

SPECIFYING THE DIRECTIONALITY OF FAULT PROPAGATION PATHS USING TRANSFER ENTROPY

Margret Bauer*, Nina F. Thornhill*, Adrian Meaburn⁺

**Department of Electronic and Electrical Engineering, Torrington Place,
University College London, London WC1E 7JE, UK.*

⁺BP Chemicals, Salt End, Hull, HU12 8DS, UK.

Abstract: In continuous chemical processes, variations of process variables usually travel along propagation paths in the direction of flow. The aim of this study was to find a data-driven method for identifying the direction of variation propagation using historical process data. Transfer entropy is a recently proposed method based on the probability density function (PDF) that measures directionality of variation with respect to time. An industrial case study illustrates the method which detects the influence of a temperature controller on downstream temperature measurements. A reversal of directionality was noted during a disturbance and a physical explanation offered.

Keywords: Chemical industry; fault diagnosis; plant-wide; probability density function; propagation; statistical process control.

1. INTRODUCTION

Plant-wide disturbances occur if a fault distributes along a fault propagation path and affects several process variables. The detection and diagnosis of plant-wide disturbances is a major problem in the process industry (Qin, 1998). It is therefore important to understand the mechanism of propagation.

A description of the propagation mechanism in causal qualitative models can identify the order of occurrence of events and specify the paths of fault propagation. Thus, the root cause of the disturbance can be retraced along the propagation path.

This paper describes work in progress on a data-driven method that serves as a basis for the automatic construction of digraphs. The entropy related

measure of transfer entropy is chosen for determining the direction of fault propagation and to answer questions of

- Directionality: Does a measurement X influence a second measurement Y more than measurement Y influences X?
- Consistency: Does the directionality change if a fault is present?

Digraph based models usually express the relationship between faults and symptoms and define the propagation paths by incorporating process knowledge of experts (Nam et al., 1996). A drawback is that extracting expert knowledge is very time consuming and the knowledge not always available. If a mathematical representation of the process exists then digraphs can be developed systematically from differential and algebraic

equations (Maurya et al., 2003a,b). However, complete descriptions are rarely available for complex chemical processes and large plants.

An incorporation of the data-driven approach in a causal map consisting of digraphs was presented by Chiang and Braatz (2003). They derive causal maps from expert knowledge and from the measurements affected by a fault. Broken dependencies along the propagation paths are identified using data-driven, entropy related measures. The direction of propagation is not automatically detected but derived from expert knowledge.

The contribution of the work reported in this paper explores a measure for the directionality of variation propagation in historical process data. The result can then serve as a basis for qualitative model generation in form of digraphs.

The next section explains the concept of transfer entropy and its use for fault propagation. To illustrate the concept of transfer entropy it is then applied to an industrial case study in Section 3. Results of directionality and consistency detected by transfer entropy are discussed.

2. DIRECTIONALITY ANALYSIS

Entropy related measures regard the measurement of a physical variable as the outcome of a random process. Information contained in the outcome is expressed by its probability density function (PDF). Assuming that the process is ergodic over a defined timeframe the PDF can be estimated by the relative frequency. This is achieved by dividing the amplitude axis into discrete amplitude bins and assigning every sample data value to a bin accordingly. The number of data points per bin accumulates to the relative frequency.

Entropy measures the information, also referred to as uncertainty or randomness, of a process by logarithmical summing over all bins. When a fault is present the PDF and consequentially the entropy can change significantly (Chiang and Braatz, 2003). Thus, entropy can be used for detecting faults as it is insensitive to outliers or temporary biases.

2.1 Transfer Entropy

A focus on PDF destroys the time trend of the sequence. To show the significance of this problem, two signals are considered: the first is a random signal whose amplitude is uniformly distributed between -1 and +1, the second a triangular wave oscillating regularly between -1 and +1. The PDFs of these two signals are identical yet they are obviously

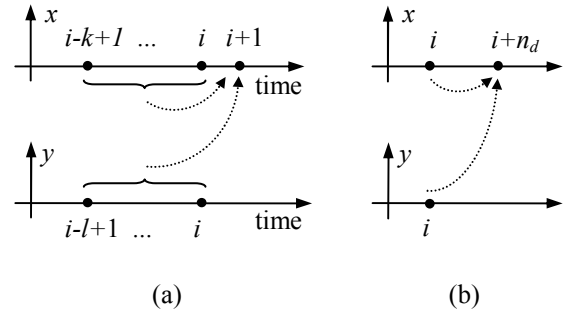


Fig. 1. Time samples involved in the computation of the transfer entropy, general case (a) and special case for $k=l=1$ with time delay n_d (b).

not the same signal. Thus, PDF and entropy cannot capture time dependency.

Transfer entropy has been recently proposed by Schreiber (2000) to measure dependency in time and thus which of two variables causes the other. As an example, the dependency between heart and breath rate was investigated, finding that the heart rate influences the breath rate rather than vice versa. Transfer entropy relates k previous samples of variable X and l previous samples of variable Y to predict the next value of X and thus incorporates the time dependency. This relationship is illustrated in Figure 1(a). Transfer entropy is then calculated by summing the joint and conditional PDF of two time sequences logarithmically:

$$T_{Y \rightarrow X} = \sum_{x_{i+1}} \sum_{\mathbf{x}_i^k} \sum_{\mathbf{y}_i^l} p(x_{i+1}, \mathbf{x}_i^k, \mathbf{y}_i^l) \log \frac{p(x_{i+1} | \mathbf{x}_i^k, \mathbf{y}_i^l)}{p(x_{i+1} | \mathbf{x}_i^k)} \quad (1)$$

where the sums are over all amplitude bins. The joint PDF $p(x_{i+1}, \mathbf{x}_i^k, \mathbf{y}_i^l)$ is the probability that the combination of x_{i+1} , \mathbf{x}_i^k and \mathbf{y}_i^l have particular values. The conditional PDF $p(x_{i+1} | \mathbf{x}_i^k, \mathbf{y}_i^l)$ is the probability that x_{i+1} has a particular value when the value of previous samples \mathbf{x}_i^k and \mathbf{y}_i^l are known.

The index $Y \rightarrow X$ indicates that the influence of sequence Y on sequence X is measured. The reverse dependency $X \rightarrow Y$ is calculated by exchanging x and y of the joint and conditional PDFs. To specify whether X influences Y more than Y influences X a measure for the directionality is introduced:

$$t_{X \rightarrow Y} = (T_{X \rightarrow Y} - T_{Y \rightarrow X}) / \min\{T_{X \rightarrow Y}, T_{Y \rightarrow X}\} \quad (2)$$

If $t_{X \rightarrow Y}$ is greater than zero then X influences Y . Scaling the difference by the minimum of both values provides a standardized index.

2.2 Time Delays

Transfer entropy uses the presence of time delays to identify which variable influences the other. In continuous processes time delays occur as the product streams along the propagation path. Thus, disturbances and variations can be observed with a time delay at two sequential measuring points. In most cases, the time delay of the stream is unknown and varies with throughput over time. This issue has to be considered when adjusting the parameters for the transfer entropy.

The distance n_d between two consecutive samples, as shown in Figure 1(b) for $k=l=1$, is an important parameter which captures the time delay t_d . The transfer entropy and also $t_{X \rightarrow Y}$ become a function of n_d . Eq. 1 reveals that no symmetry with respect to n_d holds but it can be noted that $T_{X \rightarrow Y}(n_d) = T_{Y \rightarrow X}(-n_d)$.

Ideally, the parameter n_d should equal the actual time delay t_d . Since t_d is usually not known, a weighted sum of the directionality measure t_{XY} for a number of N of values of n_d can be introduced:

$$t'_{X \rightarrow Y}(N) = \sum_{n_d=1}^N \frac{1}{n_d} t_{X \rightarrow Y}(n_d). \quad (3)$$

The number of samples N has to be sufficiently large to ensure that the time delay t_d is captured. The directionality for $n_d = t_d$ should be much larger than the average value for the directionality the sum in Eq. 3 therefore will be dominated by the $n_d = t_d$ term and will capture the dependency. The factor $1/n_d$ originates from the assumption that small time delays are generally more relevant than large time delays.

This intuitive assumption is based on the principle of Markov processes.

2.3 Challenges

A drawback of transfer entropy is that it depends strongly on the bin size of the PDF. A large number of time samples are required to find a suitable representation of the PDF with a sufficient large number S of amplitude bins. Additionally, the number of bins increases exponentially with larger k and l as the total number of the combination of bins is $S_{tot} = S^{k+1} \cdot S^l$.

The obstacle of poor resolution or too large bin size of the PDF can be overcome by using more sophisticated estimation than binning. Fewer time samples are needed to estimate the PDF when using Kernel functions (Scott, 1979) which are centred around the sample points and thus accumulate to the PDF. Alternatively, a uniform or linear distribution can be assumed and approximated by integrating over each bin (Hukkanen and Braatz, 2003). As this research is still ongoing, Kernel functions will be implemented in a next step.

3. THE INDUSTRIAL PROCESS

The process schematic of the industrial case study is shown in Figure 2. The process is the solvent recycle path in a gas purification system. Measurements of the temperature are taken at seven points along the recycle stream. The level of the liquid at the bottom of both distillation columns and the differential pressure within are also captured.

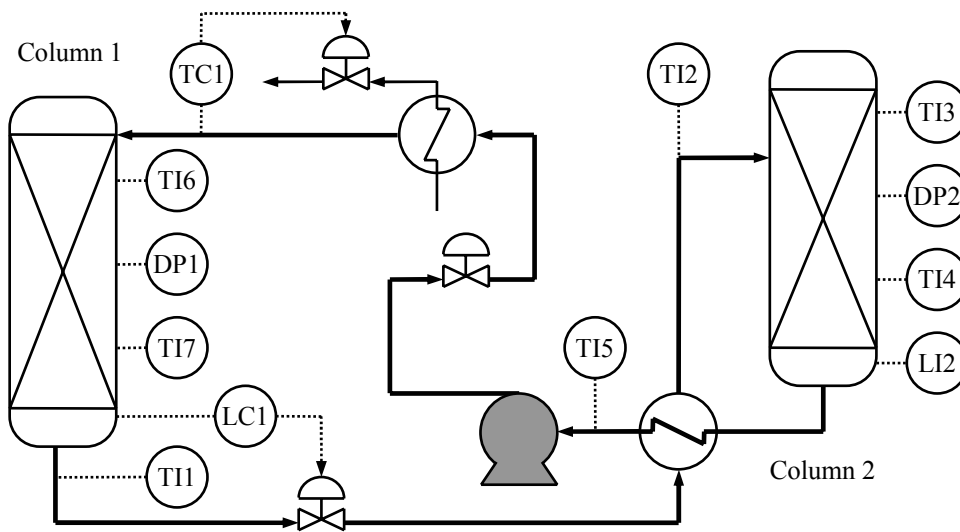


Fig. 2. Process schematic of the industrial case study with process variable indicators and controllers for temperature (T), level (L) and differential pressure (DP).

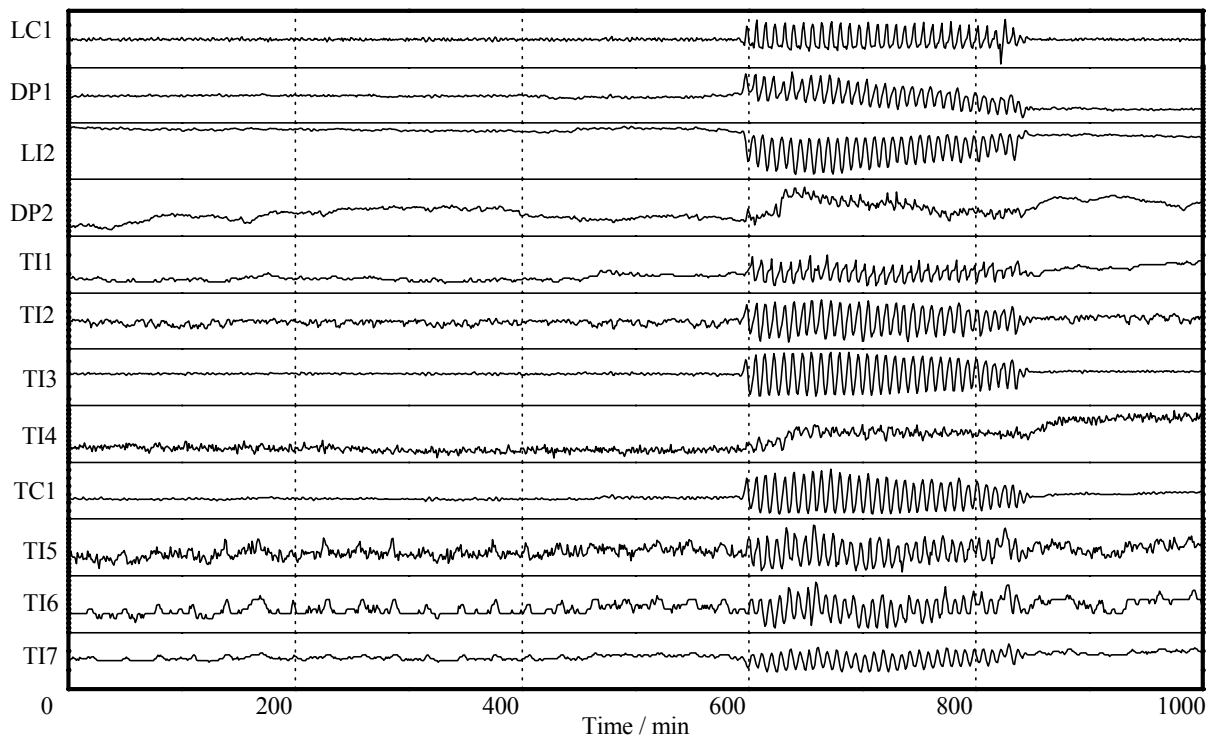


Fig. 3. Time trends of process variables.

The purpose of applying transfer entropy to case study is to provide a measure for the direction of propagation and to identify directionalities that change during the presence of a disturbance.

3.1 The Data Set

The full data set consists of 1000 samples taken at the twelve measurement points at a sample rate of once per minute. Figure 3 shows the mean centred and normalized time trend of the process variables. From visual inspection an oscillatory disturbance can be identified from sample 600 to 840. The identical

period of oscillation and the coinciding start and ending of the disturbance for all measurements suggest a plant-wide disturbance that affects more or less all process variables.

3.2 Direction of Propagation

Five sequential temperature measurements are investigated for applying the directionality measure: indicator TI5 next to the recycle heat exchanger, the process variable of controller TC1, two indicators at distillation column 1 (TI6 and TI7), and the measurement at the exit of column 1, TI1.

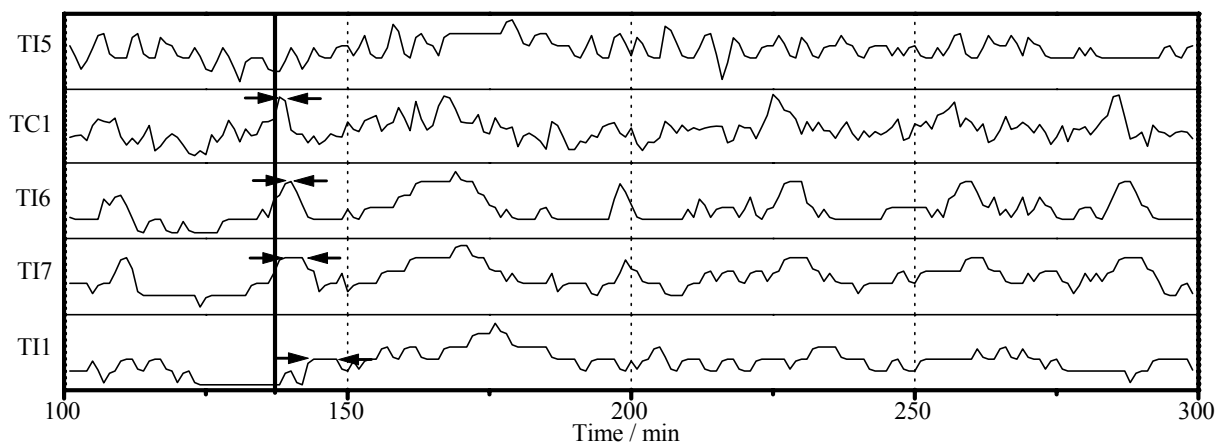


Fig. 4. Close up of temperature measurements.

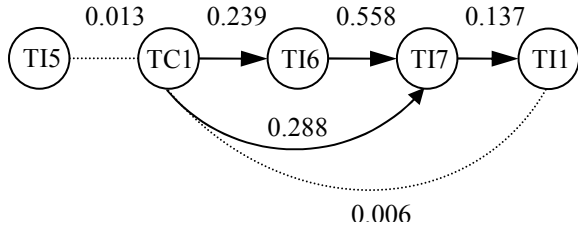


Fig. 5. Directionality $t'_{x \rightarrow y}$ for sequential temperature measurements of the industrial process ($k=2, l=1$).

A close-up of the time trends of these measurements is shown in Figure 4. Distinct features can be observed which occur with a time delay in all variables. A large peak just before 200 minutes provides an example. It occurs first in TC1 and moves to TI6, TI7 and TII. A requirement for the new transfer entropy method is that it should match this visual observation.

Transfer entropy was able to measure the directionality of sequential variables. The variables and the direction represented by arcs are shown in Figure 5. The values assigned with the arcs are values for the directionality measure $t'_{x \rightarrow y}$ ($K=5$). Large values indicate a strong directionality.

As expected, the two temperature indicators at column 1 are strongly dependent ($t'_{TI6 \rightarrow TI7} = 0.558$) and TI6 influences TI7. As a result, TI7 is approximately influenced by TC1 as much as TI6. However, the directionality between TC1 and TII is almost zero ($t'_{TC1 \rightarrow TII} = 0.006$), while TI7 influences TII distinctively ($t'_{TI7 \rightarrow TII} = 0.137$). The directionality measure proves to give a representation of the actual dependencies in the process because it follows the direction of material flow. In particular, Figure 5 shows an influence of TC1 on TI7 as much as on TI6. This means that the structure is correctly represented even if the measurement of TI6 is not available. It can be concluded correctly that TC1 influences all temperature measurements in the column.

3.3 Consistency of Propagation

Consistency is a second issue to be solved by transfer entropy since the direction of signature propagation might change when a disturbance is present. Transfer entropy showed that during normal operation, differential pressure DP1 influences level LC1 while during the disturbance LC1 influences DP1. Physical experiments suggest that the disturbance is caused by foaming in column 1. If that is so then a possible explanation for the observed change in directionality is as follows.

During normal operation (up to 600 min in Figure 3), any pressure increase also increases the liquid hold-up in the column and decreases the flow rate of the liquid. Thus, the level of liquid fluctuates. It is known that foaming affects the level measurement on this column. This in turn will cause LC1 to move the output and this will alter both the liquid flow and liquid hold-up because of the resulting movement of the control valve. The changes in hold-up and flow rates will in turn cause the differential pressure to change.

Figure 6 shows transfer entropy $T_{LC1 \rightarrow DP1}$ and $T_{DP1 \rightarrow LC1}$ over a number of values for the time delay n_d . The directionality measure t' is the weighted sum over all time delays such that during normal operation $t'_{DP1 \rightarrow LC1} = 0.805$ and during the disturbance $t'_{LC1 \rightarrow DP1} = 0.850$. In the upper plot, representing the undisturbed period, DP1 influences LC1. In the lower plot, representing the period when the disturbance is present, LC1 influences DP1. This result suggests that transfer entropy can capture consistency in fault propagation by finding directionality changes when a fault is present.

4. CONCLUSIONS

The concept of transfer entropy (Schreiber, 2000) has been proposed for finding the direction of flow during normal operation and to identify changes when a fault is present.

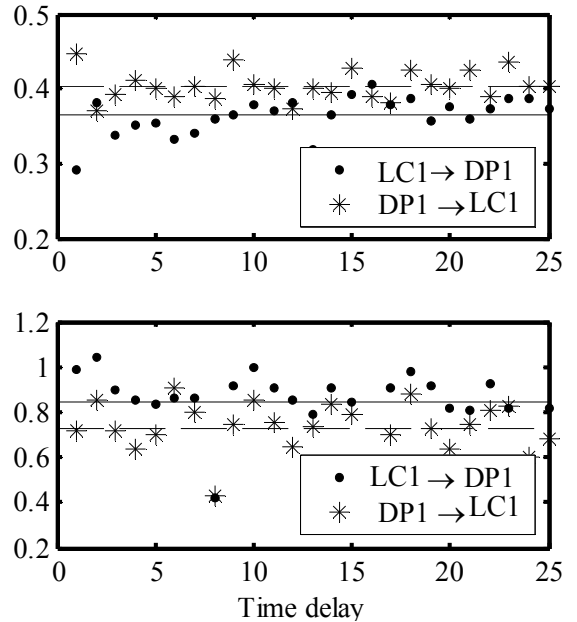


Fig. 6. Transfer entropy of level (LC1) and differential pressure (DP1) during normal operation (upper plot) and when a disturbance is present (lower plot).

The measure gives a basis for creation of digraph models using process data rather than expert knowledge. In an industrial case study, the direction of variables was identified from the process data and an inverse direction was detected during a period of disturbance. The directionality measure is showing exciting results even though the unknown time lag can only be approximated and not exactly captured. It appears to be robust within bounds, however, the next step is to optimize the parameters for the number of time delays N and the number of samples of joint and conditional probability.

ACKNOWLEDGEMENTS

The principal author gratefully acknowledges the financial support of the University College London Graduate School. The UCL authors appreciate the support of BP Chemicals, Hull.

REFERENCES

Chiang, L.H. and Braatz, R.D. (2003). Process monitoring using causal map and multivariate statistics: fault

- detection and identification. *Chemometrics and Intelligent Laboratory Systems*, **65**, 159-178.
- Hukkanen, E. J. and Braatz, R.D. (2003). Measurement of particle size distribution in suspension polymerization using in situ laser backscattering. *Sensors and Actuators B: Chemical*, **96**, 451-459.
- Maurya, M.R., Rengaswamy, R. and Venkatasubramanian, V., 2003. A systematic framework for the development and analysis of signed digraphs for chemical processes. 1. Algorithms and analysis. *Industrial and Engineering Chemical Research*, **42**, 4789-4810.
- Maurya, M.R., Rengaswamy, R. and Venkatasubramanian, V., 2003. A systematic framework for the development and analysis of signed digraphs for chemical processes. 2. Control loops and flowsheet analysis. *Industrial and Engineering Chemical Research*, **42**, 4811-4827.
- Nam, D.S., Han, C., Jeong, C.W. and Yoon, E.S. (1996). Automatic construction of extended symptom-fault associations from the signed digraph. *Computers and Chemical Engineering*, **20**, S605-S610.
- Qin, S.J. (1998). Control performance monitoring – a review and assessment. *Computers and Chemical Engineering*, **23**, 173-186.
- Schreiber, T. (2000). Measuring information transfer. *Physical Review Letters*, **85**, 461-464.
- Scott, D.W. (1979). On optimal and data-based histograms. *Biometrika*, **66**, 605-610.