

Finding the Direction of Disturbance Propagation in a Chemical Process Using Transfer Entropy

Margret Bauer, John W. Cox, Michelle H. Caveness, James J. Downs, and Nina F. Thornhill, *Senior Member, IEEE*

Abstract—In continuous chemical processes, variations of process variables usually travel along propagation paths in the direction of the control path and process flow. This paper describes a data-driven method for identifying the direction of propagation of disturbances using historical process data. The novel concept is the application of transfer entropy, a method based on the conditional probability density functions that measures directionality of variation. It is sensitive to directionality even in the absence of an observable time delay. Its performance is studied in detail and default settings for the parameters in the algorithm are derived so that it can be applied in a large scale setting. Two industrial case studies demonstrate the method.

Index Terms—Causal map, control loop performance, digraph, fault diagnosis, kernel estimator, oscillation, plantwide disturbance, probability density function, process history, time series, transfer entropy.

I. INTRODUCTION

COMMERCIAL control loop condition monitoring systems are now being widely used in the process industries [1]–[4] and are reducing running costs by successfully highlighting control loops that need attention. However, it remains a challenge to determine the root cause of a disturbance that propagates widely through a unit or plant causing secondary upsets in other measurements and control loops. In such a case, many variables and control loops may be identified as performing poorly even though they are not the cause. Isolating and diagnosing the fault is, therefore, an important task. This paper describes a solution through the application of transfer entropy, a method due to Schreiber [5] that exploits conditional probabilities to determine cause and effect relationships in process data. The possibility of a transfer entropy test was introduced in an earlier conference paper [6] which showed that it was possible to determine causes and effects using the process measurements collected during routine operation of a plant. The key advance in this paper compared to [6], is the detailed description and optimization of the method to give default parameters to facilitate its application to a large-scale plant. A further contribution of

Manuscript received May 5, 2005. Manuscript received in final form June 30, 2006. Recommended by Associate Editor S. Palanki. The work of M. Bauer was supported in part by University College London Graduate School, by the IEE (Hudswell Bequest Travelling Fellowship), and by the internship programme at Eastman Chemical Company.

M. Bauer and N. F. Thornhill are with the Department of Electronic and Electrical Engineering, University College London, London WC1E 7JE, U.K. (e-mail: n.thornhill@ee.ucl.ac.uk).

J. W. Cox, M. H. Caveness, and J. J. Downs are with the Advanced Controls Technology Group, Eastman Chemical Company, Kingsport, TN 37662-5280 USA.

Digital Object Identifier 10.1109/TCST.2006.883234

this paper is to give fundamental insights into its operation and to extend it to additional case studies.

The reviews by Venkatasubramanian *et al.* [7]–[9] made the economic argument for the importance of fault detection and diagnosis in the chemical process industries. They presented a family tree of methods for analysis of process systems with an emphasis on fault detection and diagnosis. Categories highlighted were as follows:

- quantitative and qualitative process history-based methods;
- quantitative and qualitative model-based methods.

Transfer entropy is a quantitative process history-based method, and the outcome of the analysis is a qualitative process model in the form of a digraph showing the causal relationships between variables. The textbook by Chiang *et al.* [10] also presented broad coverage of knowledge-based, data-driven, and analytical techniques.

Propagation paths in case of normal operation and disturbances have been previously investigated with statistical methods based on probability density functions. Chiang and Braatz [11] used the Kullback–Leibler information distance to identify broken relationships when a fault was present. A requirement for the identification, however, was the existence of a causal map in the form of a digraph that had to be derived from expert knowledge of the process. Signed digraphs were extensively reviewed by Maurya *et al.* [12], [13], who discussed graph-based approaches for safety analysis and fault diagnosis of chemical process systems. They highlighted the difficulties of the accurate capture of the graph representation pointing out that it is a time consuming task and error-prone. By contrast, in the work presented here, the causal map is derived from the data.

This paper is organized as follows. In Section II, an overview of data-driven methods for directionality is given together with an outline of the concept of transfer entropy. The definition and implementation of entropy-based methods is given in Section III, and implementation issues are addressed in Section IV. Section V describes an automated method for ordering and pruning the causal relationships to generate a causal map in the form of a digraph. Two case studies are introduced in Section VI to show the use of transfer entropy for tracing fault propagation paths. Settings for the parameters used in the computation are also proposed and justified.

II. MEASURING DIRECTIONALITY AND CAUSALITY

A. Directionality and Dependency

In chemical processes, a persistent fault or disturbance often appears in many places in the plant. Plantwide oscillations can

originate from a variety of causes including instrumentation failures, tuning problems, and process inherent instabilities. Other types of disturbances may be transient events or nonperiodic. Examples are presented in the case studies of Section VI-B and VI-C. Typically a disturbance originating at one location propagates through the plant and causes secondary upsets in other locations and the task is to distinguish the root cause from the secondary upsets so that maintenance can be done. The signals change as the disturbance travels in the process because of time delays, attenuation, and the presence of noise and further disturbances. The task of a directionality measure is to give a representation of the direction in which the disturbance propagates and to find cause and effect relationships.

One way of determining the cause-and-effect relationship is to look for time delays, for instance by seeking the lag at which the cross-correlation function between the time trends of two measurements is maximized. However, time lags are not easy to determine in an automated and unambiguous way when the measurements are oscillatory or when they are noisy. Moreover, the time delay may be undetectable if it is smaller than the sampling interval of the data historian. The information theory measure of transfer entropy proposed by Schreiber [5], overcomes these difficulties by quantifying the amount of information transferred from one signal to the next and vice versa. By measuring a functional relationship that is often described as nonlinear coherence, transfer entropy identifies the driver and response variables of a system and the causality.

B. Information Theory Measures

Measures for quantifying dependency for bi- or multivariate time series include the correlation coefficient, cross-correlation, and mutual information [14]. Mutual information quantifies dependencies from the joint probability density function of two random variables. It measures the reduction of uncertainty of a random variable after the knowledge of a second variable as follows:

$$I(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(x, y) \log \frac{p(x, y)}{p(x)p(y)} dx dy. \quad (1)$$

These methods are symmetrical. They quantify the amount of dependency but cannot measure its directionality or causality.

The information theory measure of transfer entropy proposed in [5] takes the concept of mutual information a step further. It examines the transition probabilities $p(x_{i+h}|\mathbf{x}_i, \mathbf{y}_i)$ and $p(y_{i+h}|\mathbf{x}_i, \mathbf{y}_i)$, where $p(\cdot|\cdot)$ denotes the conditional relationship, that is, the probability that x_{i+h} has a certain value when past values \mathbf{x}_i and \mathbf{y}_i are known. Here, x_{i+h} is the value of x , h steps in the future from x_i , and h is referred to as the prediction horizon. Full details are given in Section III-B. Refinements have been introduced since the original work of Schreiber to optimize the detection of directionality for specific applications [15], [16]. In this paper, the original method is used and its parameters optimized for use with data from chemical processes.

The described directionality methods were developed theoretically for time trends caused by dynamics for which the intensity of coupling is known. Common systems are the Hénon map, the Lorenz attractor [17], and the Rössler system [18], for all of which the coupling strength can be adjusted with sets of

parameters. Real-life measurements include physiological data, such as heart and breath rate [5] or neuro-electrical signals [19], [20], financial data [15], [16], and outputs of an electronic network [21]. The nature of the data in these applications resembles the measurements captured during process disturbances in chemical processes because it is a superposition of deterministic behavior and stochastic noise, hence, the motivation of applying the methods to directionality analysis of fault propagation in chemical plants. A variety of related methods based on a principle called generalized synchronization have also been developed in the last decade [19], [20], [22]. The methods use an embedded matrix whose columns are time-shifted measurements. These also look promising for finding directionality in process measurements and an initial exploration of their use with process data is reported in [23].

III. METHODS

A. Transition Probability

To estimate the probability density functions (PDFs) and, thus, the transition probabilities, two assumptions must be made. First, enough relevant data points must be available to estimate the probability density function. Second, the time series must be stationary, that is, the expectation values of first orders, mean, and variance, must not be a function of time [24], [25].

Two discrete signals of length N with continuous amplitude axis are considered. That is, signals x and y are sampled at time instances i , $x_i \in [x_{\min}; x_{\max}]$, $y_i \in [y_{\min}; y_{\max}]$ with $i = 1, \dots, N$. The PDF of x and y is denoted by $p(x)$ and $p(y)$, respectively, while the joint PDF of x and y is $p(x, y)$. An important property of two stochastic signals which will be required later for the definition of transition probabilities is independence. Two variables are independent if the joint PDF is the product of the PDFs of the variables, such that

$$p(x, y) = p(x)p(y). \quad (2)$$

Joint PDFs for two stationary variables sequential in time are denoted by $p(x_{i+h}, x_i)$ with the same PDF for x_i, x_{i+h} because of stationarity, that is, $p(x_i) = p(x_{i+h})$, where h is the prediction horizon of x_i . The generalization of this joint PDF is the joint PDF for $k + l$ variables giving $p(\mathbf{x}_i, \mathbf{y}_i)$, where $\mathbf{x}_i = [x_i, x_{i-\tau}, \dots, x_{i-(k-1)\tau}]$ and $\mathbf{y}_i = [y_i, y_{i-\tau}, \dots, y_{i-(l-1)\tau}]$ are embedded vectors. The parameters k and l are referred to as the embedding dimension of x and y , respectively. A time interval τ allows the scaling in time of the embedded vector. The joint PDF $p(\mathbf{x}_i, \mathbf{y}_i)$, thus, captures time dynamics as well as interactions between x and y . The transition probability is defined after the Bayesian principle through joint PDFs

$$p(x_{i+h}|\mathbf{x}_i, \mathbf{y}_i) = \frac{p(x_{i+h}, \mathbf{x}_i, \mathbf{y}_i)}{p(\mathbf{x}_i, \mathbf{y}_i)}. \quad (3)$$

Here, $(\cdot|\cdot)$ denotes a conditional relationship, that is, the probability that future value x_{i+h} has a certain value when past values \mathbf{x}_i and \mathbf{y}_i are known. The term “transitional” as used by Schreiber [5] indicates conditional in time, that is, the information is transferred from past values \mathbf{x}_i to a value in the future x_{i+h} . If x_{i+h} is independent of \mathbf{x}_i and \mathbf{y}_i , then the transition

TABLE I
TYPES AND DIMENSIONS OF JOINT PDFs REQUIRED FOR
IMPLEMENTATION OF TRANSFER ENTROPY IN (4)

Joint PDF	Dimension	$k = 1, l = 2$
$p(x_{i+h}, \mathbf{x}_i, \mathbf{y}_i)$	$k + l + 1$	4
$p(\mathbf{x}_i, \mathbf{y}_i)$	$k + l$	3
$p(x_{i+h}, \mathbf{x}_i)$	$k + 1$	2
$p(\mathbf{x}_i)$	k	1

probability simplifies to $p(x_{i+h}|\mathbf{x}_i, \mathbf{y}_i) = p(x_{i+h})$ using (2). In that case, the knowledge of past values of x and y does not change the PDF of x_{i+h} . Another special case is when $k = 0$ and $l \neq 0$, so that the transition probability $p(x_{i+h}|\mathbf{y}_i)$ measures the causal relationship between x and y in the sense that y can be identified as the cause or driver of x .

B. Transfer Entropy

The transition probabilities contain all information on causality between two variables. As a function of $k + l + 1$ dimensions, this information is required to be represented by a single measure. The measure of transfer entropy proposed by Schreiber [5] extracts the amount of information transferred from variable y to x as follows:

$$t(x|y) = \sum p(x_{i+h}, \mathbf{x}_i, \mathbf{y}_i) \cdot \log \frac{p(x_{i+h}|\mathbf{x}_i, \mathbf{y}_i)}{p(x_{i+h}|\mathbf{x}_i)}. \quad (4)$$

The sum symbol represents $k+l+1$ sums over all amplitude bins of the discrete joint PDFs and transition probabilities. Table I gives the types and dimensions of the joint PDFs required for the transfer entropy computation. Even for small embedding dimensions, for example, $k = 1$ and $l = 2$, a four-dimensional (4-D) joint PDF over x_{i+h} , x_i , y_i , and y_{i-1} has to be constructed. The numerator of the logarithmic term, describing the probability of x_{i+h} if past values \mathbf{x} and \mathbf{y} are known, is compared to the denominator, describing the probability of x_{i+h} if only \mathbf{x} is known. If $t(x|y) = 0$, no information is transferred from y to x . In terms of transition probability this is the case if x_{i+h} is independent of \mathbf{x}_i and \mathbf{y}_i since, using (2) and (3), the logarithm of 1 is 0.

Transfer entropy is derived from the information measure of entropy. Entropy was introduced by Shannon [14] as a measure of uncertainty, or unpredictability summing a weighed logarithm of the PDF. Shannon also introduced mutual information to measure the mutual amount of information contained in two signals as defined in (1). Mutual information is, however, by its definition in (1) a symmetrical measure and, thus, cannot capture causality. Transfer entropy is chosen over mutual information because of its asymmetrical property. A causality measure is derived by comparing the influence of x on y with the influence of y on x

$$t_{x \rightarrow y} = t(y|x) - t(x|y). \quad (5)$$

Positive values of $t_{x \rightarrow y}$ then mean that x causes y , while negative values mean the reverse case. No causality is detected if both $t(y|x)$ and $t(x|y)$ have a similar value and, as a consequence, $t_{x \rightarrow y}$ is close to zero.

C. Significance Level

Small values of transfer entropy $t_{x \rightarrow y}$ suggest no causality or direction of influence while large values do. To establish a threshold above which $t_{x \rightarrow y}$ is recognized as a valid result Kantz and Schreiber suggest Monte Carlo methods using surrogate data. The threshold is referred to as significance or confidence level. Monte Carlo methods are any methods that solve a problem by generating suitable random numbers or sequences and observing that a fraction of the random numbers obey some property. Here, the problem is to verify or reject a null hypothesis. In the case of the directionality estimation, the null hypothesis is that the transfer entropy measure $t_{x \rightarrow y}$ is small, that is, it implies that x does not influence y by the result $t_{x \rightarrow y} = 0$. If a large value for $t_{x \rightarrow y}$ is measured, the null hypothesis can be rejected, that is, it implies that x influences y . Suitable random numbers are values of the directionality measure that are generated by computing the measure from N_s surrogate time series such that $\lambda_i = t_{x \rightarrow y}^{\text{sur}, i}$ with $i = 1, \dots, N_s$. For surrogate time series construction, the iterative amplitude adjusted Fourier transform (iAAFT) method as described by Schreiber and Schmitz [26], is used in all the following computations. The significance level is then defined as

$$s_{x \rightarrow y} = \frac{t_{x \rightarrow y} - \mu_\lambda}{\sigma_\lambda} > 6 \quad (6)$$

where μ_λ and σ_λ are mean and standard deviation of λ_i . A six-sigma threshold for the significance level is chosen here to give a robust decision rather than a two or three sigma test as in [27]. Since $t_{x \rightarrow y}$ does not necessarily follow a Gaussian distribution under the null hypothesis, however, it is not valid to assign a three-in-one million confidence level to the 6-sigma test.

IV. IMPLEMENTATION

A. Replacing Transition Probabilities by Joint PDFs

Before calculating the causality value in (5) from transfer entropy in (4), joint PDFs and transition probabilities have to be constructed from time series. Replacing the transition probability by joint PDFs according to (3), gives the joint PDFs as summarized in Table I required for the computation of $t_{x \rightarrow y}$. Estimation of the PDF from time series x and y is most commonly done with histograms, see for example Scott [28], but due to the high order of the joint PDFs (Table I), the number of samples required for the construction via histograms is extremely large. Here, Kernel estimation is proposed for the implementation (see, e.g., Silverman, [29]).

B. Kernel Estimator for PDF

The Kernel method gives a more precise estimation of the PDF than histograms by considering the exact values of a time series x . A Kernel function K is centered at every sample point and summed to give an estimate $\hat{p}(x)$

$$\hat{p}(x) = \frac{1}{N} \sum_{i=1}^N K(x - x_i). \quad (7)$$

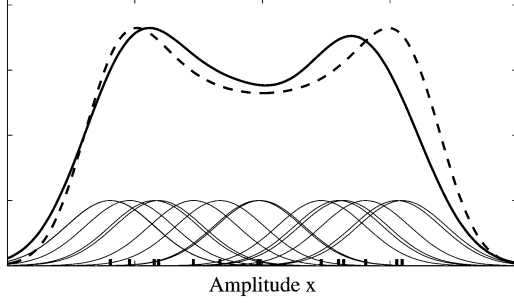


Fig. 1. Construction of Kernel estimator $\hat{p}(x)$ (solid line) for a sine wave with added noise using 15 samples of time trend. The dashed line is the analytical PDF $p(x)$.

Kernel function $K(x)$ has to fulfill $\int_{-\infty}^{\infty} K(x)dx = 1$, its maximum value must be at $x = 0$ and the limit values for plus and minus infinity are zero. Here, a Gaussian Kernel function is used which satisfies all requirements

$$K(x - x_i) = \frac{1}{\sqrt{2\pi}\theta} \exp\left(-\frac{(x - x_i)^2}{2\theta^2}\right) \quad (8)$$

where θ is the estimator width which is adjusted to the number of samples N and the standard deviation of time series x after Silverman [29]. Fig. 1 shows the construction of an estimated PDF $\hat{p}(x)$ using the Kernel method. Even for a small number of samples, only 15 in this example, the Kernel method gives a good estimation of the analytical PDF.

In [29], the amplitude axis is considered as continuous. For implementation, however, it has to be discretized to allow computation in a software code. A discrete value k_i is introduced that translates x_i onto a discrete grid $k_i \in [1 : n]$

$$k_i = \left[(n-1) \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \right]_{\text{round}} + 1 \quad (9)$$

where the $[]_{\text{round}}$ function rounds to the nearest integer. The number of amplitude bins is denoted by n and can be set independently. The PDF is constructed according to (7) with discrete Kernel function

$$K[k - k_i] = \frac{1}{a\sigma\sqrt{\pi}} \exp\left(-\frac{(k - k_i)^2}{\sigma^2}\right) \quad (10)$$

with $k = k_i - \sigma, k_i - \sigma + 1, \dots, k_i + \sigma$. Due to the finite length of $K[k - k_i]$, a scaling by factor $a = \sum_{k=-\sigma}^{\sigma} (1/\sigma\sqrt{\pi}) \exp(-k^2/\sigma^2)$ is required. For the discrete Kernel, estimator width θ is replaced by the discrete Kernel width σ . The optimal Kernel width is

$$\sigma = \left[cN^{-1/5} \sigma_x n \right]_{\text{round}} \quad \text{with } c = 0.2. \quad (11)$$

Here, σ_x is the standard deviation of time series x . For a derivation of σ (see [29]). The translation from continuous to discrete x -axis for the PDF estimation is shown in Fig. 2. The total number of bins is $n + 2\sigma$ with $\sigma \ll n$.

C. Kernel Estimator for Joint PDF

The estimation of the joint PDF can be constructed in a similar manner. The estimation of the joint PDF for x and y using

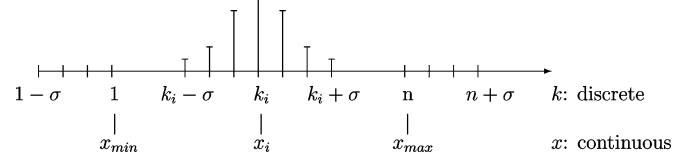


Fig. 2. Transformation from continuous to discrete amplitude axis according to (9). The total number of bins is $n + 2\sigma$.

the Kernel method is denoted by

$$\hat{p}(x, y) = \frac{1}{N} \sum_{i=1}^N K(x - x_i, y - y_i). \quad (12)$$

The Kernel function for a Gaussian Kernel is then

$$K(x - x_i, y - y_i) = \frac{1}{2\pi\theta^2} \exp\left(-\frac{(x - x_i)^2 + (y - y_i)^2}{2\theta^2}\right). \quad (13)$$

For the discrete case of joint PDFs, simultaneous discretization of amplitude axes x and y into n amplitude bins has to be carried out. Higher order joint PDFs can be constructed as an extension to (12). Transition probabilities are decomposed into joint PDFs according to (3).

D. Computational Effort

Computing speed and the finite number of samples limits choice of embedding dimensions k and l . The estimation of the computational effort is separated into two parts. First, the construction of the PDF has to be considered and, second, the calculation of the transfer entropy value using (4). For each joint PDF of dimension m , $(2\sigma + 1)^m N$ summations are required. Considering the joint PDFs required listed in Table I, this adds up to $q^k (q + 1) (q^l + 1) N$ summations with $q = 2\sigma + 1$ as the length of the Kernel. Equation (11) shows that σ increases with the number of bins n and with the number of samples N . The computation of transfer entropy requires $(n + 2\sigma)^{k+l+1}$ summations. Embedding dimensions k and l have to be chosen carefully for two reasons. First, the number of computations increase with the power of k and l . Second, also the number of samples required to estimate the high-dimensional joint PDFs increases with the power of k and l . Thus, k and l should be chosen as small as possible. Schreiber suggests to set l larger than k to focus on y influencing x rather than x influencing x . For lower embedding dimension good results can be achieved by setting k to zero. Setting l to 2 results in a maximum of three dimension of the joint PDFs after Table I and can be dealt with if the number of samples available are in the order of 1000 when using the Kernel method. For the choice of $k = 0$ and $l = 2$, which will be used in the following, a total of $4N(2\sigma^3 + 4\sigma^2 + 3\sigma + 1) + (n + 2\sigma)^3$ summations are required.

V. CAUSAL MAP

A. Prerequisites

For industrial applications, the relationship between several measurements is considered, say R measurements $x^{(r)}$ with $r = 1, \dots, R$, and a representation of the results in a causal

TABLE II
EXAMPLE FOR REORDERING OF VARIABLES $x^{(r)}$

	$x^{(1)}$	$x^{(2)}$	$x^{(3)}$	$x^{(4)}$	$x^{(5)}$
$x^{(1)}$	0	0	0	0	0
$x^{(2)}$	1	0	0	0	1
$x^{(3)}$	0	1	0	0	1
$x^{(4)}$	0	0	0	0	0
$x^{(5)}$	1	0	0	0	0

	$x^{(2)}$	$x^{(3)}$	$x^{(5)}$	$x^{(1)}$
$x^{(2)}$	0	0	1	1
$x^{(3)}$	1	0	1	0
$x^{(5)}$	0	0	0	1
$x^{(1)}$	0	0	0	0

	$x^{(3)}$	$x^{(2)}$	$x^{(5)}$	$x^{(1)}$
$x^{(3)}$	0	1	1	0
$x^{(2)}$	0	0	1	1
$x^{(5)}$	0	0	0	1
$x^{(1)}$	0	0	0	0

map or digraph is desired. The causality measure of (5) is computed for all combinations of variables $x^{(r)}$, that is, $R(R-1)/2$ relationships. The results can be denoted in a causality matrix as follows:

$$\mathbf{t} = \begin{bmatrix} 0 & t_{x^{(1)} \rightarrow x^{(2)}} & \dots & t_{x^{(1)} \rightarrow x^{(R)}} \\ -t_{x^{(1)} \rightarrow x^{(2)}} & 0 & \dots & t_{x^{(2)} \rightarrow x^{(R)}} \\ \vdots & \vdots & \ddots & \vdots \\ -t_{x^{(1)} \rightarrow x^{(R)}} & -t_{x^{(2)} \rightarrow x^{(R)}} & \dots & 0 \end{bmatrix} \cdot (14)$$

The rows represent cause variables, while the columns represent the effect variables. The task is to generate a causal map that shows the relationships between all $x^{(r)}$ and $x^{(s)}$ in an automated way. Due to symmetry, negative causality measures are ignored without losing any information. All relationships with a significance level larger than $6\sigma_\lambda$ are accepted and treated equally, while those below $6\sigma_\lambda$ are ignored. This treatment gives a matrix Λ consisting of 1's and 0's in which the 1's mean that the column variable influences the row variable. A zero entry indicates that no relationship exists. The top panel of Table II shows an example. Using 1's and 0's removes information about the sizes of the elements in (14). The numerical values are used, however, to resolve conflicts between two options of fault propagation paths as described in the following algorithm.

The automatic generation of a causal map from the causality matrix Δ is accomplished in two steps. First, the order of measurements $x^{(r)}$ is rearranged to $\bar{\Delta}$ to bring a maximum number of entries above the main diagonal. In the second step, the causal map is constructed from $\bar{\Delta}$. The automatic construction of a causal map is shown here for a model example and later on in the paper for the test case study.

Algorithm to Rearrange the Order of Variables

The new order of variables that maximizes the number of non-zero entries above the main diagonal is found by the following procedure.

- *Step 1: Initial sorting.* The measurements $x^{(r)}$ are sorted by the number of non-zero entries in the r th row to create a new causality matrix Δ' whose first row has the least number of zero entries. The new indices are r' . If there is a

tie then the number of non-zero entries in the r' th column is used as a tie breaker. In case of a further tie, the order remains unchanged.

- *Step 2: Establishing rules.* The entries above the main diagonal of Δ' are translated into inequality rules to ensure that the element remains above the main diagonal when further sorting. For example, the element in the second row and third column of Δ' in Table II gives the inequality $x^{(3)} < x^{(5)}$.
- *Step 3: Shifting elements above diagonal.* A nonzero element is below the diagonal in the r' th row and s' th column. $x^{(r')}$ and $x^{(s')}$ can be exchanged if no previously established rule is violated as a consequence of the exchange. If a rule is violated, alternative positions with row index smaller than r' for $x^{(s')}$ or larger than s' for $x^{(r')}$ can be explored in the same way. In the case that no position change without a rule violation can be found, the option with the highest transfer entropy value, that is, the summed value for all rules in question, is chosen. A new rule is established for the exchanged variables.

Repeat Step 3 for all elements below the main diagonal. The resulting matrix is the optimized causality matrix $\bar{\Delta}$, the order of the rows is denoted by \bar{r} .

The algorithm is illustrated using a model example in Table II. The top table shows the example causality matrix Δ for five process variables $x^{(1)}, \dots, x^{(5)}$. In the initial sorting step, the variables are rearranged with $x^{(2)}$ and $x^{(3)}$ both with two entries in each row and $x^{(5)}$ one entry. $x^{(4)}$ is excluded from further analysis and the modified causality matrix is shown in the middle table of Table II. Rules for the four elements above the main diagonal are established in Step 2, that is, $x^{(2)} < x^{(5)}$, $x^{(2)} < x^{(1)}$, $x^{(3)} < x^{(5)}$, and $x^{(5)} < x^{(1)}$. The element below the diagonal for $x^{(3)}$ and $x^{(2)}$ is dealt with by Step 3. $x^{(3)}$ and $x^{(2)}$ can be exchanged since none of the established rules are violated. Thus, the modified causality $\bar{\Delta}$ results as shown in the bottom table of Table II.

Algorithm to Construct the Causal Map

Once the modified causality matrix $\bar{\Delta}$ is retrieved the construction of a causal map is straightforward using the following steps.

- *Step 1: Initializing layout.* To generate a basic structure, $x^{(\bar{r})}$ are placed in a row according to new order \bar{r} . Variables that were excluded from analysis in Step 1 can be omitted since they have no causal connection with all other variables.
- *Step 2: Inserting arcs.* The relationships for all non-zero entries in causality matrix $\bar{\Delta}$ can be represented by arcs pointing from cause variable in the detected row to effect variable in the corresponding column. These arcs are used to represent the information flow.
- *Step 3: Removing shortcuts.* In some situations, if a variable A causes a variable B and B in turn causes C, a further dependency between A and C will be detected. This dependency is only secondary and not of main interest. It can be easily ignored by removing all shortcut arcs, that is, all arcs from variable $x^{(r)}$ to variable $x^{(s)}$ for which two relationships via a third variable exist.

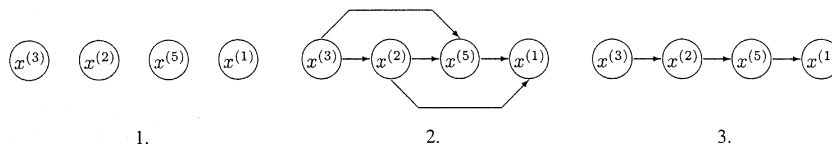


Fig. 3. Construction steps of causal map for examples in Table II. Step 1: initializing layout; Step 2: inserting arcs; Step 3: removing shortcuts.

The variables on the left of the causal map are then the suspected root cause. The procedure is illustrated by the model example in Fig. 3. In the left-hand plot representing Step 1, all variables are laid out in the new row order \bar{r} . In Step 2, arcs for all relationships are drawn representing non-zero entries of causality matrix $\bar{\Delta}$ in Table II. In the right-hand plot of Fig. 3, the third step is illustrated in which all shortcuts are removed. For example, the shortcut from $x^{(3)}$ to $x^{(5)}$ is removed since a way from $x^{(3)}$ to $x^{(5)}$ via $x^{(2)}$ exists. The resulting causal map represents the process dependency detected with the transfer entropy measure. $x^{(3)}$ is suspected to be the root cause of the investigated disturbance.

The main advantage of the algorithm is that it focuses on the order of occurrence of events which is desired for fault propagation and root cause analysis. The most important question of which variable comes first and is followed by which other variable is addressed. The situation of branches, that is, A influences B and C, can also be represented by the automatically generated causal map that are placed in the order A, B, C or A, C, B with arcs drawn from A to B and A to C.

VI. PARAMETER OPTIMIZATION

A. Default Parameter Values

The calculation of transfer entropy requires a number of parameters to be set. It will be shown in the following that when applied to chemical process data, the transfer entropy measure is insensitive to the parameter setting within bounds. Parameters will be optimized by applying them to data of a reference case study. The reference case study is an industrial process in which a disturbance occurs that travels along the fault propagation paths affecting a number of process variables. The direction of cause and effect and the relationships between two measurements is known in this case from first principles. The transfer entropy parameters are optimized using these relationships. The parameter settings are then applied to a test case study for which the root cause was unknown prior to analysis. The causal map is constructed and a suggestion of the root cause is made.

Fixed parameters are the embedding dimensions which are set to $k = 0$ and $l = 2$ due to computational restraints as argued above. The special case $k = 0$ means that only past values of y and not x are regarded when estimating a future value of x so that for instance $t(\text{TI1}|\text{TI2})$ in (4) would be calculated using past values of TI2 and the current value of TI1 while $t(\text{TI2}|\text{TI1})$ uses past values of TI1 and the current value of TI2. Adjustable parameters are the time interval τ between past value of y , such that $\mathbf{y} = [y_i, y_{i-\tau}]$, and the prediction horizon h of the future value of x , x_{i+h} in (4). The optimal parameters of both time interval and prediction horizon are a function of the process dynamics. If the process dynamics are known, the parameters can

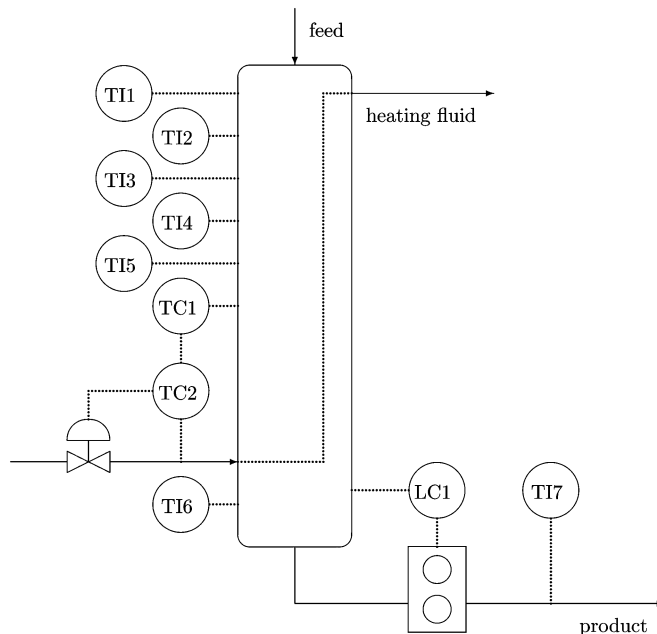


Fig. 4. Process schematic of reference case study, industrial reaction process at Eastman Chemical Company.

be set accordingly. If a dead time is detected between two measurements the optimum values of both τ and h are equal to the dead time. This can be concluded from the definition of transfer entropy in (4) which predicts a future value from the knowledge of the past values which are separated by h and τ time steps. However, if the process dynamics are unknown, small values of τ and h such as $h = \tau \leq 4$ should give good results as the transfer entropy measure should be robust for parameter changes. The minimum number of sample N required for computation is also investigated in the following.

B. Reference Case Study

A reaction process at Eastman Chemical Company was chosen in which a disturbance from the upstream feed travels along the process flow. The process schematic is shown in Fig. 4. Five temperature measurements are taken at the top part of the reactor above a point where heating fluid is pumped through the tray structure to control the temperature. Additional temperature measurements are taken at the bottom part of the reactor and further downstream. The level at the bottom tray of the reactor is a crucial variable that is closely observed. The disturbance enters the process through the reactor feed and affects, to some extent, all parts of the process. Fig. 5 shows the normalized time trends of the measurements. The sampling time of the data is 20 s. A disturbance with an average period of oscillation of 60 samples or 20 min can be seen in Fig. 5 for most of the variables. Only 1000 samples are shown here,

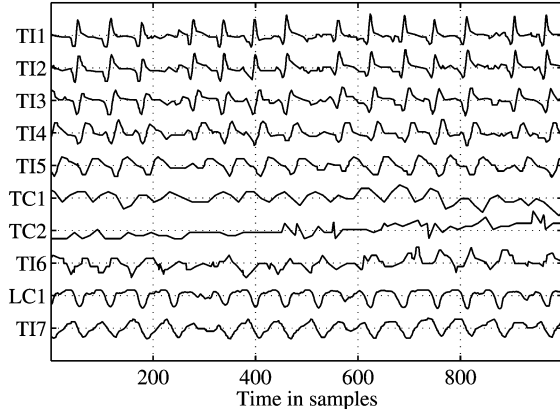


Fig. 5. Time trends of measurements in reference case study during period of disturbance.

but altogether 4100 samples of a time period exhibiting the disturbance are available.

Physical investigation of the process showed that the disturbance originated further upstream of the reactor, affecting TI1 first followed by TI2 to TI5 and then propagating further till TI7. For parameter optimization, four relationships between TI1 and TI2, TI2 and TI3, TI3 and TI4, TI4 and TI5 are selected. Another selected relationship is the one between LC1 and TI7 since their time trends show the disturbance clearly. These five relationships were known to be significant from process knowledge.

1) *Time Interval*: In a first step, the time interval is varied while the other parameters are fixed. Time interval τ was defined for the transition probabilities in (3) which uses embedded vectors $\mathbf{x}_i = [x_i, x_{i-\tau}, \dots, x_{i-(k-1)\tau}]$ and $\mathbf{y}_i = [y_i, y_{i-\tau}, \dots, y_{i-(k-1)\tau}]$. A change of the time interval corresponds to subsampling of the data series with the benefit of no data being disregarded. Adjusting the time interval ensures that the dynamics of the underlying system are represented accurately. Fig. 6 shows transfer entropy values and significance level for the five relationships of the reference case study that were known to be significant from process knowledge. The significance level is above the threshold of six sigma for most time intervals which shows that transfer entropy is robust. The maximum value of $s_{x \rightarrow y}$ is at different values of τ for the five plots, varying from $\tau = 2$ for TI1 \rightarrow TI2 to $\tau = 9$ for LC1 \rightarrow TI7. When summing all $s_{x \rightarrow y}$ over the five relationships the maximum lies at $\tau = 4$. Thus, the time interval is set to $\tau = 4$ which gives a significance level well above the threshold for all five relationships. With a sampling rate of 20 s, time interval τ is equivalent to 80 s. In general, time interval τ is dependent on the dynamics of the process. However, Fig. 6 shows that the result is robust to variations in τ so that a certain mismatch between process dynamics and time intervals can still give good results. The computation of the transfer entropy method with a standard PC is sufficiently fast to evaluate results for a range of τ and h values if desired.

2) *Prediction Horizon*: In the previous step, the prediction horizon h was set to the τ while optimizing time interval τ . It is expected that the value of the prediction horizon h lies in the same range as the time interval since dynamics of the underlying system are also reflected by the prediction horizon. Fig. 7

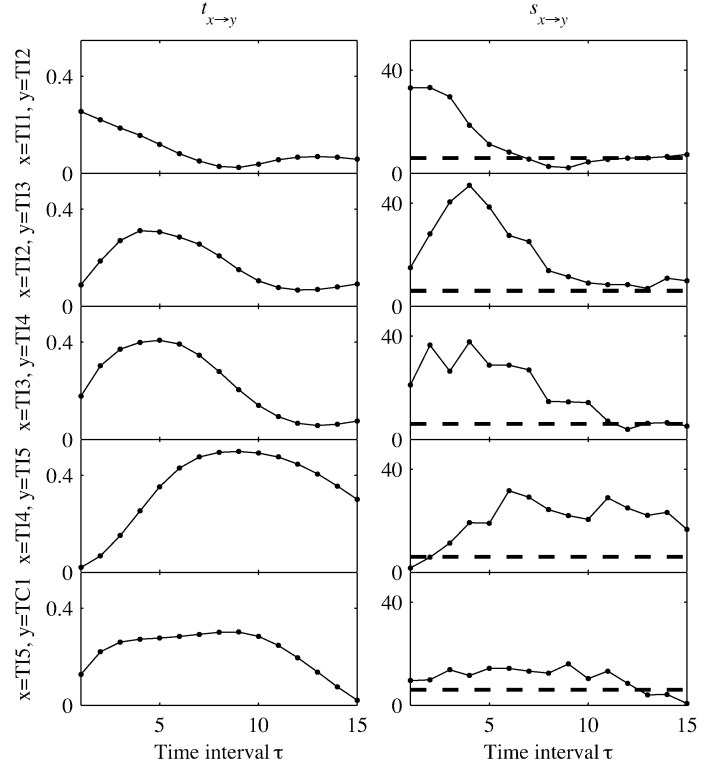


Fig. 6. Optimizing time interval τ for five dependencies; left-hand plots: transfer entropy value $t_{x \rightarrow y}$, right-hand plots: significance value $s_{x \rightarrow y}$; with $h = \tau$ and $N = 4100$.

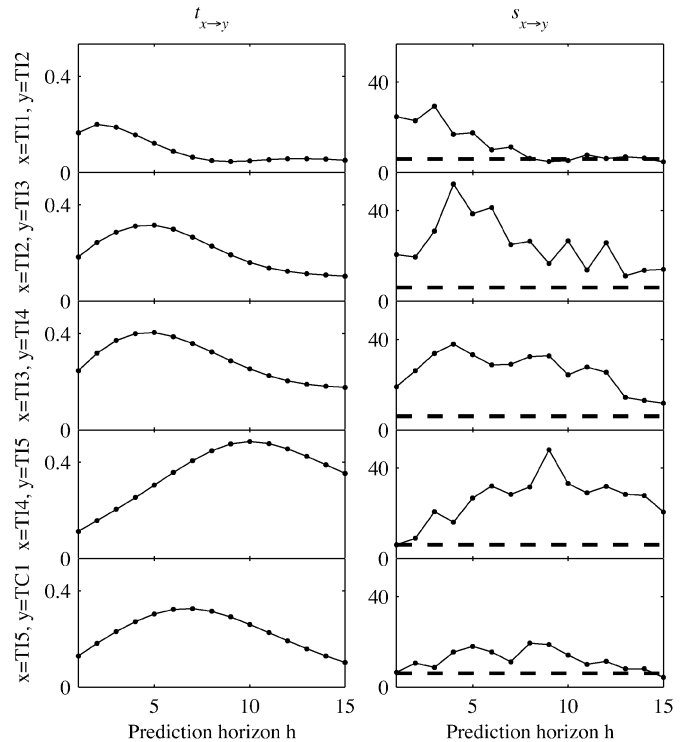


Fig. 7. Optimizing prediction horizon h for five dependencies; left hand plots: transfer entropy value $t_{x \rightarrow y}$, right hand plots: significance value $s_{x \rightarrow y}$; with $\tau = 4$ and $N = 4100$.

shows transfer entropy and significance value as a function of h . The dependency of transfer entropy on h is similar to the dependency on τ as in Fig. 6. The values of $t_{x \rightarrow y}$ and $s_{x \rightarrow y}$ are robust

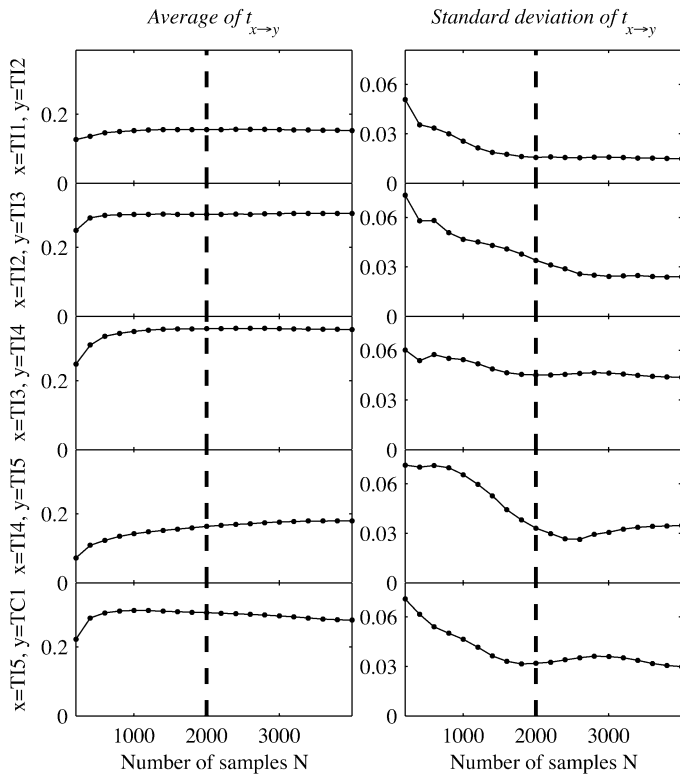


Fig. 8. Finding minimum number of samples N_{\min} for five dependencies; left-hand plots: average of transfer entropy $t_{x \rightarrow y}$, right-hand plots: standard deviation of $t_{x \rightarrow y}$; with $h = \tau = 4$.

against the choice of h . The maximum value of $s_{x \rightarrow y}$ varies for the five relationships but the sum of all significance levels has its maximum at $h = 4$. Thus, the prediction horizon is chosen to be equal to the time interval. The data investigated in the reference case study is cycling with a period of oscillation of 60 samples. Setting time interval and prediction horizon to $\tau = h = 4$ corresponds to subsampling by factor 4. Thus, the number of samples per oscillation period are reduced to 15.

3) *Minimum Number of Samples:* For practical application the number of samples to be considered for computation is a crucial parameter. The question is how many samples are required to get a significant result, that is, the minimum number of samples N_{\min} . The impact of the number of samples on the transfer entropy measure $t_{x \rightarrow y}$ is shown in Fig. 8. Average value and standard deviation of $t_{x \rightarrow y}$ are calculated using subsets of the original data set of length N . The subsets are overlapping and delayed by 100 samples. For example, with the total length being 4100 samples, 40 subsets are constructed for $N = 200$ consisting of samples 1 to 200, 101 to 300, 201 to 400, and so on, down to two subsets for $N = 4000$. The left-hand plots of Fig. 8 shows that the average value of $t_{x \rightarrow y}$ does not vary significantly when $N \geq 600$. However, the results of $t_{x \rightarrow y}$ for the subsets can vary as shown in the right-hand plots. The standard deviation only levels out for approximately $N > 2000$ samples.

The recommendation is that the minimum number of samples should be set to $N_{\min} = 2000$ if possible. If, however, fewer than 2000 samples are available, it is still worthwhile to carry out transfer entropy analysis down to 400 or 500 samples. The confidence level in the result will then be lower as a consequence.

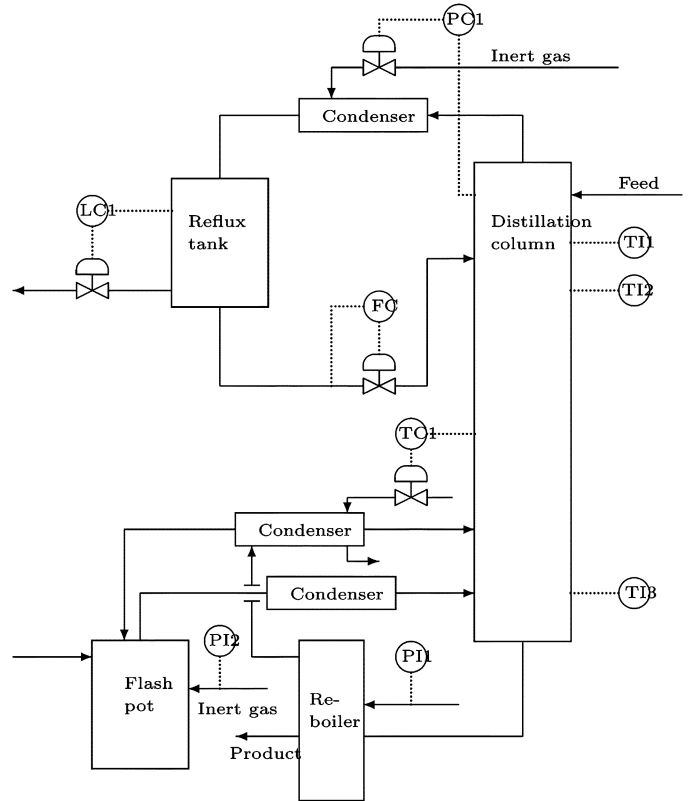


Fig. 9. Process schematic of test case study.

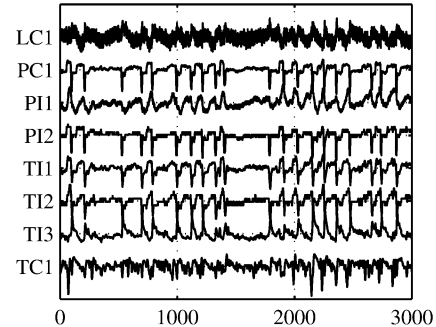


Fig. 10. Time trends of measurements in test case study.

C. Test Case Study

The test case study is part of a production process at Eastman Chemical Company. The centerpiece of the process is a distillation column in which the product is refined. The distillation column operates under vacuum and the top of the column is equipped with a total condenser and a reflux drum. The reflux drum level LC1 is controlled by manipulating the distillate valve. Pressure is controlled by adjusting the flow of inert gas to the top of the column. The temperature profile in the column is maintained by controlling a temperature in the middle of the column (TC1) through changes in the heat allowed to enter the column. It is important to maintain this temperature to target in order to achieve proper separation of impurities from the product. Several other temperatures are provided for monitoring of the temperature profile (TI1, TI2, TI3). A pressure and temperature gradient can be observed along the column with pressure and temperature being functionally related. The reboiler

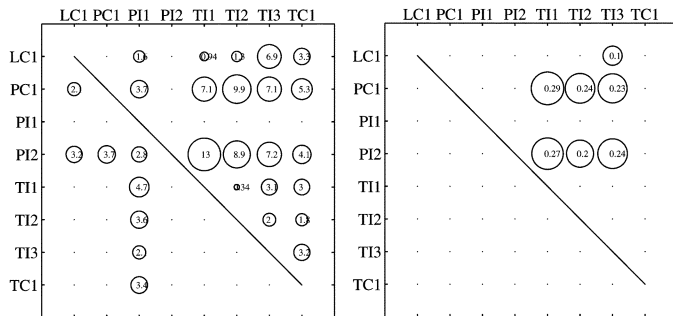


Fig. 11. Results of case study two. Left plot: significance level $s_{x \rightarrow y}$, right plot: transfer entropy values $t_{x \rightarrow y}$ for significance level above threshold of six. $k = 0, l = 2, N = 8640, \tau = h = 4$.

TABLE III

GUIDELINES FOR PARAMETER SETTING AS RESULTS OF REFERENCE CASE STUDY WITH A SAMPLING RATE OF 20 s

Parameter	Guidelines
Embedding dimensions	$k = 0, l = 2$
Minimum number of samples	$N_{min} = 2000$
Time interval	$\tau = 4$
Prediction horizon	$h = \tau = 4$

TABLE IV

CAUSALITY MATRIX OF TEST CASE STUDY AFTER REORDERING

	PC1	PI2	LC1	TI3	TI1	TI2
PC1	0	0	0	1	1	1
PI2	0	0	0	1	1	1
LC1	0	0	0	1	0	0
TI3	0	0	0	0	0	0
TI1	0	0	0	0	0	0
TI2	0	0	0	0	0	0

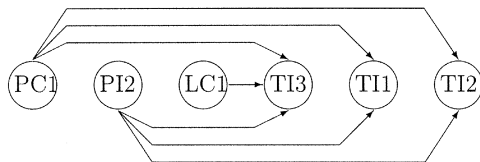


Fig. 12. Causal map of test case study.

chest pressure and the inert gas pressure further on in the process are monitored by PI1 and PI2, respectively.

The part of the process shown in Fig. 9 was selected because a troublesome disturbance affected the measurements around the column. The time trend of the disturbance is shown in Fig. 10 with 3000 samples taken every 10 s. Altogether, 8640 samples were available. A common pattern can be seen in all time trends. Sharp spikes occur around the same time in the level, pressure, and temperature measurements although with some interruptions and, thus, not necessarily regular. The root cause of the disturbance was unknown prior to the analysis.

The causality measure $t_{x \rightarrow y}$ was calculated from transfer entropy of the time trends to find conclusions towards the root cause of the disturbance. The significance level is shown in the upper panel of Fig. 11. The parameters that were found to be the best suited in the reference case study are listed in Table III and are used for the test case study. The results are shown in form of a bubble chart in which size of the bubble on the intersection of the horizontal and vertical axis gives the significance level. For

example, the bubble on the intersection of row PI2 column TI1 is 13 which means that PI2 causes TI1. The lower plot shows all transfer entropy values for which the significance level exceeds the threshold of 6 [see (6)]. Rearranging the transfer entropy matrix according to the first step described the reordering algorithm in Section V with all elements above the main diagonal, see Table IV.

The causality matrix is transferred into a causal map as described above resulting in Fig. 12. Possible root causes suggested by the causal map are both PC1 and PI2 which each affect an equal number of following variables. Referring to the process schematic in Fig. 9 shows that PC1 and PI2 are two pressures connected with the inert gas flow. Further investigation showed that the inert gas inflow into the condenser and the flash pot was coming from the same pipe controlled by a split pressure controller. Investigations revealed that the pressure controller had an oversized control valve and caused the inert gas flow to be erratic, thus, affecting the remaining process. The valve sizing was corrected and a further data trend showed that the disturbance went away. Thus, transfer entropy using the proposed settings was able to identify the root cause correctly.

VII. CONCLUSION

Disturbances in chemical processes often spread from one process variable to neighboring variables. It is of interest to measure the direction of propagation to locate the root cause of the disturbance. The data-driven method of transfer entropy measures the causality between two time series and, thus, the direction of propagation. The crucial parameters of transfer entropy have been investigated and guidelines for its application to chemical processes derived from experimental data were proposed. The reference case study showed that a data ensemble of 2000 samples gave reliable results. Transfer entropy is based on estimates of probability density functions which requires a large amount of data. However, with the capacity of modern data historians, historical process data is readily available. The successful applications to a test case study highlights the use of transfer entropy given the derived parameters. An automated algorithm for constructing a causal map from the directionality detected by the transfer entropy measure is applied to the test case study.

ACKNOWLEDGMENT

The authors would like to thank all members of the Advanced Control Technologies Group at Eastman Chemical Company for feedback and discussion.

REFERENCES

- [1] L. Desborough and R. Miller, "Increasing customer value of industrial control performance monitoring—Honeywell's experience," in *Proc. AIChE Symp. Ser.*, 2002, vol. 98, pp. 153–186.
- [2] T. Kourti, "Process analysis and abnormal situation detection: from theory to practice," *IEEE Contr. Syst. Mag.*, vol. 22, no. 5, pp. 10–25, Oct. 2002.
- [3] ExperTune, Inc., Hartland, WI, Plant Triage, 2005. [Online]. Available: <http://www.expertune.com/plantrriage.html>
- [4] Matrikon, Inc., Edmonton, AL, Canada, Process Doctor, (2006). [Online]. Available: <http://www.matrikon.com/products>
- [5] T. Schreiber, "Measuring information transfer," *Phys. Rev. Lett.*, vol. 85, pp. 461–464, 2000.

- [6] M. Bauer, N. F. Thornhill, and A. Meaburn, "Specifying the directionality of fault propagation paths using transfer entropy," presented at the DYCOPS 7, Boston, MA, Jul. 2004, paper 62, CD ROM.
- [7] V. Venkatasubramanian, R. Rengaswamy, K. Yin, and S. N. Kavuri, "A review of process fault detection and diagnosis—Part I: Quantitative model-based methods," *Comput. Chem. Eng.*, vol. 56, pp. 293–311, 2003.
- [8] —, "A review of process fault detection and diagnosis—Part II: Qualitative models and search strategies," *Comput. Chem. Eng.*, vol. 56, pp. 313–326, 2003.
- [9] —, "A review of process fault detection and diagnosis—Part III: Process history based methods," *Comput. Chem. Eng.*, vol. 56, pp. 327–346, 2003.
- [10] L. H. Chiang, E. L. Russell, and R. D. Braatz, *Fault Detection and Diagnosis in Industrial Systems*. London: Springer-Verlag, 2001.
- [11] L. H. Chiang and R. D. Braatz, "Process monitoring using causal map and multivariate statistics: Fault detection and identification," *Chem. Intell. Lab. Syst.*, vol. 65, pp. 159–178, 2003.
- [12] M. R. Maurya, R. Rengaswamy, and V. Venkatasubramanian, "A systematic framework for the development and analysis of signed digraphs for chemical processes. 1. Algorithms and analysis," *Ind. Eng. Chem. Res.*, vol. 42, pp. 4789–4810, 2003.
- [13] —, "A systematic framework for the development and analysis of signed digraphs for chemical processes. 2. Control loops and flowsheet analysis," *Ind. Eng. Chem. Res.*, vol. 42, pp. 4811–4827, 2003.
- [14] C. E. Shannon and W. Weaver, "A mathematical theory of communications," *Bell Syst. Tech. J. (reprint)*, vol. 27, pp. 379–423, 1948.
- [15] A. Kaiser and T. Schreiber, "Information transfer in continuous processes," *Physica D*, vol. 166, pp. 43–62, 2002.
- [16] R. Marschinski and H. Kantz, "Analysing the information flow between financial time series—an improved estimator for transfer entropy," *Eur. Phys. J. B*, vol. 30, pp. 275–281, 2002.
- [17] L. M. Pecora, T. L. Carroll, and J. F. Heagy, "Statistics for mathematical properties of maps between time series embedding," *Phys. Rev. E*, vol. 52, pp. 3420–3439, 1995.
- [18] D. Prichard and J. Theiler, "Generalized redundancies for time series analysis," *Physica D*, vol. 84, pp. 476–493, 1995.
- [19] J. Bhattacharya, E. Pereda, and H. Petsche, "Effective detection of coupling in short and noisy bivariate data," *IEEE Trans. Syst., Man, Cybern. B*, vol. 33, no. 1, pp. 85–95, Feb. 2003.
- [20] U. Feldmann and J. Bhattacharya, "Predictability improvement as an asymmetrical measure of interdependence in bivariate time series," *Int. J. Bifurc. Chaos*, vol. 14, pp. 505–514, 2004.
- [21] B. Bezruchko and V. Ponomarenko, "Characterizing direction of coupling from experimental observations," *Chaos*, vol. 13, pp. 179–184, 2003.
- [22] T. I. Netoff, L. M. Pecora, and S. J. Schiff, "Analytical coupling detection in the presence of noise and nonlinearity," *Phys. Rev. E*, vol. 69, p. 017201, 2004.
- [23] M. Bauer, J. W. Cox, and N. F. Thornhill, "Measuring cause and effect between process variables," presented at the EEE APC Appl. Ind. Workshop, Vancouver, Canada, May 2005, CD ROM.
- [24] B. Girod, R. Rabenstein, and A. Stenger, *Signals & Systems*. New York: Wiley, 1990.
- [25] X.-R. Li, *Probability, Random Signals and Statistics: A Textgraph With Integrated Software for Electrical and Computer Engineers*. Boca Raton, FL: CRC, 1999.
- [26] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1997.
- [27] T. Schreiber and A. Schmitz, "Surrogate time series," *Physica D*, vol. 142, pp. 346–382, 2000.
- [28] D. W. Scott, "On optimal and data-based histograms," *Biometrika*, vol. 66, pp. 605–610, 1979.
- [29] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*. London, U.K.: Chapman & Hall, 1986.



Margret Bauer was born in Germany in 1976. She received the M.Sc. degree in electrical engineering from the University Erlangen-Nuremberg, Germany, in 2000, and the Ph.D. degree from University College London (UCL), U.K., in 2005.

From 2001 to 2002, she worked for Infineon Technologies in Germany and Sweden. She currently holds a Claude Leon Postdoctoral Fellowship at the University of Pretoria, South Africa. Her research interests include data-driven process monitoring techniques and the economic assessment of advanced

process control.

Dr. Bauer was a recipient of the UCL Graduate School Research Scholarship.



John W. Cox received the B.S. degree in chemical engineering from Tennessee Technological University, Cookeville, in 1992 and the M.S. degree in chemical engineering from the University of Tennessee, Knoxville, in 1998.

For the past 14 years, he has worked as a Process Control Engineer in the Advanced Controls Technology Group at Eastman Chemical Company, focusing his efforts on the application and improvement of regulatory and advanced control for a variety of Eastman processes. In addition, he codeveloped Eastman's company-wide controller assessment system. His research interests include controller performance assessment, diagnosis of control valve stiction, and oscillation detection and diagnosis for plantwide datasets.



Michelle H. Caviness received the chemical engineering degree from the Georgia Institute of Technology, Atlanta, in 1996.

She accepted a position with Eastman Chemical Company's Advanced Control Technology Group in 1996. Her projects have included reactor and column control strategy design, process modeling for control strategy analysis, and process evaluation for engineering support in polymer and chemical processing facilities. She has been responsible for the design of control strategies in several capital expansion projects and has provided start-up support for processes at locations in both the U.S. and Asia. Her recent efforts include design and implementation of advanced plant-wide control strategies to increase production and minimize energy consumption and serving as the project lead for design of a distillation based solvent recovery facility. She has co-taught several offerings of an advanced control technology course and distillation control course to the industrial process control community.



James J. Downs is manager of the Advanced Controls Technology group at Eastman Chemical Company. He has 25 years of experience in the design, startup, and support of industrial processes. His work has included the design of advanced control systems for reaction systems, distillation processes, polymer processes, extraction and other separations processes, gas handling systems and compressors, and other unit operations. He developed regulatory and overall plantwide operation strategies for numerous capital expansion projects combining plant design and control system design techniques. He developed rigorous distillation simulation programs capable of handling simultaneous distillation and chemical reaction for both steady state and dynamic (transient) modeling of distillation columns. His current research interests include plantwide control strategy design, plantwide process optimization, and the process design/process control interface. He has participated in many workshops and symposia concerning process control research and process control education.

Dr. Downs was recognized by the AIChE for his contribution to the profession by receiving the CAST Computing Practice Award in 1996.



Nina F. Thornhill (SM' 05) received the B.A. degree in physics from Oxford University, Oxford, U.K., in 1976, the M.Sc. degree in control systems from Imperial College London, U.K., in 1983, and the Ph.D. degree from University College London (UCL), U.K.

She worked for ICI from 1976 to 1982, then joined UCL in 1984, and 1992 to 1995 held an EPSRC/Royal Society Industry Fellowship with British Petroleum. She is currently a professor in the Department of Electrical and Electronic Engineering and a member of the Imperial College/UCL Centre for Process Systems Engineering. Her research interests include process data analysis and plantwide performance assessment with applications in oil, chemicals, and pharmaceuticals.