Authos' final draft of paper published in Journal of Process Control, 18, 2008, pp 707-719

A Practical Method for Identifying the Propagation Path of Plant-Wide Disturbances

Margret Bauer+ and Nina F. Thornhill*

+ABB Corporate Research Centre, Wallstadter Str. 59, D-68526 Ladenburg, Germany

**Department of Chemical Engineering, Imperial College London, South Kensington Campus, London SW7 2AZ*

Abstract

Disturbances in a continuous process often travel along the product stream and affect the performance of the process. To isolate the root cause and find the path along which the disturbance propagates requires an understanding of the cause-and-effect relationships between multiple variables. One way of identifying these relationships is through the time delays between process variables. A practical and robust approach is proposed that uses the cross-correlation function to estimate the time delay between process measurements and derive a qualitative model of the propagation path in the form of a causal map. The approach was applied to an industrial case study of a process affected by a plant-wide disturbance and was able to decide between two alternative root cause explanations. It was also applied successfully to a process with a recycle. The advantages of the proposed method are its ease of implementation and automation as well as simple interpretation of the resulting causal map.

Key words: Causal map, cross correlation function, digraph, propagation path,

fault diagnosis, plant-wide disturbance, root cause, time delay.

1 Introduction

Persistent dynamic plant-wide disturbances are a common problem in continuous chemical processes. Examples include oscillations caused by sticking valves, disturbances caused by interacting controllers and hydrodynamic instabilities such as slugging flows. Because of the interlinking of process equipment the disturbance may propagate through the plant and affect a large number of process variables, evolving into a plant-wide problem. The widespread nature of the disturbance then makes it difficult to identify its origin. Standard performance measures, such as control loop performance indices [1], are not able to deal with plant-wide problems because they deal with individual loops one at a time and do not take account of propagation.

A number of data-driven methods have been developed for root cause analysis of persistent dynamic disturbances, as reviewed in [2]. The analysis is often grouped in two stages: disturbance detection and disturbance diagnosis [3]. Detection is addressed by upper or lower threshold violation, increase in variance or oscillation detection [4], and a key step in the detection of a plant-wide disturbance is the identification of clusters of measurements having similar dynamic behaviour. Diagnosis has two objectives, the identification and the isolation of the disturbance. Identification finds the nature and location of the root cause and has been previously investigated, for example by the nonlinearity index [5]. The techniques mentioned so far require only data from routine operation of the process to give an explanation about the root cause, however the disturbance propagation paths generally have had to be

determined manually by inspection of the process schematic.

There have been strong calls from industrial commentators [6,7] for model-free causal analysis. The method proposed in this article uses historical data to achieve model-free causal analysis by taking advantage of the time delays between measurements closest to the root cause and those further away. Some process monitoring applications such as Expertune's PlantTriage and Matrikon's ProcessDoctor already include estimates of time delays between pairs of variables by means of detection of a peak in the cross-correlation function (CCF). The contribution of this article is that it constructs a robust and verified causal graph from time delay estimates. The verification is achieved (i) by statistical hypothesis testing to ensure that the detected peak in the CCF is significant and (ii) by consistency checking of the entries in the time delay matrix. The causal map is then generated from the verified time delay matrix. The paper gives procedures for these tasks that are generally applicable and illustrates them with industrial examples.

The paper is structured as follows. First, the method of time delay detection for isolation analysis is proposed using the maximum magnitude of the cross-correlation and a directionality index. In Section 3.3 threshold values for statistical significance for these two indices are derived. A causal graph for the propagation path is derived from the detected time delays in Section 4 which also presents a method for consistency checking of the time delay matrix. The approach is then applied to industrial continuous processes that are affected by a plant-wide disturbance to demonstrate its functionality.

2 Background and context

2.1 Causal analysis

Holland [8] and Pearl [9,10] introduced the area known as *learning causality from data* and described how causal graphs can be estimated from data. One approach is to hypothesize possible causal structures among the variables and reject those which are strongly incompatible with the observed correlations. The outcome is a graph which identifies the cause and effect variables. Establishing cause-and-effect (as opposed to correlation) in experimental data is difficult. However, measurements from industrial operations are usually in the form of time series and causal direction can be inferred much more easily when information about time is available. A simple pair-wise hypothesis that any pair of measured variables could have a causal relationship yields a cause-and-effect matrix, and with some assumptions about the structure of the system, the matrix leads to a causal graph. One way to generate the pair-wise hypotheses uses estimates from the data of conditional probabilities (a Bayesian network). Transfer entropy [11] adds information about time into a Bayesian network because it tests hypotheses concerning the joint and conditional probabilities of past and current values in a time series. Cause-and-effect analysis based on transfer entropy has been shown to work well in industrial case studies [12], but it does not give an explicit estimate of the time delay.

In this article, time delay estimates using cross-correlation are used to create the pair-wise hypotheses. The outcome of each test includes (i) a direction (whether X precedes Y or Y precedes X), (ii) an estimate of the delay, and (iii) the strength of the delayed correlation. Benefits of the cross correlation method

compared to transfer entropy include the faster and simpler computations and familiarity with the concepts of time delay and correlation among practicing engineers. Transfer entropy is, however, a more general method because it does not require the correlations between measurements to be linear.

2.2 Time delay estimation using cross-correlation

The cross-correlation function [13] describes the statistical properties of bivariate stationary processes by quantifying their similarity over time. It is a function of lag, κ. The cross correlation function (CCF) is estimated from two time series x and y which are captured at discrete times x_i and y_i for $i = 1 \ldots N$ where N is the number of samples.

$$\phi_{xy}[\kappa] = \frac{1}{N-\kappa} \sum_{i=1}^{N-\kappa} \hat{x}_i \hat{y}_{i+\kappa} \tag{1}$$

where $\hat{x}$ and $\hat{y}$ are derived from x and y by mean centering and scaling to unit standard deviation. Estimates of time delay and direction are derived from the value and sign of the κ at which ϕ_{xy} has its maximum magnitude, provided the magnitude is statistically significant.

Cross correlation has been widely used for time delay estimates. It was used to calculate time delays in radar signals reflected from Venus in a test of the theory of General Relativity [14]. Other reports include radar, sonar and acoustic applications [15], blood flow velocity monitoring and other ultrasound applications [16,17], and analysis of volcanic tremors [18]. Tabaru et. al. [19] presented a process application using cross correlation of wavelet coefficients in the time-frequency domain to identify the root cause of a plant-wide disturbance.

A persistent disturbance arising at one location in a process spreads through the plant along the product flow or through the effects of controllers. The origin, also referred to as root cause, may be detected by identifying time delays between the time-series of the measurements at different points in the process. If a direction can be determined, then a time delayed correlation is taken as evidence of causality because the data are derived from a physical process where time delays are known to be due to physical causation. As discussed below, time delays in some measurements depend on throughput and the estimated time delays will be different at different operating points. The causal hypothesis is, however, robust to changes in operating point because it depends only on the directions of the time delays and not on their absolute values.

In continuous processes the product, a liquid or gas, travels through the equipment with a flow rate which depends on the throughput of the plant. A feature in an upstream measurement, such as one of the peaks shown in Figure 7 in Section 5, will result in a similar peak in a downstream measurement after a time delay Δt. Physical properties of process streams such as temperature or composition travel at the flow rate of the stream and therefore can result in significant time delay. In a pipe of cross sectional area A and length l this is given by the solution of $\int_0^{\Delta t} F(t)\mathrm{d}t$ for Δt, where $F(t)$ is the volumetric flow rate. The same principles apply in process equipment such as distillation columns; there is a delay because of the time the product takes to travel from one measuring point to the next. Variations in pressure, by contrast, travel with the speed of sound which depends on the medium and not on the flow

rate. Also, propagation of variations in flow of an incompressible fluid in a pipe is effectively instantaneous.

Although it would seem from the above discussion that estimated time delay as a hypothesis of causality is best suited to temperature and composition measurements, causality can also be inferred from the cross correlation function even if the time delay is negligible compared to the data acquisition sampling interval, which is the situation typical of pressure and liquid pipeline flow measurements. The reason for this is that process dynamics also contribute to a shift in peak of the cross-correlation between the input and output. This was shown in [19] which examined the effect on the time delay estimate of linear dynamics of processes such as first or second order lags showing that the peak of the cross-correlation function has an additional delay θ where θ is the magnitude of the slope of the phase graph of the transfer function at the dominant frequency of the disturbance.

3 Mathematical formulation

3.1 Time delay estimation

The proposed method estimates the time delay between two process measurements by searching for the maximum (highest peak) and minimum (lowest valley) of the cross correlation function. If the CCF $\phi_{xy}[\kappa]$ is at its maximum for a certain time index $\kappa = \kappa^{max}$ then the second sequence shifted by κ^{max} is similar to the first sequence. In some cases, the correlation will be negative and the absolute value of the lowest valley is larger than that of the highest peak. Then the two sequences are most similar when the second one is inverted

and shifted by κ^{min}, where κ^{min} is the time index of the lowest valley. Thus, both the minimum and maximum of the CCF are recorded for the algorithm:

$$\begin{aligned} \phi^{max} &= \max_{\kappa}\{\phi_{xy}[\kappa]\} \\ \phi^{min} &= \min_{\kappa}\{\phi_{xy}[\kappa]\} \end{aligned} \tag{2}$$

where ϕ^{max} is positive and ϕ^{min} is negative. The corresponding time delays κ^{max} and κ^{min} at which the CCF has its maximum and minimum are also recorded. The choice between κ^{min} and κ^{max} as the detected time delay between x and y depends on whether ϕ^{max} or ϕ^{min} has a larger absolute value. The detected time delay λ is therefore

$$\lambda = \begin{cases} \kappa^{max}, & \phi^{max} + \phi^{min} \geq 0 \\ \kappa^{min}, & \phi^{max} + \phi^{min} < 0 \end{cases} \tag{3}$$

The time delay correlation between x_i and $y_{i+\lambda}$ has to exceed a certain threshold of statistical significance. The maximum time delayed correlation between the two time series is

$$\rho = \max\{\phi^{max}, |\phi^{min}|\}. \tag{4}$$

The CCF of any arbitrary finite length time sequences x and y inevitably has a minimum value and a maximum value even if there is no time delay present. It is therefore necessary to establish that the observed value of ρ has significance and is not merely a statistical fluctuation. A significance test for ρ is outlined below.

Given that even arbitrary sequences have both ϕ^{max} and ϕ^{min}, a further aspect has to be considered in order to confirm directionality by checking that one has a significantly larger magnitude than the other. If the magnitudes are similar, the result is ambiguous and no decision can be made. A *directionality index* is therefore introduced which measures the difference between minimum and maximum:

$$\psi = 2\frac{|\phi^{max} + \phi^{min}|}{\phi^{max} + |\phi^{min}|} \tag{5}$$

If the directionality index ψ is small then no decision can be made because ϕ^{max} and ϕ^{min} are similar in magnitude. A significance test for the directionality index ψ is presented in section 3.3.2. If ψ passes the test then a decision can be made about the direction of the time delay.

3.2 Dealing with oscillatory time series

The cross correlation function of a time delayed harmonic oscillation (sine or cosine) is periodic with the same frequency as the oscillation. Therefore there are ambiguities because of phase wrapping. For instance, if the period of oscillation is T_p it is not possible to distinguish between time series y lagging x by $0.75T_p$ or y leading x by $0.25T_p$, or indeed y leading by $nT_p + 0.25T_p$ where n is an integer. No estimate of time delay should be offered in this case because of the ambiguities. For harmonic time series, ϕ^{max} and ϕ^{min} of the CCF are of equal magnitude. Therefore one reason for failure of the directionality test is that an oscillation is present. Directionality can only be inferred in a time delayed oscillation if some additional dynamic features are present in the time series. In practice, sufficient additional dynamic features

often are present. Time delays can be detected if the oscillation occasionally has slightly shorter or slightly longer cycles. Even strongly cyclic data from chemical processes show variations in the oscillation period because the underlying physical mechanisms driving the oscillation have some randomness. For instance, in a limit cycle caused by stiction in a valve there is generally some variability in the force at which the valve stem starts to slip. A situation in which such additional features might *not* be present is when a process is being driven by rotating machinery with frequency control such that all cycles are the same length, for example where the cyclic behaviour is caused by the rotations of the drum in a dryer.

3.3 Significance testing and threshold settings

A contribution of this paper is the selection of the threshold for tests of the statistical significance of both maximum correlation ρ and directionality index ψ. The thresholds for both indices depend on the number of samples N since they are estimated from these sample values.

3.3.1 Maximum correlation ρ

The maximum correlation ρ is the correlation coefficient of first variable x and second variable y shifted by λ samples. The approach is to determine the distribution of ρ under the null hypothesis of no causal relationship between x and y i.e. the hypothesis is that they are unrelated random sequences. The reason why the statistic ρ has a distribution is because ρ is different each time it is calculated from a realization of two random sequences of length N. While the sampling statistics, mean and standard deviation, of the correlation

coefficient itself at time delay λ are known [20], the question being addressed here is different because the statistic is the maximal correlation between x and y achievable across all time delays κ. Points to note are that ρ is always positive, that is, it has to be tested by a one-sided hypothesis test, and the mean value of its distribution is not zero.

As far as the authors can ascertain, the sampling statistics of ρ as a function of N are not known so the approach chosen here establishes the distribution by empirical estimation using pairs of uncorrelated random sequences. Then, when working with plant data, an estimated time delay is considered significant if the magnitude of its ρ is significantly larger than a typical value for a pair of uncorrelated series. Figure 1 shows the estimated probability density function (PDF) constructed from 20,000 pairs of random sequences of sample length $N = 200$ with both normal and uniform distribution for the random values in the sequences. The results for both distributions are almost identical. The mean of ρ is larger than zero, as expected, approximately $\mu_\rho = 0.2$, and the shape of the PDF is asymmetrical.

If two time series do have a time delayed correlation, they are not random and their index ρ is unlikely to have come from the distribution shown in Figure 1. It is a one-sided test, so that the threshold value results in rejection of the null hypothesis and acceptance that the detected time delay is significant, if

$$\rho \geq \rho_{th} = \mu_\rho + 3\sigma_\rho \tag{6}$$

The test reflects the commonly used 3σ threshold. Referring to Figure 1, correlation between two time series would be inferred if the magnitude of ρ is larger than $\mu_\rho + 3\sigma_\rho = 0.2 + 3 \cdot 0.04 = 0.32$ for the case that $N = 200$.

The estimated PDF of ρ and therefore also mean μ_ρ and standard deviation σ_ρ are functions of the sample length N. To provide an analytical expression for ρ_{th}, the mean and standard deviation as functions of N were estimated empirically. Figure 2 shows empirical results for μ_ρ and σ_ρ when the number of samples N is varied from 0 to 1000 in intervals of 10 samples. For each N, 1000 pairs of random sequences were generated and their correlation indexes ρ computed. From those 1000 values, mean and standard deviation are estimated. Both μ_ρ and σ_ρ follow a dependence on N which can be described, exemplary for μ_ρ by the following function:

$$\mu_\rho(N) = \alpha_1 N^{-\alpha_2} \tag{7}$$

where α_1 and α_2 are two parameters to be defined using the experimental results from Figure 2. This curve fitting procedure is achieved by substituting the logarithms $r = \log \mu_\rho$ and $s = \log N$, $\beta_1 = \log \alpha_1$, $\beta_2 = \alpha_2$. The curve fitting problem is then the linear problem $r = \beta_1 + \beta_2 s$ from which the parameters β_1 and β_2 are estimated through sample points from the N versus μ_ρ results by linear regression.

The functions estimated from the curve fitting are:

$$\mu_\rho(N) = 1.85 N^{-0.41} \tag{8}$$

and

$$\sigma_\rho(N) = 0.79 N^{-0.53}. \tag{9}$$

Thus, the threshold for the correlation index ρ, above which a time delay is

detected, is a function of N as follows:

$$\rho_{th}(N) = 1.85N^{-0.41} + 2.37N^{-0.53}. \tag{10}$$

3.3.2 Directionality index ψ

Since the directionality index ψ is constructed from the CCF, it will also depend on the number of samples N. The same procedure as for the correlation index is now also applied to ψ. First, the probability density function for ψ is estimated from 1000 random sequence pairs of fixed length $N = 200$ (Figure 3). Since ψ is defined in Equation (5) as the absolute value of the difference between ϕ^{max} and ϕ^{min}, the result will always be positive and generally small. This is reflected in the PDF which is one sided with a peak close to zero. Thus, only the standard deviation σ_ψ is considered in the significance test. Figure 4 shows empirical results for σ_ψ for random sequences as the number of samples N is varied. The curve fitting method gave the following relationship:

$$\sigma_\psi(N) = 0.46N^{-0.16} \tag{11}$$

The significance threshold set for σ_ψ is based on a one-σ test. A directionality index is accepted as significant if:

$$\psi(N) \geq \sigma_\psi(N). \tag{12}$$

The reason for using a one-σ test rather than 3σ is pragmatic rather than based on theory. Experience with industrial data showed that a 3σ test for $\psi(N)$ gave false negative results and failed to detect some cases where the directionality was known and indeed where it could be seen by visual inspection of the time series.

3.3.3 Number of Samples N

A minimum number of samples has to be available to estimate the CCF. Box et al. [13] derive that a minimum of $N = 50$ pairs of samples of x and y are required to obtain a useful estimate of the cross correlation function. The reference shows only time delays below $\lambda < \frac{N}{4}$ can be reliably estimated and detected. The number of samples N in the data ensemble must therefore be at least four times the maximum expected time delay.

4 Propagation path models

The detection of the time delay facilitates the investigation of plant-wide disturbances which affect a number, say p, of process variables. The time delays between all $p(p-1)/2$ permutations of the p variables have to be considered. The size of the problem, which grows nearly quadratically with p, can be limited by considering only the time series from measurement points known to be involved with the plant-wide disturbance, for instance by use of one of the detection and clustering methods reviewed in [2]. To derive the path along which the disturbance travels, causal maps are introduced in the following section. The derived causal maps represent a qualitative model of the process for the case of a plant-wide disturbance. Two alternative topologies are discussed for a systematic derivation of the causal maps. Additional time delays which are not incorporated in the causal map are instead used for a consistency check which validates matching results of the CCF method.

4.1 Causality matrix

To understand the mechanism by which the disturbance proceeds from one part of the process to the next, the order of occurrence of the disturbance in the process variables has to be established. Causal maps have been previously employed for arguing cause and effect between process variables by Iri et al. [21] and more recently by Chiang and Braatz [22]. The propagation in these applications, however, was derived from expert knowledge rather than historical process data and the disturbance type had to be known in advance in order to analyze the propagation path.

The results of the CCF method are arranged in a causal matrix Λ showing the detected time delays from Equation (3) that have exceeded the thresholds from Equations 6 and 12. There is a maximum of $\frac{p(p-1)}{2}$ entries. The entry $\lambda_{1,2}$ in the second column of the first row, for example, is the estimated time delay of variable 2 relative to variable 1. The order of the process variables is arranged in such that the number of entries above the main diagonal is at its maximum. The effect of rearrangment is that whichever variable ends up in the first row is the candidate for the root cause because all other variables have time delays relative to that one. An algorithm for the rearrangement was presented in [12]. The causality matrix then has the following form:

$$\Lambda = \begin{bmatrix} - & \lambda_{1,2} & \dots & \lambda_{1,p} \\ & - & \ddots & \vdots \\ & & \ddots & \lambda_{p-1,p} \\ & & & - \end{bmatrix} \tag{13}$$

where $\lambda_{m,n}$ represents the time delay between variable x_m and x_n. This matrix representation for cause and effect relationships in the context of disturbance

isolation has been previously studied in [23]. The entries in the causality matrix are positive.

4.2 Consistency check

Once the causality matrix Λ is computed and the topology is selected, a consistency check can be carried out to verify and ascertain the results. The principle of the check is that the expected time delay $\lambda_{m,q}$ can be expressed as the sum of the already estimated time delays $\lambda_{m,n}$ and $\lambda_{n,q}$. When the measured time delays fulfil this criterion, they increase the confidence in the detected values. If the relationship is not fulfilled, then one or more of the time delays is incorrect. Uncertainty in the estimated time delay λ can be taken into account when conducting the consistency check by means of a user-defined accuracy parameter C:

$$\begin{aligned} \lambda_{m,q} + \lambda_{q,n} &\geq (1 - C)\lambda_{m,n} \\ \lambda_{m,q} + \lambda_{q,n} &\leq (1 + C)\lambda_{m,n} \end{aligned} \tag{14}$$

with accuracy $C \geq 0$ and $m < q < n$. If C is set to the lower boundary zero then the two time delays $\lambda_{m,q}$, $\lambda_{q,n}$ have to add up exactly to $\lambda_{m,n}$. A reasonable choice for C is to allow 20% deviation. For example, if $\lambda_{m,n}$ is five time samples then $\lambda_{m,q}$ and $\lambda_{q,n}$ must add up to a value between four and six to be accepted by the consistency check.

For $p = 3$ process variables there is one relationship to be checked for consistency. For larger values of p, more possibilities for consistency checks per

variable exist. For p variables, there are

$$N_\lambda = \frac{p(p-1)}{2} \tag{15}$$

time delays λ to consider. It can be shown through a summing procedure that the number of consistency checks to verify as described in Equation (14) is:

$$N_E = \frac{p(p-1)(p-2)}{6}. \tag{16}$$

Since there are three time delays per equation, the total number of possible consistency checks per detected time delay is

$$N_{CC} = \frac{3N_E}{N_\lambda} = p - 2. \tag{17}$$

If $p = 10$, as is the case in industrial case study I, there are $N_\lambda = 45$ time delays to detect while the number of equations to verify are $N_E = 120$ resulting in $N_{CC} = 8$ consistency checks per detected time delay.

The verification results from Equation (14) are stored in a matrix Λ^v with elements $\lambda^v_{m,n}$, where $\lambda^v_{m,n}$ is the number of verified consistency checks for time delay $\lambda_{m,n}$. A realization of the consistency checking procedure for all equations is achieved by the following implementation:

```
for i = 1:p-2
  for j = 1:p-2
    for k = j+i+1:p
      if (λ_{j,j+i} + λ_{j+i,k} ≤ (1 + C)λ_{j,k})
      and (λ_{j,j+i} + λ_{j+i,k} ≥ (1 − C)λ_{j,k})
        λ^v_{j,j+i} = λ^v_{j,j+i} + 1;
        λ^v_{j+i,k} = λ^v_{j+i,k} + 1;
        λ^v_{j,k} = λ^v_{j,k} + 1;
      end
    end
```

```
    end
  end
```

If all consistency checks are fulfilled, then Λ^v has the form

$$\Lambda^v = \begin{bmatrix} - & N_{CC} & \dots & N_{CC} \\ & - & \ddots & \vdots \\ & & \ddots & N_{CC} \\ & & & - \end{bmatrix} \tag{18}$$

and a high confidence can be assigned to all detected time delays. If, however, the $\lambda^v_{n,m} << N_{CC}$, then the detected time delay is omitted and replaced by an empty cell. Time delays for which the directionality or correlation index were found to be below the significance threshold or for which the estimated delay cannot be reconciled with other delays are represented by an empty entry in the causality matrix Λ.

4.3 Causal map topologies

4.3.1 Topologies I and II

A causal map is a graph with directed arcs between nodes where nodes represent process variables. To construct the causal map in a systematic way two generic topologies are considered as shown in Figure 5. In the left hand panel of Figure 5, the process variables are arranged in a row, in the following referred to as Topology I. This represents a chain of events which is expressed verbally as follows. The disturbance happened first in x_1 and then after time span $\lambda_{1,2}$ in x_2, after a further time span $\lambda_{2,3}$ the disturbance occurred in x_3 and so on. The nodes or process variables are arranged like a string of beads. The root cause is most likely close to the first variable in the string. An al-

ternative Topology II is shown in the right hand panel of Figure 5 where one process variable is clearly identified as the root cause which then affects all other variables. In other words, this means that the disturbances happens first in x_1 which then affects all other variables directly; after time delays $\lambda_{1,2}$ and $\lambda_{1,3}$ it can be observed in x_2 and x_3, and so on. The focus of Topology II is on the proximity of each variable to the variable closest to the root cause.

The choice between Topologies I and II depends on the results obtained from the time delay analysis and the consistency check. For Topology I all entries above main diagonal of Λ^v are considered while for Topology II the entries in first row of Λ^v are considered. To motivate the choice between Topology I and II, the number of entries above the main diagonal, that were not replaced by an empty entry, are compared to the number of entries in the first row. If the time delays in the first row increase the further they are away from the root cause, Topology I is chosen. On the other hand, if the time delays in the first row are in the same range or no increasing pattern can be detected, Topology II is chosen. These motivation points will be addressed later on in the industrial case studies.

4.3.2 Other topologies

While there is no ambiguity about the root cause variable in the causal matrix, it is noted that the time delay matrix would generally allow many other candidate causal maps besides Topologies I and II. The justification for only considering Topologies I and II in this work is that these represent two extremes that are very common in process plants. Topology I matches the case where an upstream disturbance propagates downstream, while Topology II

would be indicative of a site utility such as the steam system upsetting many units simultaneously. An intermediate case would arise, however, when a disturbed feed enters two separation columns or vessels in parallel.

A method for generation of a causal map for intermediate cases would start with a fully connected graph and prune redundant links from the graph. However, as explained above in the choice between Topologies I and II, there are also judgements to be made such as whether time delays are in the same range or not. The time delays are estimated quantities, so such questions do not necessarily have clear-cut answers. In the opinion of the authors, an attempt to generate a causal map of arbitrary topology from the time delay matrix alone is likely to generate more than one candidate graph, and the final selection will require external information. A manual method of resolution is to use the process flow diagram to inform the choice between candidates. In future, the increased availability of electronic representations of the process connectivity will open up possibilities for automated linkage of causality analysis with process connectivity analysis [2].

4.3.3 Processes with recycle

A time delay matrix with one or more positive entries remaining below the main diagonal, such as is shown below, indicates that the process has a recycle.

$$\Lambda = \begin{bmatrix} - & \lambda_{1,2} & & \\ & - & \ddots & \\ & & \ddots & \lambda_{p-1,p} \\ \lambda_{p,1} & & & - \end{bmatrix} \tag{19}$$

The causal map has an arc connecting a later measurement point to an earlier one. For example, assuming Topology I in the left hand panel of Figure 5 to give a good representation of the forward direction, the causal map derived from the above matrix would include an arc from x_p to x_1. The industrial case study in Section 5.2 shows an example. Discussion of why the matrix takes this form is deferred to Section 5.2.

The consistency checking procedure should only consider the λ entries above the main diagonal because the consistency checking expressions in Equation (14) do not apply when $m = n$, i.e. when the start and end point are the same.

5 Industrial case studies

Applications of time delay estimation, consistency checking and generation of the causal map are now presented. The first study gives a detailed worked example showing execution of the various steps of the method, while the second study demonstrates its effectiveness in detecting the presence of a recycle in a large plant.

5.1 Case study I

The industrial case study shown in this section is part of a larger plant at the Eastman Chemical Company. The process was selected because it was affected by an oscillatory disturbance of an unknown origin. The time series during the period of disturbance was available for all measurements along the process. In the following, the process and the disturbance are described and the alternative root causes are discussed prior to the analysis. The propagation

path identification method is applied to detect the time delays between all process variables and to derive the causal map.

5.1.1 Process and disturbance

Figure 6 shows the process schematic. A feed enters the top of the column and is separated into the desired product that exits the column at the bottom and a by-product that exits the column at the side draw not shown in Figure 6. A heating fluid is pumped through a piping system along the length of the column, without coming into contact with the product, and exits at the top. The heating fluid flow is controlled by the heating fluid temperature as the heating fluid is a shared facility with a varying temperature. The temperature in the column is controlled by a cascade loop for which the master controller (TC1) measures the temperature in the middle of the column and the slave controller (TC2) uses the temperature of the heating fluid to adjust the flow. The flow out of the column is the manipulated variable for the bottom tray level (LC1) and is adjusted through a pump. Temperatures are measured along the upper part of the column (TI1 to TI5), at the bottom tray (TI6) and further downstream from the column (TI7).

Figure 7 shows a close up of the measurement time series during the period of disturbance. The time series are mean centered and scaled to unit variance. All ten measurements show an oscillatory disturbance with the same oscillation period. The disturbance is less prominent in the controlled variables. The operators noticed the oscillation initially at level controller LC1 which showed a variation of 15% from the average. The disturbances in the temperature measurements were typically $\pm 0.5^0$C.

Two alternative root causes were discussed between operators and process engineers. The first explanation was that the level measurement was showing the strongest impact of the disturbance causing all other measurements to be upset, upstream and downstream in the process. A second explanation of the root cause was that the disturbance was caused further upstream, entered the process through the feed at the top of the distillation column and travelled through the unit. The propagation path is different for the two alternative explanations of the root cause. Thus, deriving the propagation path through time delay detection is expected to favor one of the two alternatives.

5.1.2 Time delay estimation

The detected time delay λ is derived from the CCF algorithm in Equation (13). The resulting causality matrix is computed as follows for all ten process variables.

$$\Lambda = \begin{array}{c} \begin{array}{cccccccccc} \text{TI1} & \text{TI2} & \text{TI3} & \text{TI4} & \text{TI5} & \text{TC1} & \text{TC2} & \text{TI6} & \text{LC1} & \text{TI7} \end{array} \\ \left[\begin{array}{cccccccccc} - & 2 & 8 & 14 & 25 & 50 & 163 & 39 & 58 & 64 \\ & - & 6 & 12 & 23 & 47 & 160 & 38 & 56 & 61 \\ & & - & 5 & 16 & 38 & 153 & 51 & 67 & 53 \\ & & & - & 11 & 34 & 266 & 48 & 42 & 49 \\ & & & & - & 22 & 246 & 37 & 30 & 37 \\ & & & & & - & 153 & 15 & 35 & 45 \\ & & & & & & - & 359 & 528 & 388 \\ & & & & & & & - & 19 & 28 \\ & & & & & & & & - & 8 \\ & & & & & & & & & - \end{array}\right] \end{array} \begin{array}{c} \\ \text{TI1} \\ \text{TI2} \\ \text{TI3} \\ \text{TI4} \\ \text{TI5} \\ \text{TC1} \\ \text{TC2} \\ \text{TI6} \\ \text{LC1} \\ \text{TI7} \end{array}$$

Sample results for TI4 and TI5 are as follows. The maximum of the CCF is at $\kappa^{max} = 11$. The amplitude of ϕ^{max} is almost twice the amplitude of ϕ^{min},

that is, $\psi = 0.54$ and the correlation index of TI4 and TI5 is $\rho = 0.61$. Both directionality and correlation index for the two time series are therefore above the thresholds derived from Equations 6 and 12 for $N = 1000$.

5.1.3 Consistency check

After establishing the time delays for all combinations of process variables the estimated time delays are verified by the consistency check described in Section 4.2. The accuracy in Equation 14 is set to $C = 20\%$. Since the number of variables is $p = 10$, the maximum number of consistency checks that can be fulfilled is $N_{CC} = 10-2 = 8$ (Equation 17). The number of fulfilled consistency checks for all combinations are:

$$\Lambda^v = \begin{array}{c} \begin{array}{cccccccccc} \text{TI1} & \text{TI2} & \text{TI3} & \text{TI4} & \text{TI5} & \text{TC1} & \text{TC2} & \text{TI6} & \text{LC1} & \text{TI7} \end{array} \\ \left[\begin{array}{cccccccccc} - & 8 & 6 & 6 & 6 & 4 & (2) & (3) & 5 & 6 \\ & - & 6 & 6 & 6 & 4 & (2) & (3) & 5 & 6 \\ & & - & 6 & 6 & 6 & (2) & (4) & 2 & 4 \\ & & & - & 8 & 5 & (1) & (3) & (4) & 5 \\ & & & & - & 5 & (1) & (3) & (4) & 5 \\ & & & & & - & (0) & 5 & (3) & 2 \\ & & & & & & - & (1) & (0) & (1) \\ & & & & & & & - & 5 & 5 \\ & & & & & & & & - & 6 \\ & & & & & & & & & - \end{array}\right] \end{array} \begin{array}{c} \\ \text{TI1} \\ \text{TI2} \\ \text{TI3} \\ \text{TI4} \\ \text{TI5} \\ \text{TC1} \\ \text{TC2} \\ \text{TI6} \\ \text{LC1} \\ \text{TI7} \end{array}$$

The matrix Λ^v can be interpreted as follows. For example, the entry in the second row and third column of Λ^v, representing the relationship between TI2 and TI3 $\lambda_{2,3} = 6$, is six, that means that six out of eight consistency checks are fulfilled. One consistency check is that $\lambda_{1,2} + \lambda_{2,3} \simeq \lambda_{1,3}$, that is, $2 + 6 = 8$ following from causality matrix Λ. A second out of the six consistency checks

is $\lambda_{2,3}+\lambda_{3,5}=6+16=22$ which is nearly equal to $\lambda_{2,5}=23$. The equality lies within the boundaries determined by $(1\pm C)\lambda_{2,5}$. The consistency numbers in brackets indicate that the time delay estimates did not pass one or both of the correlation and directionality tests in Equations 6 and 12 for $N=1000$. Now, all time delays in the causality matrix that either did not pass the thresholds or fulfilled fewer than 50% of the consistency checks are replaced by empty values. Furthermore, since no valid relationship between TC2 and any other variable could be detected, the row and column for TC2 are removed so that that the causality matrix Λ has the following form:

$$\Lambda = \begin{array}{c} \begin{array}{ccccccccc} \text{TI1} & \text{TI2} & \text{TI3} & \text{TI4} & \text{TI5} & \text{TC1} & \text{TI6} & \text{LC1} & \text{TI7} \end{array} \\ \left[\begin{array}{ccccccccc} - & \mathbf{2} & 8 & 14 & 25 & & & 58 & 64 \\ & - & \mathbf{6} & 12 & 23 & & & 56 & 61 \\ & & - & \mathbf{5} & 16 & 38 & & & \\ & & & - & \mathbf{11} & 34 & & & 49 \\ & & & & - & \mathbf{22} & & & 37 \\ & & & & & - & \mathbf{15} & & \\ & & & & & & - & \mathbf{19} & 28 \\ & & & & & & & - & \mathbf{8} \\ & & & & & & & & - \end{array}\right] \end{array} \begin{array}{c} \text{TI1} \\ \text{TI2} \\ \text{TI3} \\ \text{TI4} \\ \text{TI5} \\ \text{TC1} \\ \text{TI6} \\ \text{LC1} \\ \text{TI7} \end{array}$$

5.1.4 Causal map

The causal map can now be derived from the verified causality matrix above. The choice between Topology I and Topology II as described in Section 4.3 depends on the number of valid entries in the first row and the number of valid entries above the main diagonal. The causality matrix of the industrial case study has eight entries above the main diagonal and six in the first row. Also, the values in the first row are increasing. Thus, Topology I is chosen for

constructing the causal map. The estimated time delays that will be used for the construction are highlighted in bold font in the above causality matrix.

Figure 8 shows the causal map and the chain of events. The disturbance occurs first in temperature TI1 at the top of the column. It then travels through the column and affects the consecutive measurements, TI2 after two time samples, six samples later TI3, then after another five samples TI4 and so on. After a total of 2+6+5+11+22+15+19+8=88 time samples after the disturbance first occurred in TI1, it shows up in the last variable TI7. The cause of events and the propagation path of the disturbance could therefore be identified. Physical investigation verified the result and showed that the disturbance was caused in an upstream unit and entered the feed at the top of the column.

5.2 Case study II

The industrial case study in this section is a large plant with a recycle, provided courtesy of BP. An outline schematic is shown in Figure 9, while Figure 10 shows the time series of the measurements at the measurement points indicated in Figure 9. The sampling interval is one minute, and there is an oscillatory disturbance in most of the measurements with a period of about 56 minutes.

5.2.1 Causal map

Application of the methods gave the following causality matrix for the plant with recycle :

$$\Lambda = \begin{array}{c} \begin{array}{cccccc} \text{LC3} & \text{PC1} & \text{TC1} & \text{LC4} & \text{LC1} & \text{LC2} \end{array} \\ \left[\begin{array}{cccccc} - & \mathbf{2} & 3 & 4 & 11 & \\ & - & \mathbf{0} & 4 & 10 & \\ & & - & \mathbf{2} & 9 & \\ & & & - & \mathbf{6} & 12 \\ & & & & - & \mathbf{7} \\ \mathbf{7} & & & & & - \end{array}\right] \end{array} \begin{array}{c} \text{LC3} \\ \text{PC1} \\ \text{TC1} \\ \text{LC4} \\ \text{LC1} \\ \text{LC2} \end{array}$$

A feature of the causality matrix is the entry in the bottom left corner for $\lambda_{\text{LC2,LC3}}$. This entry stays there even after execution of the algorithm in [12] which moves as many positive time delays as possible above the main diagonal. A way to appreciate this point is to recognize that the same time delay information could alternatively be presented as a negative time delay $\lambda_{\text{LC3,LC2}} = -7$ in the top right corner. There is no rearrangement of rows and columns that could move the negative delay below the diagonal while retaining all the positive delays above the diagonal. No delays are reported for $\lambda_{\text{PC1,LC2}}$ and $\lambda_{\text{TC1,LC2}}$. The cross correlation values for these measurements were significant (Equation 6) but no direction could be established because the maximum peak and minimum valley in the cross correlation function were of similar magnitude (Equation 12).

As discussed earlier, consistency checking is limited to the entries above the main diagonal. Also, the accuracy parameter, C, for the consistency checks for the small time delays $\lambda_{\text{LC3,PC1}}$, $\lambda_{\text{LC3,TC1}}$ and $\lambda_{\text{LC3,TC1}}$ was increased to 50% rather than 20% in this analysis. The reason for this is that the data had a one minute sampling interval and the numerical rounding errors for small time delays are significant.

The causality matrix has five entries above the main diagonal and four in the first row and again Topology I is chosen for constructing the causal map. The estimated time delays used for the construction are highlighted in bold font in the matrix. The seven minute delay between LC2 and LC3 is included in the causal map in Figure 11 as an arc connecting these two measurement points in the reverse direction giving a cyclic causal map. The total time delay around the recycle is 24 minutes

5.2.2 Discussion of the recycle case study

Finding the root cause of a disturbace in a process with recycle is not always a well defined problem because the disturbance may not have a localized origin and instead be due to the dynamics of the recycle as mass and energy flow round in a coordinated way. As shown in [24], disturbances return to their point of origin and initiate a further disturbance and the end result is a disturbance that can be detected everywhere. In this study, the causality matrix shows the disturbance propagating all round the recycle. It also places LC3 in the first row because it is the measurement point with the largest number of detected time delays to other measurement points, which suggests the root cause is localized in or near the reflux drum.

The company reported that it was possible to settle the plant by making the level control in the separator reflux tank (LC3) less tight, thereby including buffering capacity into the recycle. This observation shows that both findings of the causality matrix offer significant insights into the nature of the problem. Firstly, the causality matrix shows the presence of the recycle and that there are coordinated disturbances at many measurement points in the recycle. Sec-

ondly, it is well known that buffering helps to counteract the lively dynamic responses of recycles, and in this case the causality matrix pointed to LC3 as the best place to increase the buffer capacity because it affected the largest number of other measurement points.

Further questions raised by the causal map are: (i) what is the mechanism by which the disturbance propagates upstream from the reactor (LC4) to the buffer tank (LC1), and (ii) how is it possible for a cross correlation analysis to discover a cyclic map?

The causal map suggests disturbances from the reactor propagate upstream to the buffer tank. The explanation is that the control valve for the reactor level control is in the inflow pipe between the buffer tank and reactor and the valve position is changed by the LC4 controller as necessary to maintain a constant level in the reactor. The mechanism of propagation from LC4 to LC1 is because changes in flow influence the buffer tank level. The analysis also shows the buffer tank level can influence other measurements points such as flash tank level (LC2). In fact, it is the unmeasured flow from the buffer tank into the reactor which affects the flow from the reactor into the flash tank. The flow rate is not measured, however, and the buffer tank level acts as a proxy for the unmeasured flow.

The causal map follows the direction of flow around the recycle, as would be expected, but there are two routes between two points in process with a recycle. Therefore some questions remain about why the CCF analysis highlighted one route and and not the other. For instance, why did the analysis leading to the causality matrix detect the delay of 10 minutes from PC1 to LC1 and the delay of 7 minutes from LC1 to LC2 rather than delays of 14 minutes between

LC1 and PC1 and of 17 minutes from LC2 to LC1? According to the process schematic, both results should be equally valid.

An inspection of the time series of the measurements gives an insight into this phenomenon. Whilst all the time trends show evidence of the 56 minute oscillation, they have other features as well. For instance, LC1 and LC4 have the large valley towards the right hand side of Figure 10 at around 2850 minutes. There is also evidence of this feature in PC1, although it is inverted. Another example is the time series of LC1, LC2 and TC1 which have high frequency noise superimposed on the oscillations. The cross correlation function uses such dynamic features to aid the detection of time delays and to resolve ambiguities that may arise in the detection of time delays in oscillating time trends. The CCF method tends to pick up shorter time delays that are less prone to interference from other disturbances.

6 Conclusions

In this paper, a practical approach of finding the root cause by retracing the propagation path of the disturbance has been described. The time delays that often arise between two consecutive measurements as the disturbance travels along the process flow were detected by using the cross-correlation function. In the multivariate case of a larger number of process variables, the time delays were arranged in a causality matrix. The matrix formed the basis for the construction of a qualitative model in form of a causal map. The approach was illustrated using industrial case studies, one with an oscillatory disturbance entering the plant from upstream and the other with a recycle. The proposed method could successfully retrace the propagation paths and identify the root

causes of the disturbances.

Acknowledgments

The authors would like to thank all members of the Advanced Control Technologies Group at Eastman Chemical Company for feedback and discussion. The principal author gratefully acknowledges the financial support of the UCL Graduate School, of the IEE (Hudswell Bequest Travelling Fellowship) and the internship programme at Eastman Chemical Company.

References

[1] S.J. Qin. Control performance monitoring - a review and assessment. *Computers & Chemical Engineering*, 23:173–186, 1998.

[2] N.F. Thornhill and A. Horch. Advances and new directions in plant-wide controller performance assessment. *Control Engineering Practice*, 15:1196–1206, 2007.

[3] R. Isermann and P. Ballé. Trends in the application of model-based fault detection and diagnosis of technical processes. *Control Engineering Practice*, 5:709–719, 1997.

[4] T. Hagglund. A control-loop performance monitor. *Control Engineering Practice*, 3:1543–1551, 1995.

[5] M.A.A.S. Choudhury, S.L. Shah, and N.F. Thornhill. Diagnosis of poor control loop performance using higher order statistics. *Automatica*, 40:1719–1728, 2003.

[6] M.A. Paulonis and J.W. Cox. A practical approach for large-scale controller performance assessment, diagnosis and improvement. *Journal of Process Control*, 13:155–168, 2003.

[7] L. Desborough and R. Miller. Increasing customer value of industrial control performance monitoring - Honeywell's experience. *AIChE Symposium Series*, 326(98):13–186, 2002.

[8] P.W. Holland. Statistics and causal inference. *Journal of the Americal Statistical Associations*, 81(396):945–960, 1986.

[9] J. Pearl. Causal diagrams for empirical reserach. *Biometrika*, 82(4):669–688, 1995.

[10] J. Pearl. *Causality*. Cambridge University Press, 2000.

[11] T. Schreiber. Measuring information transfer. *Physical Review Letters*, 85:461–464, 2000.

[12] 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*, 15:12–21, 2007.

[13] G.E.P. Box, G.M. Jenkins, and G.C. Reinsel. *Time Series Analysis: Forecasting and Control*. Prentice Hall, New Jersey, 1994.

[14] I.I. Shapiro, G.H. Pettengill, M.E. Ash, M.L. Stone, W.B. Smith, R.P. Ingalls, and R.A. Brockelman. Fourth test of General Relativity: preliminary results. *Physical Review Letters*, 22:1265–1269, 1968.

[15] C.H. Knapp and G.C. Carter. Generalized correlation method for estimating time-delay. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 24(4):320–327, 1976.

[16] X. Lai and H. Torp. Interpolation methods for time-delay estimation using cross-correlation method for blood velocity measurements. *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control*, 46(2):277–290, 1999.

[17] F. Viola and W.F. Walker. A comparison of the performance of time-delay estimators in medical ultrasound. *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control*, 50(4):392–401, 2003.

[18] G. Dubosclard, F. Donnadieu, P. Allard, R. Cordesses, C. Hervier, M. Coltelli, E. Privitera, and J. Kornprobst. Doppler radar sounding of volcanic eruption dynamics at Mount Etna. *Bulletin of Vulcanology*, 66(5):443–456, 2004.

[19] T. Tabaru, K. Nakano, S. Shin, and T. Matsuo. Plant data analysis by Gabor wavelet. In *Proceedings of SICE Annual Conference*, pages 950–955, Okayama, Japan, August 2005.

[20] W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery. *Numerical Recipes*. Cambridge University Press, Cambridge, UK, 2002.

[21] M.K. Iri, E. Aoki, E. O'Shima, and H. Matsuyama. An algorithm for diagnosis of system failures in the chemical process. *Computers & Chemical Engineering*, 3:489–493, 1979.

[22] L.H. Chiang and R.D. Braatz. Process monitoring using causal map and multivariate statistics: fault detection and identification. *Chemometrics and Intelligent Laboratory Systems*, 65:159–178, 2003.

[23] R. Sehgal, O.P. Gandhi, and S. Angra. Fault location of tribo-mechanical systems - a graph theory and matrix approach. *Reliability Engineering and System Safety*, 70:1–14, 2000.

[24] R.B. Gopaluni, H. Raghavan, R.S. Patwardhan, S.L. Shah, and G.A. Dumont. Identification of delay dominant recycle systems. *Journal of Process Control*, 16:903–912, 2006.

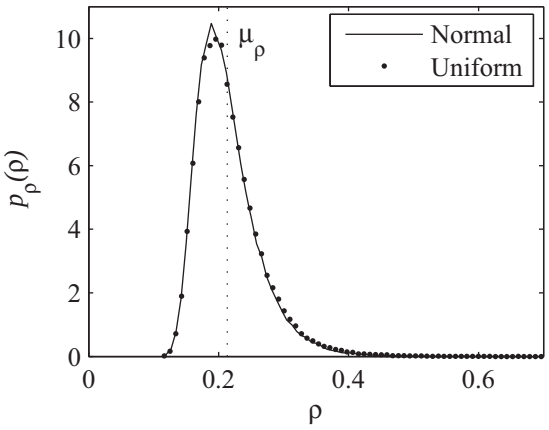

Fig. 1. Probability density function of correlation index ρ for two uncorrelated random sequences of length $N = 200$, estimated for both uniform and normal random distribution.

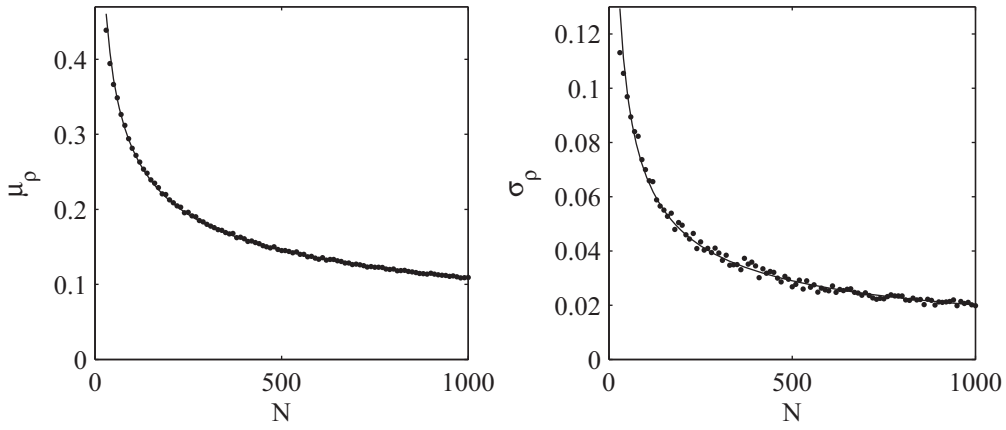

Fig. 2. Mean (left panel) and standard deviation (right panel) of correlation index as a function of sample length N. The solid line indicates the approximation function from Equations (8) and (9).

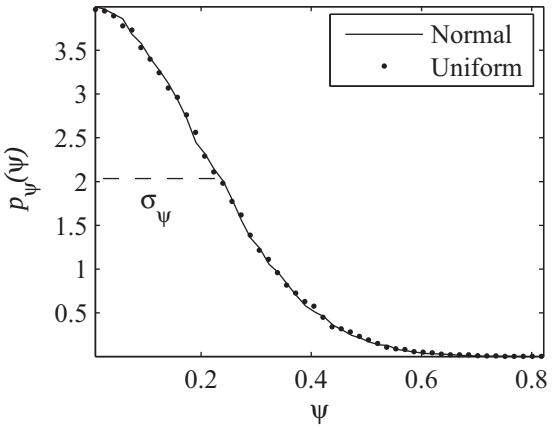

Fig. 3. Probability density function of directionality index ψ for two uncorrelated random sequences of length $N = 200$, estimated for both uniform and normal random distribution.

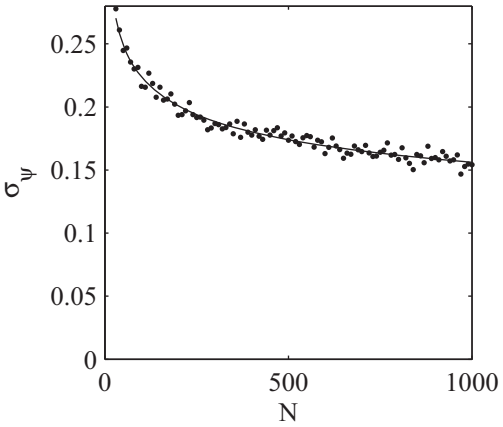

Fig. 4. Standard deviation of directionality index ψ as a function of sample length N. The solid line indicates the approximation function as given in (11) while the dots give the experimentally measured values.

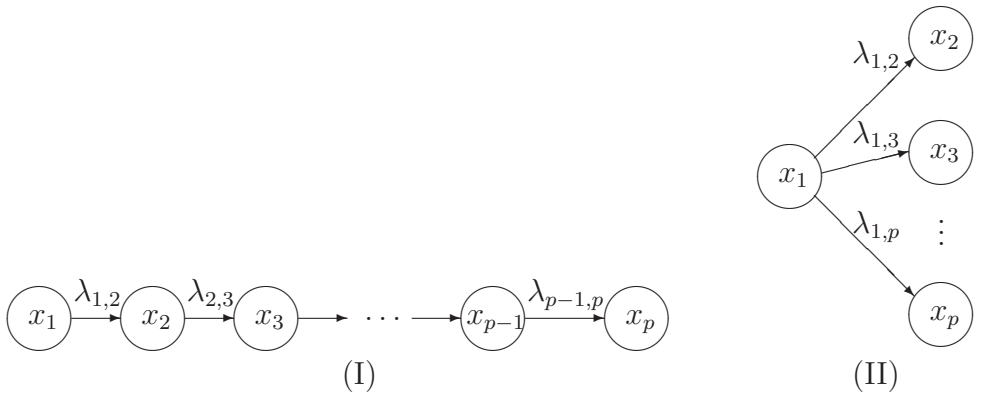

Fig. 5. Two alternative topologies for construction of a causal map of p process variables from $p-1$ detected time delays λ: (I) variables in series and (II) dependent on one root cause variable.

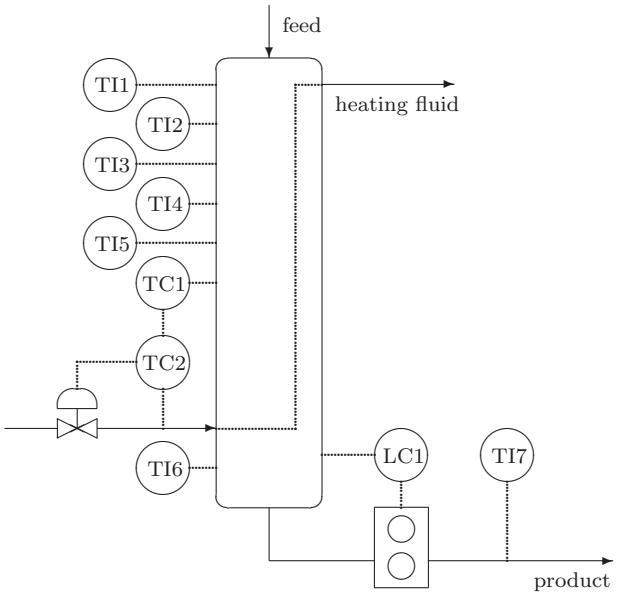

Fig. 6. Process schematic of the industrial case study I, process at Eastman Chemical Company.

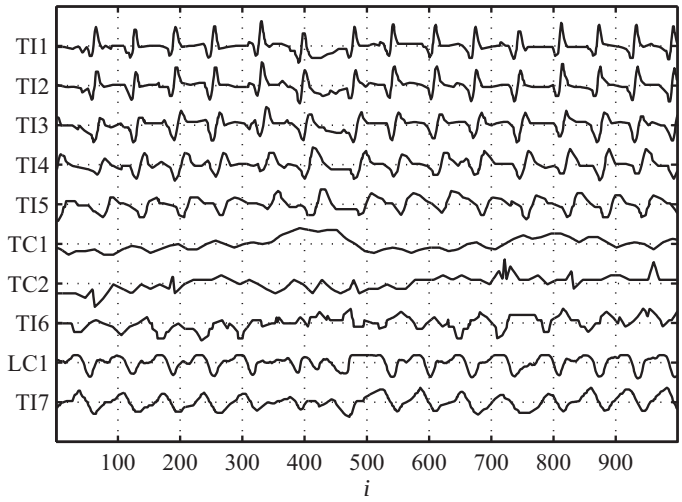

Fig. 7. Time series for process measurements of industrial case study I.

Fig. 8. Causal map of industrial case study I.

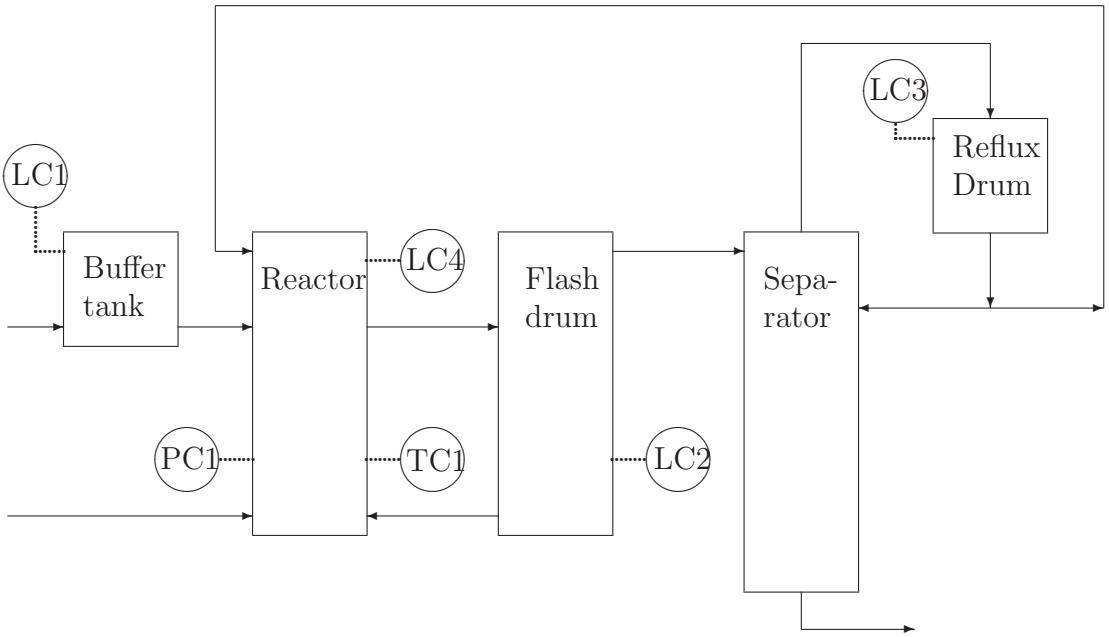

Fig. 9. Process schematic of industrial case study II.

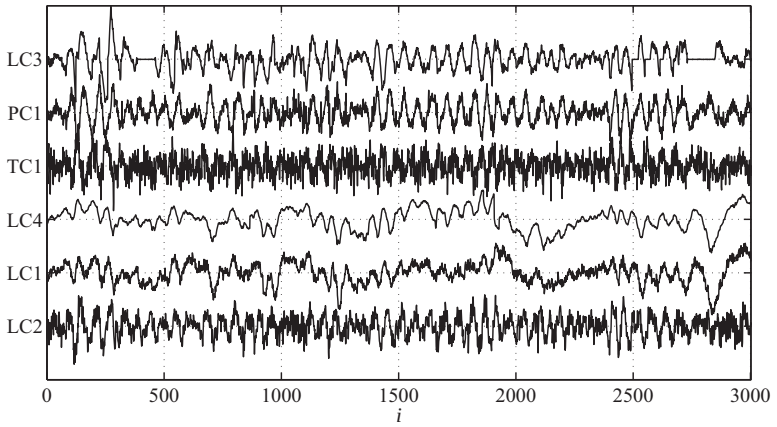

Fig. 10. Time series for process measurements of industrial case study II.

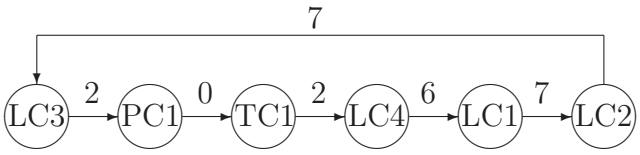

Fig. 11. Causal map of industrial case study II.

Figure Captions

Figure 1: Probability density function of correlation index ρ for two uncorrelated random sequences of length $N = 200$, estimated for both uniform and normal random distribution.

Figure 2: Mean (left panel) and standard deviation (right panel) of correlation index as a function of sample length N. The solid line indicates the approximation function from Equations (8) and (9).

Figure 3: Probability density function of directionality index ψ for two uncorrelated random sequences of length $N = 200$, estimated for both uniform and normal random distribution.

Figure 4: Standard deviation of directionality index ψ as a function of sample length N. The solid line indicates the approximation function as given in (11) while the dots give the experimentally measured values.

Figure 5: Two alternative topologies for construction of a causal map of p process variables from $p - 1$ detected time delays λ: (I) variables in series and (II) dependent on one root cause variable.

Figure 6: Process schematic of the industrial case study I, process at Eastman Chemical Company.

Figure 7: Time series for process measurements of industrial case study I.

Figure 8: Causal map of industrial case study I.

Figure 9: Process schematic of industrial case study II.

Figure 10: Time series for process measurements of industrial case study II.

Figure 11: Causal map of industrial case study II.