A. Convergence in CCM

Another important aspect of CCM is the convergence of ρ and *rms* error with increase in length of time series N . By convergence we mean that ρ increases (e_{rms} decreases) as length increases. This is so because with increased information the trajectories forming the attractor fill in, thus giving closer neighbors and higher correlation (lower *rmse* error) [17]. The convergence will be limited by observational error, noise and time series length¹. Nevertheless the convergence of the estimates is a key to establish the directionality of information flow. In case of bidirectional coupling, apart from the absolute value of the correlation ρ , its convergence rate can also be used to judge the extent of causation. The greater the causality effect the greater will be the convergence rate and vice versa. This can be explained from the following illustrative example of CCM.

B. Illustrative Example

The interpretation of results from CCM analysis can be explained using the following example from [17]. Consider two interacting time series given by

$$\begin{aligned} x(k+1) &= x(k)[3.8 - 3.5x(k) - \beta_{xy}y(k)] \\ y(k+1) &= y(k)[3.8 - 3.8y(k) - \beta_{yx}x(k)] \end{aligned} \quad (3)$$

where β_{xy} (β_{yx}) controls the contribution of $y(k)$ ($x(k)$) on $x(k)$ ($y(k)$).

¹In perfect deterministic and noise free settings $\rho \rightarrow 0$ as length $L \rightarrow \infty$ [17]

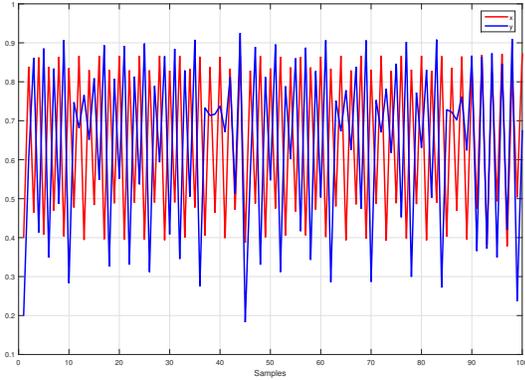


Fig. 1. Time series x and y for CCM illustrative example, $\beta_{yx} = 0.1, \beta_{xy} = 0.02$

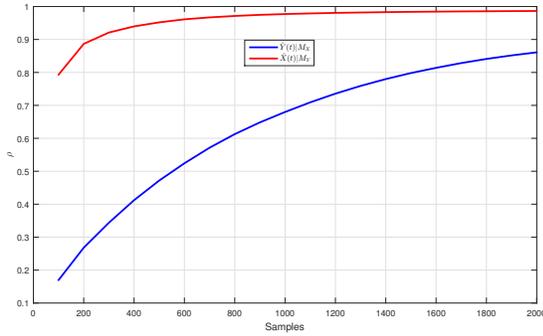


Fig. 2. CCM results (illustrative example)

Time series and results of CCM analysis for the case $\beta_{yx} > \beta_{xy}$ are shown in Figure 1 and Figure 2 respectively. It has been shown by Sugihara *et. al* [17] that the causal relationship in this example cannot be captured by Granger causality as the "separability" assumption cannot be fulfilled. The estimated state of x has greater correlation with the original time series and $\hat{x}|M_y$ converges faster than $\hat{y}|M_x$ thus confirming that x has larger influence on y .

C. Convergent Cross Mapping Algorithm

The causation detection from $x \rightarrow y$ using the CCM method is summarized in the following steps [17]

- 1) Consider two time series $x(t)$ and $y(t)$ of length N
- 2) Construct time delayed embedding vectors $Y_i(t)$, the shadow manifold M_y , for time series $y(t)$ with given embedding dimension m and time delay τ using

$$Y_i = [y(i), y(i - \tau), y(i - 2\tau), \dots, y(i - (m - 1)\tau)] \quad (4)$$

where $i = (m - 1)\tau + 1, (m - 1)\tau + 2, \dots, N$

- 3) For each vector Y_i find the $m + 1$ nearest neighbors on the manifold M_y denoted by $n_j \forall j = 1, \dots, (m + 1)$
- 4) Using the nearest neighbors $n_j \forall j = 1, \dots, (m + 1)$ from M_y manifold locate the corresponding values in $x(t)$ and mark them $x_j, x_{j+1}, \dots, x_{m+1}$

- 5) Generate cross mapped estimate of $x(t)$ given by $\hat{x}|M_y$ from weighted mean of x_{j_s} using

$$\begin{aligned} \hat{x}_i &= \sum_{j=1}^{m+1} w_j x_j \\ w_j &= \frac{w_j}{\sum w_k} \quad k = 1, \dots, m + 1 \\ w_j &= \exp\left(\frac{-\|Y_i - Y_j\|}{\|Y_i - Y_{n_1}\|}\right) \end{aligned} \quad (5)$$

where $\|Y_i - Y_j\|$ and $\|Y_i - Y_{n_1}\|$ are the Euclidean distance in M_y between i^{th} and j^{th} vector and nearest neighbor respectively. In case distance to nearest neighbor is zero then $w_1 = 1$ and $w_j = 0 \forall j = 2 : m + 1$

- 6) Calculate the correlation coefficient $\rho_{\hat{x}|M_y}$ using Equation 2.
- 7) Repeat Steps 1-6 for different time series length N to check for convergence of $\rho_{\hat{x}|M_y}$ as a function of time series length.

The cross mapping from x to y can be evaluated similarly. the directionality or the strength of causality can be determined from the difference in $\rho_{\hat{x}|M_y}$ and $\rho_{\hat{y}|M_x}$. If $\rho_{\hat{x}|M_y} > \rho_{\hat{y}|M_x} \implies x \rightarrow y$ and vice versa.

III. SIGNIFICANCE TEST

It is discussed in the preceding section that the directionality or the extent of causative effects can be determined from the relative value of correlation (ρ) between the original and estimated time series. However, one should question whether the difference between $\rho_{\hat{x}|M_y}$ and $\rho_{\hat{y}|M_x}$ is significant to reliably infer the causal relationship and direction of information flow. The same is true for inferring the convergence of the $\rho_{\hat{x}|M_y}$ and $\rho_{\hat{y}|M_x}$. To this end, it is necessary to test the null hypothesis

$H_0 ::$ There is no significant difference in correlations i.e. $\rho_{\hat{x}|M_y} = \rho_{\hat{y}|M_x}$

The sample correlation doesn't follow the normal distribution so in order to test the null hypothesis, i.e., the significance between the relative correlation, Fisher's z -transformation is used. Fisher's z -transformation maps the correlation coefficient, to normally distributed z -scores by the relation.

$$z_{\hat{x}|M_y} = 0.5 * \ln \left[\frac{1 + \rho_{\hat{x}|M_y}}{1 - \rho_{\hat{x}|M_y}} \right] \quad (6)$$

where \ln is the natural logarithm. Similarly $z_{\hat{y}|M_x}$ can be calculated from $\rho_{\hat{y}|M_x}$. The difference between z -transformed correlation coefficients can be given by

$$Z^* = \frac{z_{\hat{y}|M_x} - z_{\hat{x}|M_y}}{\sqrt{\frac{1}{N_1 - 3} + \frac{1}{N_2 - 3}}} \quad (7)$$

where N_1 and N_2 are sample sizes for two correlations. In this case we have $N_1 = N_2 = N^*$. The NULL hypothesis is rejected at the confidence level α , for a two tailed-test, if $|Z^*| > Z_{\alpha/2}$. A confidence level of $\alpha = 0.01$ is used in this work, which gives $Z_{\alpha/2} = 2.58$.

A. Significance of Convergence

In order to test convergence of the correlation, between original and estimated time series, the CCM is performed for increasing lengths of time series N_{min} to N_{max} . The convergence is inferred if difference between $\rho_{N_{min}}$ and $\rho_{N_{max}}$, using Equations (6) and (7), is found to be significant. The only difference would be to plug in $\rho_{N_{min}}$ and $\rho_{N_{max}}$ to calculate z -scores in Equation (6) and finding whether the corresponding Z^* gives $|Z^*| > Z_{\alpha/2}$ for confidence level α .

IV. DETERMINING EMBEDDING DIMENSION

The approach used in this work is based on the phase space reconstruction of attractor manifolds \mathbf{M}_x and \mathbf{M}_x via delayed embedding. An important parameter in this approach the embedding dimension of the attractor. The embedding dimension is to be chosen such that the attractor unfolds enough to describe the dynamic behavior of the underlying dynamical system. Moreover, if the embedding dimension m is less than the required minimum m_0 i.e $m < m_0$, then the points appearing as near neighbor of a point in state space may not be true neighbors in actual state space [13]. They only appear near neighbors in low dimensional phase space because the attractor may not have unfolded fully.

Therefore any analysis based on the nearest neighbor approach is prone to error. This is true for the work presented here and other related methods such as given in [3]. The adverse impact of choosing a larger embedding dimension than required is the increased computational effort without improving the result much. Thus, it is quite essential to have an automatic way of determining the minimum embedding dimension to make any method based on phase space reconstruction reliable with minimum computational effort. In order to address this concern an automatic method for determining the minimum embedding dimensions based on false near neighbors is incorporated with the CCM based method to make it robust and effective.

The minimum embedding dimensions needed to for phase space reconstruction are calculated using the method of false nearest neighbors proposed by Cao [5]. The method relies on the fact that if m is the true embedding dimension of the phase space, then points near in m - dimensional space will be still be close in $m + 1$ dimensional space. Consider and time series $z(t) = z_1, z_2, \dots, z_N$ of N samples. Then delayed vectors, with embedding dimension m and time delay τ are given by

$$Z_i = (z_i, z_{i+\tau}, \dots, z_{i+(m-1)\tau}) \quad \forall i = 1, 2, \dots, (N - m)\tau \quad (8)$$

For i^{th} vector Z_i distance measure $a(i, m)$ is defined as

$$a(i, m) = \frac{\|Z_i(m+1) - Z_{n(i,m)}(m+1)\|}{\|Z_i(m) - Z_{n(i,m)}(m)\|} \quad (9)$$

$$i = 1, 2, \dots, N - m\tau$$

where $\|\cdot\|$ represents vector norm operation; $Z_i(m+1)$ and $Z_i(m)$ are i^{th} delay vectors in m and $m + 1$ dimensional space. $Z_n(m)$ is the nearest neighbor of $Z_i(m)$ in the m

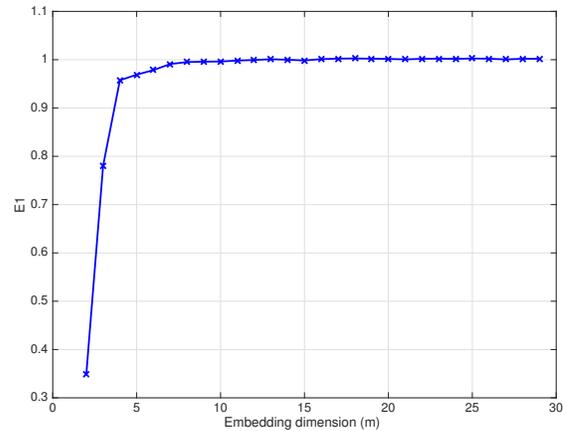


Fig. 3. Automatic determination of embedding dimension (m) using false near neighbour methods

dimensional space, where $n(i, m)$ is integer in set $(1 \geq n \geq N - m\tau)$.

The value of $a(i, m)$ is calculated for all the delay vectors and its mean $E(m)$ is given by

$$E(m) = \frac{1}{N - m\tau} \sum_{i=1}^{N-m\tau} a(i, m) \quad (10)$$

The variation in E for the increase in embedding dimension to $m + 1$ is given by the quantity $E1$ defined as

$$E1(m) = \frac{E(m+1)}{E(m)} \quad (11)$$

The quantity $E1(m)$ in (11) approaches unity and stops changing for some $m > m_0$ when the near neighbors in m dimensions stay near in $m + 1$ dimension as well. If this is the case then the minimum embedding dimension is given by $m_0 + 1$.

In order to illustrate the procedure described above for determination, the $E1(m)$ quantity for one of the data sets is plotted in Figure 3. It can be seen that $E1$ approaches unity and stops changing for $m > 8$, thus giving $m = 9$ as the minimum embedding dimension for phase space reconstruction.

V. PROPOSED METHOD

The method proposed looks for the causal relationship and direction of information flow between two time series $x(k)$ and $y(k)$ each of length N . Both time series are recorded from a dynamical system and share a common manifold \mathbf{M} in phase space. The procedure involves cross mapping /estimation of one time series from the other thereby giving the direction of maximum influence or information flow. The main steps involved in the proposed method are as follows:

- 1) Mean center and normalize both time series to unit standard deviation.

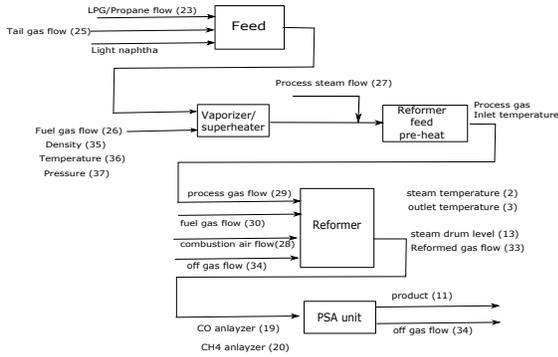


Fig. 4. Process schematic for industrial case study [20]

- 2) Determine the minimum embedding dimension for both time series using method given in section IV. The two time series can have different embedding dimension.²
- 3) Compute $\rho_{\hat{x}|M_y}$ and $\rho_{\hat{y}|M_x}$ using CCM method described in section II.
- 4) Test the NULL hypothesis using significance test given in section III.
- 5) Repeat steps 1-4 for increasing length of time series starting from N_{min} to N_{max} . Variables passing the convergence test, given in Section III-A, are considered for analysis.
- 6) In case the the NULL hypothesis is rejected and there is significant difference in $\rho_{N_{Max}}$ and $\rho_{N_{Min}}$ then $\rho_{\hat{x}|M_y} > \rho_{\hat{y}|M_x}$ gives $x \rightarrow y$ otherwise $y \rightarrow x$
- 7) Repeat steps 1-6 for all pairs of variables affected by the plant-wide disturbance
- 8) Variables influencing all the affected variables is designated as the source of disturbance.
- 9) Alternatively, an information flow graphs can be constructed with variables as nodes and arrows directing the flow of information as determined in step 6. This is the same concept as is employed in [8].

VI. INDUSTRIAL CASE STUDY

The proposed CCM based method is applied to industrial from South East Asian refinery. The same data set has been used previously for the detection of plant-wide oscillations([21],[19]) and root cause analysis ([20], [?], and [23]). The data consists of 37 different tags each having $N = 512$ samples at sampling rate of 1min. It has been reported that the plant suffers from a plant wide oscillation, due to valve nonlinearity, with oscillation period of 16.7 min^{-1} . Nine tags are identified to be affected by this oscillation ([21]) (shown in Figure 5). The nonlinearity signatures are found to be present in 4 tags, namely 11, 13, 33 and 34 [20]. Tag 34 is found to be the most nonlinear and is designated as the source of this plant wide oscillation. The non-linearity indices calculated by Thornhill [20] are also shown in the Figure 5.

²But for the sake of simplicity in this work same embedding dimensions are used for both time series

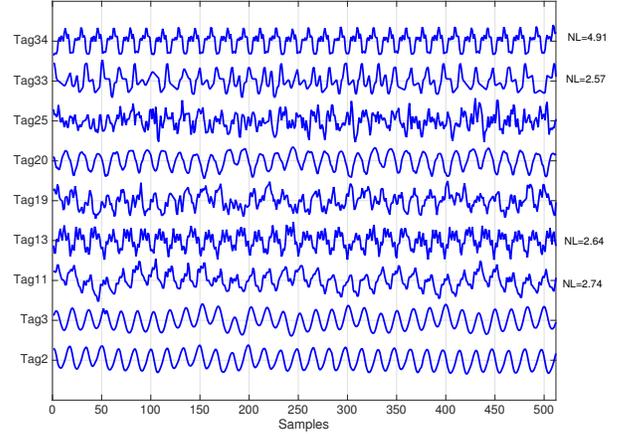


Fig. 5. Time trends and nonlinearity indices for SEA refinery data (Case study)

The same data set is studied using the proposed CCM method to analyzed how well it can identify the origin of the plant wide disturbance. Two scenarios are considered here a) In first scenario the known reason for disturbance is taken into account and tags showing non-linearity only are analyzed with the proposed method to identify the origin of the nonlinearity. b) The second scenario assumes that the cause is not known in advance and all 9 tags affected by plant-wide oscillation are analyzed. This scenario will test the broader scope of the proposed method.

A. Parameter Settings

The following parameters settings are used to analyze the industrial data.

- The analysis is started with sample size of $N_{min} = 100$ samples that is increased in steps of 100. The maximum number of samples used are $N_{max} = 500$
- Minimum embedding dimensions are computed using the method given in section IV and it is found that $m = 20$ fits well for all the time series. One may choose different m for different tags, but same m is used for all the loops.

B. Analysis with Known Root Cause

As discussed the plant wide oscillations in the industrial data under study are due to valve non-linearity. Only four tags are identified as the ones with signatures of nonlinearity. These four tags 11, 13, 33 and 34 are analyzed using the proposed CCM based method and results are summarized in the following.

It can be seen from the correlation plots for the CCM analysis in Figure 6 that Tag 34 has greater influence on all the other loops. The same effect is shown in the form of information flow graph in Figure 7. The information flow graph shows only those connections that fulfill the significance criteria given in section III and show $\Delta\rho > 0.02$. Thus the proposed method has been successful in designating the origin of plant wide disturbance.

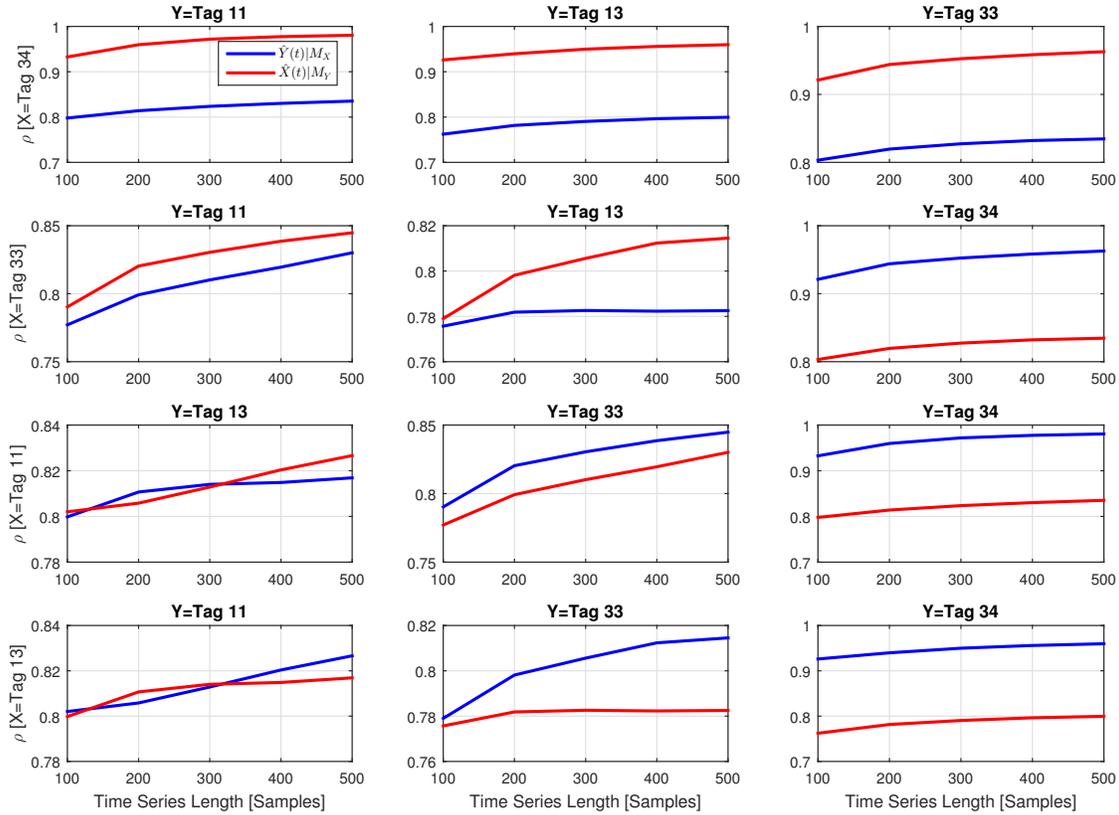


Fig. 6. CCM results Industrial case study (with known root cause)

C. Analysis with Unknown Root Cause

In this scenario it is assumed that the root cause is not known and all the loops affected by plant-wide oscillation are analyzed using proposed method. The results are summarized in Table I. Here again Tag 34 is found to be influencing all the other variables and thus designated as the source of plant-wide disturbance. It is interesting to note that the Tag25 has been influenced by tags other than Tag34. It might be so because the Tag 25 is recycled waste gas from another unit and the disturbance propagated through that unit [20].

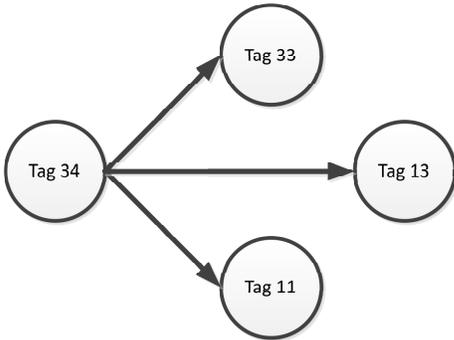


Fig. 7. CCM results (Information flow graph): industrial case study (known root cause)

TABLE I

CCM RESULTS FOR ALL LOOPS SUFFERING FROM PLANT-WIDE OSCILLATION (INDUSTRIAL CASE STUDY)

Cause	Influenced Loops
Tag 34	All
Tag 33	-
Tag 25	-
Tag 20	3
Tag 19	-
Tag 13	-
Tag 11	25
Tag 3	25
Tag 2	25

VII. CONCLUSIONS

In this paper convergent cross mapping based approach for isolating the source of plant wide disturbance is presented. The method is simple and effective and requires tuning of embedding dimension only. The proposed method is made reliable by appending it with automatic determination of embedding dimension. Moreover, the application of the proposed approach to an industrial case study gives promising results in identifying the origin of disturbance.

VIII. ACKNOWLEDGMENTS

The financial support from Siemens AS, Trondheim is gratefully acknowledged. The authors would like to thank Prof A.K. Tangirala for providing the industrial data.

REFERENCES

- [1] J. Arnhold, P. Grassberger, K. Lehnertz, and C. E. Elger, "A robust method for detecting interdependencies: application to intracranially recorded eeg," *Physica D: Nonlinear Phenomena*, vol. 134, no. 4, pp. 419–430, 1999.
- [2] M. Bauer, J. W. Cox, M. H. Caveness, J. J. Downs, and N. F. Thornhill, "Finding the direction of disturbance propagation in a chemical process using transfer entropy," *IEEE transactions on control systems technology*, vol. 15, no. 1, pp. 12–21, 2007.
- [3] —, "Nearest neighbors methods for root cause analysis of plantwide disturbances," *Industrial & Engineering Chemistry Research*, vol. 46, no. 18, pp. 5977–5984, 2007.
- [4] S. L. Bressler and A. K. Seth, "Wiener–granger causality: a well established methodology," *Neuroimage*, vol. 58, no. 2, pp. 323–329, 2011.
- [5] L. Cao, "Practical method for determining the minimum embedding dimension of a scalar time series," *Physica D: Nonlinear Phenomena*, vol. 110, no. 1, pp. 43–50, 1997.
- [6] D. Chicharro and R. G. Andrzejak, "Reliable detection of directional couplings using rank statistics," *Physical Review E*, vol. 80, no. 2, p. 026217, 2009.
- [7] A. A. S. Choudhury, S. L. Shah, and N. F. Thornhill, *Diagnosis of process nonlinearities and valve stiction: data driven approaches*. Springer Science & Business Media, 2008.
- [8] P. Duan, T. Chen, S. L. Shah, and F. Yang, "Methods for root cause diagnosis of plant-wide oscillations," *AIChE Journal*, vol. 60, no. 6, pp. 2019–2034, 2014.
- [9] R. A. Fisher, "Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population," *Biometrika*, vol. 10, no. 4, pp. 507–521, 1915.
- [10] C. W. Granger, "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica: Journal of the Econometric Society*, pp. 424–438, 1969.
- [11] R. Huffaker, "Phase space reconstruction from time series data: where history meets theory," *Proceedings in Food System Dynamics*, pp. 1–9, 2010.
- [12] H. Kantz and T. Schreiber, *Nonlinear time series analysis*. Cambridge university press, 2004, vol. 7.
- [13] M. B. Kennel, R. Brown, and H. D. Abarbanel, "Determining embedding dimension for phase-space reconstruction using a geometrical construction," *Physical review A*, vol. 45, no. 6, p. 3403, 1992.
- [14] E. Pereda, R. Rial, A. Gamundi, and J. Gonzalez, "Assessment of changing interdependencies between human electroencephalograms using nonlinear methods," *Physica D: Nonlinear Phenomena*, vol. 148, no. 1, pp. 147–158, 2001.
- [15] C. D. F. A. S. SAMPLE, "correlation coefficients covering the cases (i) the frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population, biometrika, vol. 10, pp. 507v521, 1915. here the method of defining the sample by the coordinates of," 1921.
- [16] T. Sauer, J. A. Yorke, and M. Casdagli, "Embedology," *Journal of statistical Physics*, vol. 65, no. 3, pp. 579–616, 1991.
- [17] G. Sugihara, R. May, H. Ye, C.-h. Hsieh, E. Deyle, M. Fogarty, and S. Munch, "Detecting causality in complex ecosystems," *science*, vol. 338, no. 6106, pp. 496–500, 2012.
- [18] F. Takens, "Detecting strange attractors in turbulence," in *Dynamical systems and turbulence, Warwick 1980*. Springer, 1981, pp. 366–381.
- [19] A. Tangirala, S. Shah, and N. Thornhill, "PSCMAP: A new tool for plant-wide oscillation detection," *Journal of Process Control*, vol. 15, no. 8, pp. 931–941, 2005.
- [20] N. F. Thornhill, "Finding the source of nonlinearity in a process with plant-wide oscillation," *Control Systems Technology, IEEE Transactions on*, vol. 13, no. 3, pp. 434–443, 2005.
- [21] N. F. Thornhill, S. L. Shah, B. Huang, and A. Vishnubhotla, "Spectral principal component analysis of dynamic process data," *Control Engineering Practice*, vol. 10, no. 8, pp. 833–846, 2002.
- [22] T. Yuan and S. J. Qin, "Root cause diagnosis of plant-wide oscillations using granger causality," *Journal of Process Control*, vol. 24, no. 2, pp. 450–459, 2014.
- [23] X. Zang and J. Howell, "Isolating the root cause of propagated oscillations in process plants," *International Journal of Adaptive Control and Signal Processing*, vol. 19, no. 4, pp. 247–265, 2005.