Finding the Source of Nonlinearity in a Process With Plant-Wide Oscillation

Nina F. Thornhill

***Abstract*—A plant-wide oscillation in a chemical process often has an impact on product quality and running costs and there is, thus, a motivation for automated diagnosis of the source of such a disturbance. This brief describes a method of analyzing data from routine operation to locate the root cause oscillation in a dynamic system of interacting control loops and to distinguish it from propagated secondary oscillations. The novel concept is the application of a nonlinearity index that is strongest at the source. The index is large for the nonsinusoidal oscillating time trends that are typical of the output of a control loop with a limit cycle caused by nonlinearity. It is sensitive to limit cycles caused both by equipment and by process nonlinearity. The performance of the index 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 such as a refinery or petrochemical plant. Issues arising from artifacts in the nonlinearity test when applied to strongly cyclic data have been addressed to provide a robust, reliable and practical method. The technique is demonstrated with three industrial case studies.**

***Index Terms*—Fault diagnosis, harmonics, limit cycles, nonlinearities, spectral analysis, surrogate data, time series.**

I. Introduction

IT IS important to diagnose and cure oscillations in a controlled process because a system running steadily without oscillation is more profitable and safer [1]. A feedback control loop containing a nonlinearity such as a sticking valve often exhibits self-generated and self-sustained limit cycle oscillation [2]–[4], and many surveys have shown that these oscillations are a significant industrial problem [5], [6]. The situation is made worse when the oscillation propagates throughout a dynamic system such as a chemical plant where it can become widespread due to physical coupling and recycles. Rapid determination of the source of a system-wide oscillation allows maintenance activity to focus on the root cause [7]. This article presents a method aimed at that objective and presents three industrial case studies in which the method successfully found the root cause.

The time trend of measurements from a limit cycle oscillation is a nonlinear time series, i.e., it cannot be described as the output of a linear system driven by white noise. A nonlinearity test from Kantz and Schreiber [8] has been adapted for the detection of limit cycle oscillations and guidelines for its application to process data have been devised. The underpinning idea in root cause diagnosis is that the nonlinearity is greatest at the source of the problem. By source is meant a measurement associated with the single-input–single-output controller that has been caused to oscillate by a nonlinearity in the loop. Justification of that assumption is given in the next section.

The possibility of a nonlinearity test was outlined in an earlier conference publication [9] and demonstrated in an application at Eastman Chemical Company [10]. The key advance in this brief compared with [9] is the exploitation of the cyclic nature of the measurements to optimize the method. In [10], the focus was on the solution to a particular industrial application and a set of parameters was selected for the algorithm but without a discussion of their optimality. A contribution of this brief is to give fundamental insights into the method and to extend it to additional case studies involving the detection of valve, sensor and process nonlinearities. It gives an in-depth explanation of how the method works when applied to oscillating disturbances and the criteria by which the parameters of the algorithm may be selected. Default parameters are suggested to facilitate routine application of the method to a large-scale plant.

II. Diagnosis of Nonlinearity

A. Nonlinear Time Series Analysis

The waveform in a limit cycle is periodic but nonsinusoidal and therefore has harmonics. A distinctive characteristic of a nonlinear time series is the presence of phase coupling which creates coherence between the frequency bands occupied by the harmonics such that the phases are nonrandom and form a regular pattern. Nonlinearity may, thus, be inferred from the presence of harmonics and phase coupling.

Methods for nonlinearity detection in the time series include the techniques using surrogate data [8], [11] which have been used in applications ranging from analysis of EEG recordings of people with epilepsy [12] to the analysis of X-rays emitted from a suspected astrophysical black hole [13]. Surrogate data are times series having the same power spectrum as the time series under test but with the phase coupling removed by randomization of phases. A key property of the test time series is compared to that of its surrogates and nonlinearity is diagnosed if the property is significantly different in the test time series.

Another method of nonlinearity detection uses higher order spectra because these are sensitive to certain types of phase coupling. For instance, the bispectrum [14]–[16] responds to quadratic phase coupling in a signal such as $x(t)$ below in which the phase of the frequency component at f_1+f_2 is $\phi_1+\phi_2$, but there is no bispectral response if ϕ_3 is a random phase

$$x(t) = a_1\cos(2\pi f_1 t+\phi_1) + a_2\cos(2\pi f_2 t+\phi_2) + a_3\cos(2\pi(f_1+f_2)t+\phi_3).$$

The bispectrum and the related bicoherence have been used to detect the presence of nonlinearity in process data [17], [18].

Manuscript received February 18, 2004. Manuscript received in final form June 14, 2004. Recommended by Associate Editor A. T. Vemuri. This work was supported in part by the Royal Academy of Engineering under the Foresight Award and in part by the NSERC-Matrikon-ASRA Industrial Research Chair in Process Control at the University of Alberta.

The author is with the Imperial College/UCL Center for Process Systems Engineering, Department of Electronic and Electrical Engineering, University College London, London WC1E 7JE, U.K. (e-mail: n.thornhill@ee.ucl.ac.uk).

Digital Object Identifier 10.1109/TCST.2004.839570

1063-6536/$20.00 © 2005 IEEE

A potential disadvantage of the bispectrum for detection of nonlinear limit cycle oscillations is that limit cycles may have symmetrical waveforms (e.g., a square wave or triangular wave) and the bispectrum of a symmetrical waveform is zero. Zang and Howell [19] have investigated the types of limit cycles that are amendable to bispectrum analysis.

The presence of harmonics in a time series has also been used successfully for diagnosis of SISO control loop faults [7], [20], [21]. Finding harmonics requires signal processing to isolate the spectral frequencies of interest and inspection to confirm that the frequencies are integer multiples of a fundamental. The inspection is often undertaken by visual examination of the spectra and is therefore unsuitable for a large-scale implementation involving several hundred or even a thousand or more plant measurements. Moreover, it is possible that components at the harmonic frequencies are not phase coupled in which case the harmonic signature will be a misleading indicator of nonlinearity.

B. Propagation of a Nonlinear Limit Cycle

Repair of a faulty control loop requires that the engineer knows which control loop should be maintained. In the case of a plant-wide oscillation, it can be a very difficult problem to know which loop to work on because the disturbance from a control loop in a limit cycle typically propagates plant-wide to cause numerous secondary oscillation in other control loops. An automated means is therefore needed to determine which among all the oscillating control loops is the root cause and which are secondary oscillations. Successful studies have used the presence of prominent harmonics to distinguish the source of a limit cycle oscillation from the secondary oscillations in a distillation column in a refinery [22] and in a pulp and paper mill [23]. The reason why secondary oscillations have lower nonlinearity is that as the signal propagates away from its source it passes through physical processes which give linear filtering and which generally add noise. (The filter may be assumed linear if the system is oscillating around a fixed operating point). Such a filter destroys the phase coherence of the time trends and often reduces the magnitudes of the harmonics. Thus, nonlinearity reduces as the disturbance propagates away from the source and the time trend with the highest nonlinearity is the best candidate for the root cause. The nonlinearity statistic to be discussed in Section III can be used for such root-cause diagnosis.

C. Surrogate Data Analysis

A time series with phase coupling is more structured and more predictable than a similar time series known as a *surrogate* having the same power spectrum but with random phases [11]. The spread of values of some statistical property of a group of surrogate data trends provides a reference distribution against which the properties of the test time series can be evaluated.

The techniques of surrogate data analysis have been widely applied for detection of nonlinearity in time series [12], [13]. In the process area, Aldrich and Barkhuizen [24] detected nonlinearity in process data by comparing a singular spectrum analysis of the test data with those from linear surrogate data. Barnard *et al.* [25] showed that identification of systems is possible by using surrogate methods to classify the data, as well as to validate models derived from these data.

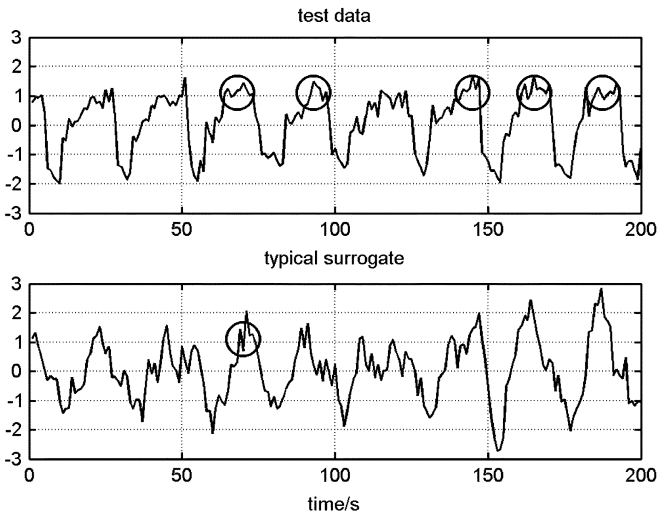


Fig. 1. Test data and typical surrogate. The time trends are mean centered and scaled to unit standard deviation.

Issues have been identified with the use of surrogate data with cyclic time series [26], [27]. The surrogate is derived by taking the discrete Fourier transform (DFT) of the test data, randomization of the arguments followed by an inverse DFT. Nonlinearity testing based on strongly cyclic data can give rise to false detection of nonlinearity because when the time trend is strongly cyclic then artifacts in the DFT due to end-matching effects influence the surrogates. A demonstration of the consequences for strongly cyclic data are demonstrated in this article although in practical applications the effect was found to have a minimal impact, as will be discussed later.

III. Method

A. Overview

The basis of the test is a comparison of the predictability of the time trend compared to that of its surrogates. Fig. 1 illustrates the concept. The top panel is an oscillatory time trend of a steam flow measurement from a refinery. It has a clearly defined pattern and a good prediction of where the trend will go after reaching a given position, for example at one of ringed peaks, can be achieved by finding similar peaks in the time trend and observing where the trend went next on those occasions.

The lower panel shows a surrogate of the time trend. By contrast to the original time trend the surrogate lacks structure even though it has the same power spectrum. The removal of phase coherence has upset the regular pattern of peaks. For instance, it is hard to anticipate where the trajectory will go next after emerging from the region highlighted with a circle because there are no other similar peaks.

Predictability of the time trend relative to the surrogate gives the basis of a nonlinearity measure. Prediction errors for the surrogates define a reference probability distribution under the null hypothesis. A nonlinear time series is more predictable than its surrogates and a prediction error for the test time series smaller than the mean of the reference distribution by more than three standard deviations suggests the time trend is nonlinear.

B. Construction of the Data Matrix

Nonlinear prediction of time series was described by Sugihara and May [28] to distinguish determinism from random noise, and the field of nonlinear time series analysis and prediction has been reviewed by Schreiber [29]. Rhodes and Morari [30] gave an early process application of nonlinear prediction where the emphasis was on modeling of nonlinear systems when noise corrupts a deterministic signal.

Nonlinear prediction uses a data matrix called an embedding having E columns each of which is a copy of the original data set delayed by one sampling interval. For instance, a data matrix with $E = 3$ is

$$\mathbf{Y} = \begin{pmatrix} y(1) & y(2) & y(3) \\ y(2) & y(3) & y(4) \\ y(3) & y(4) & y(5) \\ \cdots & \cdots & \cdots \\ y(\ell-2) & y(\ell-1) & y(\ell) \end{pmatrix}.$$

Rows of the matrix $\mathbf{Y}$ represent time trajectories that are segments of the original trend. Since the original data formed a continuous time trend the trajectories in adjacent rows are similar. They are called near-in-time neighbors. Also, if the time trend is oscillatory then the trajectories in later rows of $\mathbf{Y}$ will be similar to the earlier rows after one or more complete cycles of oscillation. For instance, if the period of oscillation is 50 samples per cycle then $\|\underline{\mathbf{y}}_{51} - \underline{\mathbf{y}}_1\|$ will be small, where $\underline{\mathbf{y}}_{51}$ is the 51st row vector of $\mathbf{Y}$ and $\underline{\mathbf{y}}_1$ is the first. Those rows are called *near neighbors*.

C. Calculation of Prediction Error

Predictions are generated from near neighbors. Near-in-time neighbors are excluded so that the neighbors are only selected from other cycles in the oscillation. When k nearest neighbors have been identified then those near neighbors are used to make an H step-ahead prediction. For instance, if row vector $\underline{\mathbf{y}}_1$ were identified as a near neighbor of $\underline{\mathbf{y}}_{51}$ and if H were 3 then $y(4)$ would give a prediction of $y(54)$. A sequence of prediction errors can, thus, be created by subtracting the average of the predictions of the k nearest neighbors from the observed value. The overall prediction error is the rms value of the prediction error sequence.

The analysis is noncausal and any element in the time series may be predicted from both earlier and later values. Fig. 2 illustrates the principle using a time series from the SE Asia refinery case study where the embedding dimension E is 16 and the prediction is made 16 steps ahead. The upper panel shows the 100th row of the data matrix $\mathbf{Y}$ which is a full cycle starting at sample 100, marked with a heavy line. Rows of $\mathbf{Y}$ that are nearest neighbors of that cycle begin at samples 67, 133, 166, 199, and 232 and are also shown as a heavy lines in the lower panel. The average of the points marked ◯ in the lower panel are then used as a prediction for the value marked $\mathbf{x}$.

D. Data Preprocessing

Detection of plant-wide oscillation is now a solved problem and is starting to be offered by vendors [31], [32]. The periodic nature of the detected oscillation may be exploited in order to give robust default settings for the parameters. A summary list is presented here and the detailed reasoning behind the recommendations will be presented in Section IV. With the data preprocessing steps indicated here the default parameters can be used for any oscillating time trend.

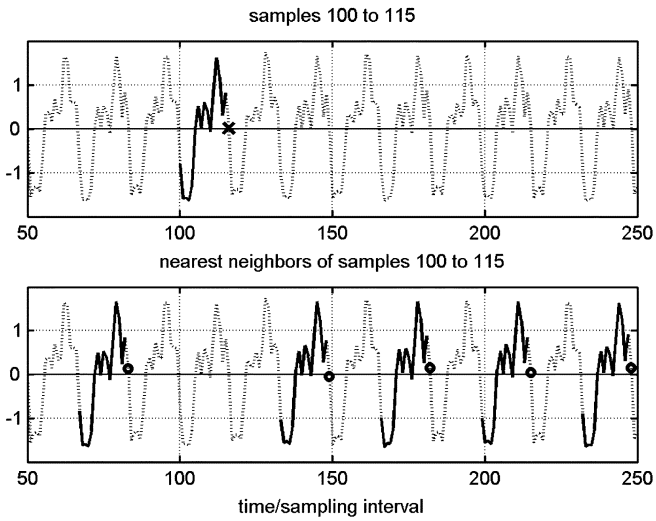


Fig. 2. Illustration of the nearest neighbor concept. The highlighted cycles in the lower plot are the five nearest neighbors of the cycle in the upper plot. The average of the points marked o gives a prediction for the point marked $\mathbf{x}$.

1) The period of the plant-wide oscillation is determined.
2) The number of samples per cycle S is adjusted to be no more than 25. The time trends are subsampled if necessary.
3) The number of cycles of oscillation in the data set should be at least 12 for a reliable nonlinearity estimate.
4) The selected data are end-matched to find a subset of the data containing an integer number of full cycles. The algorithm and other issues associated with end-matching are explained in detail in Section IV-F.
5) The end-matched data are mean centered and scaled to unit standard deviation. The sequence $y(1)\cdots y(\ell)$ denotes end-matched and preprocessed data in the following sections.

E. Surrogate Data

Surrogate data are derived from the preprocessed and end-matched time trend. Surrogate data have the same power spectrum as the time trend under test. The magnitudes of the DFT are the same in both cases but the arguments of the DFT of the surrogate data set are randomized. Thus, if the DFT in frequency channel i is

$$|Y(j\omega_i)|e^{j\angle Y(j\omega_i)}$$

then the DFT of the surrogate is

$$|Y(j\omega_i)|e^{j(\angle Y(j\omega_i)+\phi_i)}$$

where ϕ_i is a phase selected from a uniform random distribution in the range $-\pi < \phi_i \leq +\pi$. The aliased frequency channels above the Nyquist sampling frequency have the opposite phase added. If the number of samples ℓ is even and if the frequency

channels are indexed as $i = 1$ to ℓ the Nyquist frequency is in channel $\ell/2 + 1$ and the alias of the ith frequency channel is channel $\ell - i + 2$. Then

$$\phi_1 = 0, \quad \phi_{\ell/2+1} = 0, \quad \text{and} \quad \phi_{\ell-i+2} = -\phi_i (i = 2 \text{ to } \ell/2),$$

If ℓ is odd

$$\phi_1 = 0 \quad \text{and} \quad \phi_{\ell-i+2} = -\phi_i (i = 2 \text{ to ceil}(\ell/2))$$

where ceil$(\ell/2)$ is the rounded-up integer value of $\ell/2$.

Finally, the surrogate data set is created from the inverse Fourier transform of the phase randomized DFT.

F. Nonlinearity Test

The nonlinearity test requires the determination of mean square prediction errors of M surrogates. The statistical distribution of those errors gives a reference distribution. If the test data prediction error lies on the lower tail of the reference distribution then the test signal is more predictable and nonlinearity is diagnosed using the following statistic based on a three-sigma test

$$N = \frac{\bar{\Gamma}_{\text{surr}} - \Gamma_{\text{test}}}{3\sigma_{\Gamma_{\text{surr}}}}$$

where Γ_{test} is the mean square error of the test data, $\bar{\Gamma}_{\text{surr}}$ is the mean of the reference distribution and $\sigma_{\Gamma_{\text{surr}}}$ its standard deviation. If $N > 1$ then nonlinearity is inferred in the time series. Larger values of N are interpreted as meaning the time series has more nonlinearity, those with $N < 1$ are taken to be linear.

It is possible for the test to give small negative values of N. Negative values in the range $-1 \leq N < 0$ are not statistically significant and arise from the stochastic nature of the test. Results giving $N < -1$ do not arise at all because the surrogate sequences which have no phase coherence are always less predictable than a nonlinear time series with phase coherence.

G. Algorithm Summary

Step 1) Form the embedded matrix from a preprocessed and end matched subset of the test data $y(1) \cdots y(\ell)$

$$\mathbf{Y} = \begin{pmatrix} y(1) & y(2) & \cdots & y(E) \\ y(2) & y(3) & \cdots & y(E+1) \\ y(3) & y(4) & \cdots & y(E+2) \\ \cdots & \cdots & \cdots & \cdots \\ y(\ell-E+1) & y(\ell-E+2) & \cdots & y(\ell) \end{pmatrix}.$$

Step 2) For each row $\underline{\mathbf{y}}_i$ of $\mathbf{Y}$ find the indexes j_p $(p = 1 \ldots k)$ of k nearest neighbor rows $\underline{\mathbf{y}}_{j_p}$ having the k smallest values of $\|\underline{\mathbf{y}}_{j_p} - \underline{\mathbf{y}}_i\|$ subject to a near-in-time neighbor exclusion constraint $|j_p - i| > E/2$.

Step 3) Find the sum of squared prediction errors for the test data

$$\Gamma_{\text{test}} = \sum_{i=1}^{\ell-H} \left(y(i+H) - \frac{1}{k} \sum_{p=1}^{k} y(j_p + H) \right)^2.$$

Step 4) Create M surrogate prediction errors Γ_{surr} by applying steps 1 through 3 to M surrogate data sets.

TABLE I
SUGGESTED DEFAULT VALUES FOR PARAMETERS

description	value
samples per cycle, S	$7 \leq S \leq 25$
number of columns in embedded matrix, E	$E = floor(S)$
prediction horizon, H	$H = E$
number of cycles of oscillation, C	$C \geq 12$
number of near neighbors, k	$k = 8$
number of surrogates, M	$M = 50$

Step 5) Calculate the nonlinearity from

$$N = \frac{\bar{\Gamma}_{\text{surr}} - \Gamma_{\text{test}}}{3\sigma_{\Gamma_{\text{surr}}}}.$$

IV. DEFAULT PARAMETER VALUES

A. Default Parameter Values

Empirical studies have been carried out to ascertain the sensitivity of the nonlinearity index to the parameters of the algorithm. Reliable results have been achieved using the default values shown in Table I. The *floor* function in the third row indicates that for noninteger values of S then E is set to the rounded-down integer value of S.

The next subsections explore each one of these recommendations showing why they were selected. The time trends from Fig. 3. were used for the evaluation (they are from the industrial case study in Section V-B). Fig. 3 shows mean centered data normalized to unit standard deviation while the spectra are scaled to the same maximum peak height.

The time series of the first three measurements are nonlinear because they are close to the root cause. Their spectra have harmonics and the phase patterns are not random. The last two are far from the root cause and are linear. The data have an oscillation period of 16.7 sampling intervals and the conditions used were varied around default values of $C = 12, E = H = 16, k = 8$ and $M = 50$.

B. Number of Samples Per Cycle

It is practical to limit the number of samples per cycle S. There is a tradeoff between the number of samples needed to properly define the shape of a nonsinusoidal oscillation on the one hand and the speed of the computation on the other. The algorithm requires a distance measure to be ascertained between every pair of rows in the embedded matrix and the time taken for the computation increases as ℓ^2, where $\ell = C \times S$ is the total number of samples in the time trend. Therefore the number of samples per cycle S and the number of cycles C cannot be increased arbitrarily. Data sets of 200 samples $(S = 8.2, C = 24)$ and 418 samples ($S = 16.7$ and $C = 25$) gave successful results in the industrial case studies reported in Section V.

It would be infeasible to operate with fewer than seven samples per cycle because harmonics would not be satisfactorily captured. With $S = 7$, any third harmonic present is sampled at

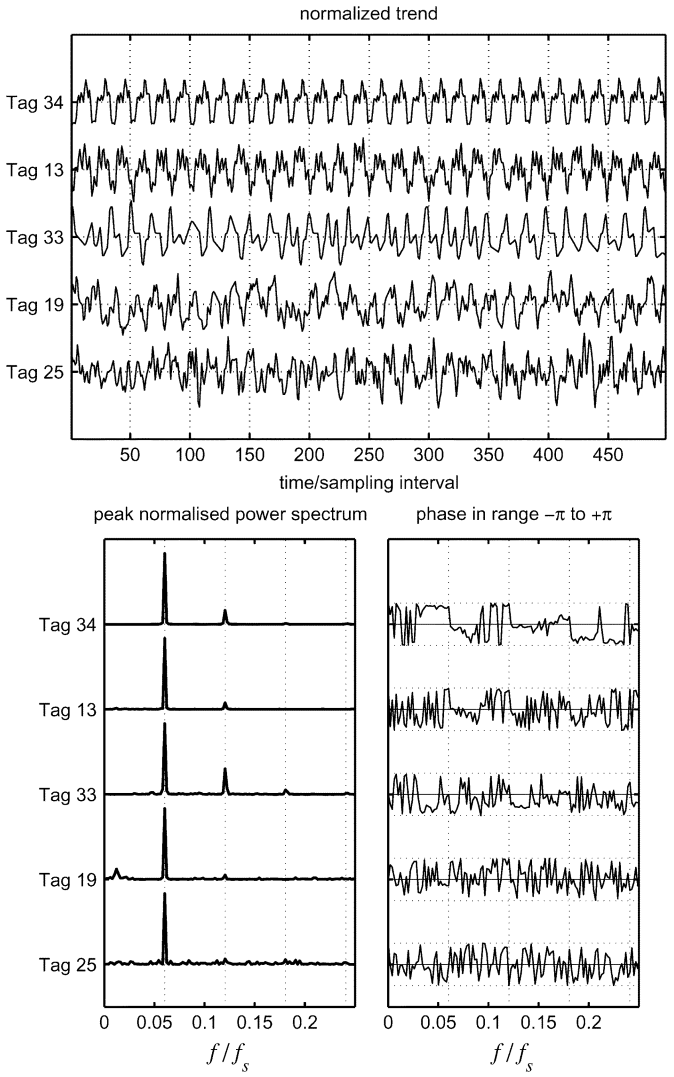


Fig. 3. Time trends and spectra of the data used for detailed evaluation.

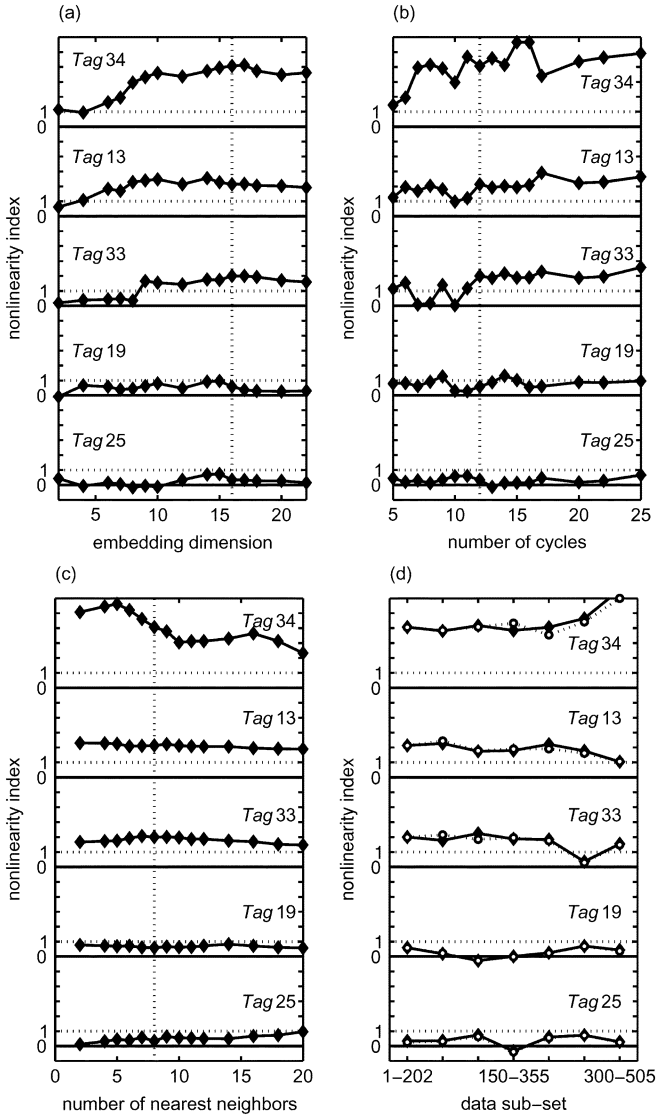


Fig. 4. Effects of parameters of the algorithm on the calculated nonlinearity index. The vertical dashed lines in (a)–(c) show the recommended default values.

2.33 samples per cycle which just meets the Nyquist criterion of two samples per cycle. The reason for focusing the recommendation on the third harmonic is that it is the most prominent harmonic in symmetrical oscillations having square or triangular waveforms.

C. Embedding Dimension and Prediction Horizon

Fig. 4(a) shows the effect of changing E and H. They were kept equal to each other and both were varied together. The threshold of nonlinearity $(N = 1)$ is also shown in the plot (horizontal dashed line) as well as the $E = H = 16$ default for the data set (vertical dashed line). Once E becomes larger than half a cycle of the oscillation, in this case when $E > 8$, the results for the nonlinearity index N become quite steady while for small values of E the index falls toward the $N = 1$ threshold.

An aim of the work presented here is to give reliable default values that are easy to determine. Determination of the period of oscillation S is becoming a standard component of controller performance tools [31], [32]. Therefore the recommendation to set $E = \text{floor}(S)$ is robust because it is in the steady region of Fig. 4(a) and is easy to implement because S is already known.

The poor performance with small values of E arises because of the phenomenon of false near neighbors [33], especially when the time trends have high frequency features or noise. The upper panel in Fig. 5 shows an example of what can happen when the embedding dimension is small, in this case $E = 2$. The rows of the $\mathbf{Y}$ matrix starting at sample 159 comprises just two samples, 159 and 160, shown as small square symbols. Near neighbors are shown in the lower panel, these are two-sample segments of the time trend whose values are similar to samples 159 and 160. However, some of them are false neighbors because they are not from matching parts of the trend. The average of the points marked ◯ are used as a prediction for the value marked $\mathbf{x}$, but some of them such sample 223 which is based on a false neighbor are not accurate.

D. Number of Cycles and Near Neighbors

Fig. 4(b) shows a plot of the number of cycles of oscillation presented for analysis versus the nonlinearity value.

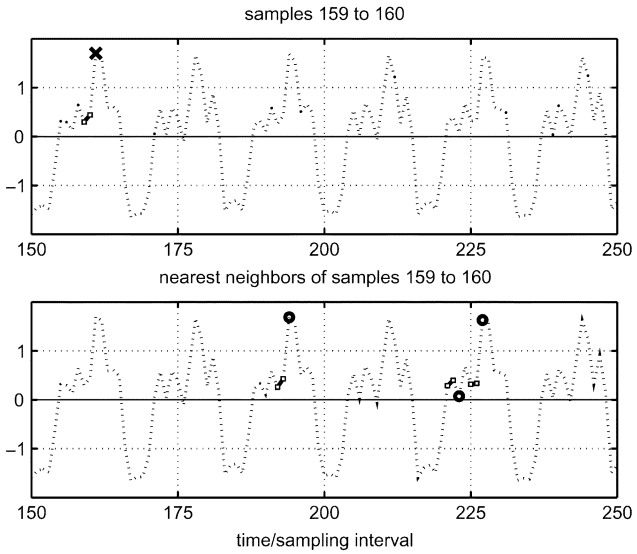


Fig. 5. Illustration of false near neighbors when $E = 2$. The samples at 221 and 222 are similar to those at 159 and 160 but are from a different part of the cycle. A prediction of sample 161 based on sample 223 will be inaccurate.

The value of the nonlinearity statistic fluctuates up and down but when the number of cycles becomes too few the results start to become unreliable. For instance, when more than twelve cycles are used in the analysis then tag 33 is consistently and correctly reported as nonlinear but when fewer than 12 cycles are present its nonlinearity index and that of other tags drops toward the $N = 1$ threshold. On the basis of these examples it seems necessary to use 12 cycles of oscillation or more. The nonlinearity index becomes consistent if this condition applies because the ranking order of the tags is maintained, for instance Fig. 4(b) shows that tag 34 consistently has the highest nonlinearity index if $C \geq 12$.

Fig. 4(c) fixes the number of cycles of oscillation at 12 and varies k, the number of near neighbors used for prediction. The results are quite steady over a wide range of values of k although some of the tags with nonlinearity show a drop toward the $N = 1$ threshold for large k. A recommended value for k can be based on the number of cycles of oscillation. From common sense reasoning, it is sensible to make sure that number of near neighbors is smaller than the number of cycles because each near neighbor is one whole cycle if the $E = S$ recommendation is adopted. Given that one or two cycles may be lost during end-matching a conservative choice is $k = 8$ when the number of cycles is 12. The same conservative reasoning suggests that for other cases k should never be greater than $C - 4$ and in practice it has been found quite satisfactory to just set the value to $k = 8$.

The reason why the nonlinearity tests gives less reliable results for large k when the number of cycles C is fixed at 12 is that the useful near neighbors run out. For instance, if $k = 20$ and if the data set has twelve cycles then any one cycle has eleven near neighbors that are closely matching cycles starting at the same position in the oscillation, like those in Fig. 2. The remaining near neighbors will have to be selected from other rows of the embedded matrix and will not be such good matches.

E. Assessment of Variability of the Index

Fig. 4(d) shows the variability in the nonlinearity index as the data subset is varied. Each time trend had 512 points. The data set was divided into seven overlapping subsets each having 12 cycles of oscillation. The first subset comprised samples 1 to 202, the second was 50 to 252, and so on. Fig. 4(d) shows that different parts of the data trend exhibit varying amounts of nonlinearity. It is noted, however, that the same conclusion about the tag with the highest nonlinearity applies regardless of the data subset. Tag 1 (34) has the highest nonlinearity right across the board, tags 13 and 33 are next highest and 19 and 25 have little nonlinearity.

The white dots in Fig. 4(d) show the effect of varying the number of surrogates. Since it is a statistical test there must be enough surrogates to properly define the reference distribution. The difference between the results from 50 surrogates (black diamonds) and 250 surrogates (white circles) is about ± 0.1 overall and is less than the variability caused by the data subset. It is therefore concluded that 50 surrogates are enough.

F. End-Matching Step

1) End-Matching Criterion: Surrogate data analysis requires a subset of the data such that the starting gradient and value match well to the final gradient and value. Hegger, *et al.* [34] recommend finding a subset of the nonmatched data (denoted by samples x) with n samples starting at x_i and ending at x_{i+n-1} which minimizes the sum of the normalized discontinuities $(d_0 + d_1)$ between the initial and end values and the initial and end gradients, where

$$d_0 = \frac{(x_i - x_{i+n-1})^2}{\sum_{j=i}^{i+n-1}(x_j - \bar{x})^2}$$

$$d_1 = \frac{((x_{i+1} - x_i) - (x_{i+n-1} - x_{i+n-2}))^2}{\sum_{j=i}^{i+n-1}(x_j - \bar{x})^2}$$

with $\bar{x}$ being the mean of the sequence $x_i \cdots x_{i+n-1}$.

The end matching procedure is modified for use with an oscillating signal as recommended in [26] to avoid the artifacts due to spectral leakage described in the next section. End matching of an oscillating time trend as described previously creates a time trend where the last value is the first sample of another cycle. An end matched sequence which contains an exact number of cycles is $x_i \ldots x_{i+n-2}$ derived from the $x_i \ldots x_{i+n-1}$ sequence by omitting the last sample.

2) False-Positive Results With Strongly Cyclic Data: It is known that calculation of reliable surrogates can be problematical for regular and smooth cyclic time series. Unless care is taken with the end matching the test may give false positive results and report nonlinearity for a linear time series.

The reason for the false positive results is the phenomenon of spectral leakage in the DFT caused by the use of a finite length data sequence. Fig. 6 illustrates the effect of spectral leakage. The upper panel shows the DFT of a sine wave having eight samples per cycle when the total data length is an exact multiple of the period, in this case exactly eight cycles. The DFT is zero in all frequency channels expect the one at $f/f_s = 0.125$ corresponding to the frequency of the oscillation. By contrast,

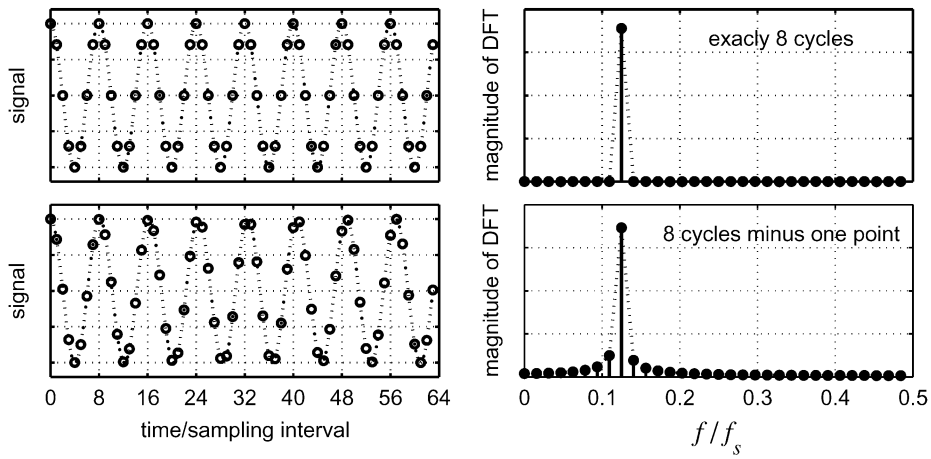


Fig. 6. Illustration of the importance of end matching. For a strongly cyclic time trend the data set should be an exact number of cycles of oscillation otherwise the Fourier transform will give spectral leakage into adjacent frequency channels.

the lower panel shows the DFT when the total data length is a complete number of cycles minus one sample. It has a nonzero magnitude in frequency channels adjacent to the channel containing the main spectral peak. A phase randomized surrogate derived from the DFT in the lower panel therefore contains frequencies that were not present in the original signal and will, thus, be less predictable than the original sine wave giving a false indication of nonlinearity. The true surrogate of a sine wave is a phase shifted sine wave at the same frequency and is equally predictable.

It is therefore necessary to take special precautions when analyzing cyclic time series. Stam *et al.* [26] used the end-matching step that ensures the data length of the time series is an exact multiple of the period of the cycle to avoid false nonlinearity detection. Small and Tse [27] proposed the calculation of special constrained surrogates that pay particular attention to frequencies in the data set having periods longer than the period of the strong cycles. That solution is not applicable to industrial data where the nonlinearity of interest is distortion of the periodic waveform, however.

3) Application to Industrial Data: Experimental laboratory data will give problems of the type outlined previously when the experimental system is driven by a cyclic source such as a laboratory signal generator. This is termed in the literature a "strong cyclic component." Industrial data, however, even when cyclic do not often suffer from the problems described previously because the cyclic behavior is not normally "strong." Although they might be readily detectable the cycles are normally not completely regular and the spectral power in the test signal is already is spread across several frequency channels. For instance, in Fig. 7 showing oscillatory data from a separation column the time trend labeled AC1 has two shorter cycles at around samples 230 to 250. The effects of spectral leakage in the DFT are less severe in this signal because a wider range of frequency channels are already occupied. Therefore the surrogates more accurately represent the real signals than in the case of strong cyclic data.

Table II shows nonlinearity calculations for AC1 and a pure sine wave of the same period of oscillation as the average of AC1. The correct result for the sine wave is $\mathrm{N} < 1$, a result

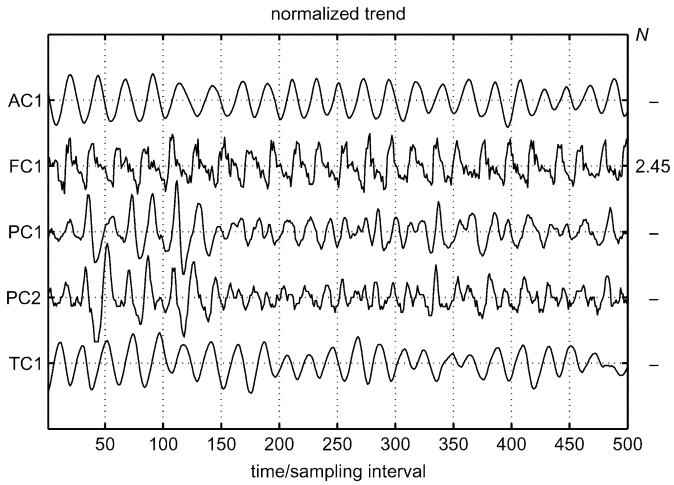


Fig. 7. Time trends from industrial study No 1. Nonlinearity indexes greater than one are shown on the right.

TABLE II
EFFECT OF END MATCHING ON FALSE-POSITIVE RESULTS

end matching	*N* for sine wave	*N* for industrial data AC1
one point extra	**4.72**	0.89
end matched	0.11	0.86
one point fewer	**5.07**	0.83
two points fewer	**5.19**	0.84

which is achieved when the subset of the data is an exact number of cycles of oscillation. If the subset is longer or shorter, even by one sample, there is a false nonlinearity detection because of spectral leakage contaminating the surrogates. By contrast, the industrial data such as AC1 with its less regular cycles are not sensitive to minor variations in the end-match. No false nonlinearity was detected for the industrial data even when mismatched at the ends by one or two samples.

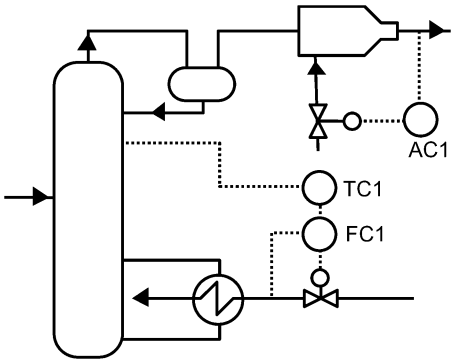


Fig. 8. Process schematic for industrial study No 1. Loops PC1 and PC2 (not shown) are the upstream and downstream pressures.

V. Industrial Case Studies

A. Unit-Wide Oscillation Caused by a Sensor Fault

Fig. 8 shows a separation column courtesy of a BP refinery. The sampling interval was 20 s and the controller errors show the presence of a unit-wide oscillation in FC1, TC1, and AC1 with a period of 21 sampling intervals or 7 min. Measurements from upstream and downstream pressure controllers PC1 and PC2 also show evidence of the same oscillation along with other disturbances and noise.

It is known that the cause of the oscillation was a faulty steam sensor in the steam flow loop FC1. It was an orifice plate flow meter but there was no weep-hole in the plate. Condensate collected on the upstream side until it reached a critical level when the accumulated liquid would periodically clear itself by siphoning through the orifice. The challenge for nonlinearity detection is to identify FC1 as the source of the unit-wide oscillation. The average oscillation period was 21 samples and the settings for the algorithm were therefore chosen as: $E = H = 21$ and $k = 8$. A data set comprising 500 samples and 24 cycles of oscillation was used.

Fig. 7 plots mean centered time trends of controller errors normalized to unit standard deviation. The nonlinearity indexes are presented at the right-hand side. The only nonlinearity index greater than 1 was that of FC1. Therefore the nonlinearity analysis has correctly identified the steam flow control loop causing the unit-wide oscillation.

B. Plant-Wide Oscillation Caused by a Valve Fault

Fig. 9 shows an outline schematic of a hydrogen reformer from a SE Asian refinery and the mean centered normalized controller errors (1-min samples) are presented in Fig. 10. The measurements from this plant have been discussed before [35] where the main disturbance was shown by spectral principal component analysis to be a 16.7-min oscillation in the reformer and pressure swing absorption (PSA) unit.

The exact root cause was not communicated although it has been emphasized that it is a valve fault and pressure cycle swings of the PSA unit were not the cause. The aim of the analysis is to determine which of the oscillating measurements is closest to the root cause. The average oscillation period was 16.7 samples, the settings for the algorithm were therefore

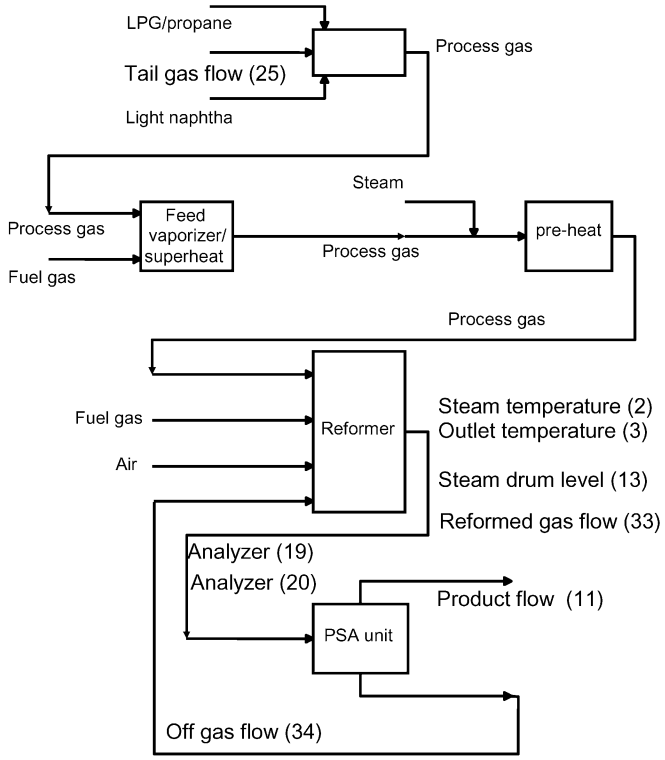


Fig. 9. Process schematic for industrial study No 2.

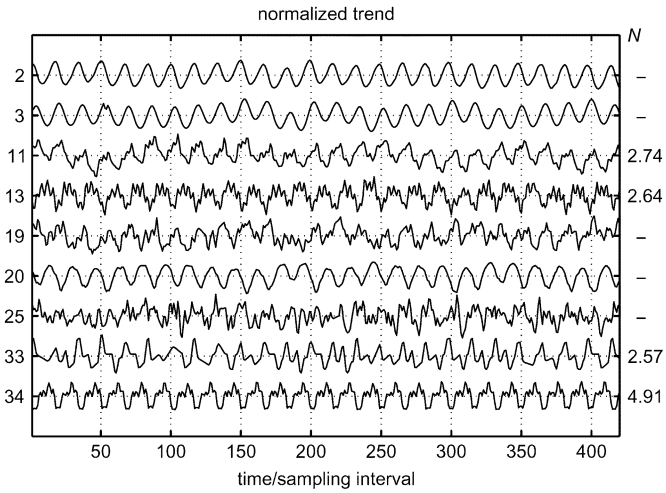


Fig. 10. Time trends from industrial study No 2. Nonlinearity indexes greater than one are shown on the right.

$E = H = 16$ and $k = 8$. A data set comprising 25 cycles of oscillation was used.

The nonlinearity indexes are shown on the right-hand side of Fig. 10 where the largest nonlinearity index is for tag 34. Therefore the flow measured by tag 34 is identified as the physical root cause and the means of propagation of the oscillation is disturbance to the offgas recycle to the reformer.

Tag 25 is upstream yet it was still influenced by the oscillation [35]. The means of propagation from the root of the disturbance in the PSA unit to tag 25 is thought to be that tag 25 is waste gas recycled from another unit to which the oscillation has propagated. That Tag 25 is very far from the root cause is clear because its time trend shows no nonlinearity.

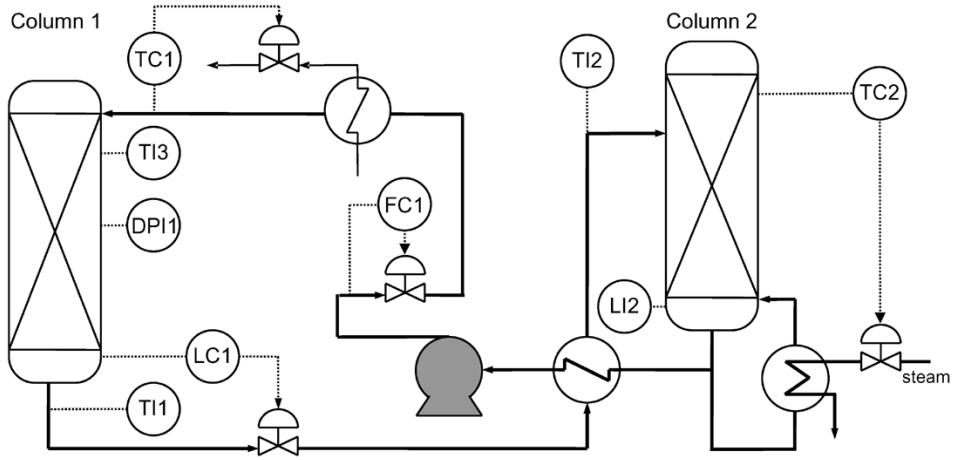


Fig. 11. Process schematic for industrial study No 3.

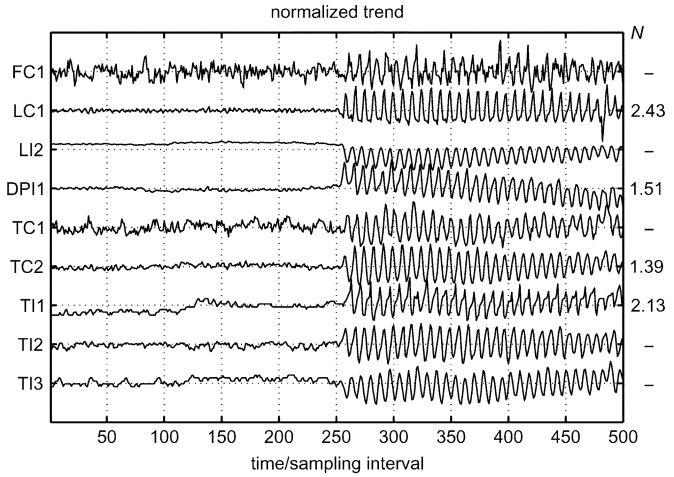


Fig. 12. Time trends from industrial study No 3. Nonlinearity indexes greater than one are shown on the right for samples 261–460 in the oscillatory episode. No nonlinearity was detected before sample 250 in any tag.

The question arises whether multiple sources of nonlinearity can be isolated by the proposed method. It is possible to detect multiple sources if the disturbances have different characteristics such as a different oscillation frequency. A cluster of tags oscillating at the same frequency is assumed to be a plant-wide disturbance with a single root cause. For instance, the tags in Fig. 10 share a 16.7-min oscillation. In [35], a second plant-wide disturbance with a different oscillation period was also reported. It was investigated separately and found to be an external disturbance originating in another unit.

C. Plant-Wide Oscillation Caused by Process Nonlinearity

Fig. 11 shows an outline schematic of the solvent recycle in a gas purification system, courtesy of BP Chemicals. The solvent absorbs a component from a mixed gas process stream in column 1. The absorbed gas is stripped out in column 2 and the regenerated solvent recycles to column 1.

Fig. 12 shows mean centered and normalized time trends from several measurements in the two columns and the challenge was to identify the source of an oscillation that periodically bursts into life, an example of which can be seen starting at sample 250. The prevailing hypothesis was that the oscillation was due to foaming in column 1 because addition of antifoam would stop the oscillation. A successful analysis of this data set should therefore to point to column 1 as the source of oscillation and rule out a competing hypothesis that the oscillation was driven by the steam utility system through the steam valve in TC2.

The nonlinearity results at the right-hand side of Fig. 12 are from the episode of upset operation, samples 261 to 460. The downward sloping linear trend in DPI1 was removed before analysis. The greatest nonlinearity was in Tag LC1 in column 1. The likely mechanism for generation of the oscillation is a periodic buildup and breakdown of foam in column 1 that affects the level sensor LC1, the differential pressure DPI1 and exit temperature TI1. The propagation of the oscillatory disturbance is from its root cause in column 1 to the top of column 2. Tag TC2 (the column 2 temperature controller) participates in the disturbance but its nonlinearity is not high so the steam system is ruled out as the root cause. The reason that TC2 has nonlinearity while TI2 (solvent stream temperature) shows no nonlinearity is because TC2 is affected by variations in the unmeasured flow rate of solvent into column 2 as well as by its temperature. Many other measurements in the recycle path are upset by the oscillating disturbance but no others had any nonlinearity.

This example shows that the nonlinearity index can locate the root causes of a limit cycle caused by process nonlinearity as well as cases where sensors or actuators in control loops cause limit cycling.

VI. Conclusion

Plant-wide oscillations in a system of interacting control loops often originate from a self-sustained limit cycle oscillation in just one control loop. Such a disturbance propagates to other parts of the plant and causes secondary oscillation. This brief has presented a method for locating the root cause of a plant-wide oscillation using a nonlinearity test based on the relative predictability of test data and surrogate data. The performance of the procedure was analyzed as key parameters in the algorithm varied, and default parameters were specified so that the test can be applied to new data sets.

The method was demonstrated using three industrial case studies having a sensor fault, a valve fault and a process nonlinearity caused by hydrodynamic instability. The nonlinearity test located the root cause in all three cases.

ACKNOWLEDGMENT

The author would like to thank A. Meaburn, Z. Rawi, and K. Landells of BP Chemicals, also S. Shah of the University of Alberta and A. Visnubhotla of Matrikon Inc. for providing data and process insights to support the brief.

REFERENCES

[1] J. P. Shunta, *Achieving World Class Manufacturing Through Process Control*. Englewood Cliffs, NJ: Prentice-Hall, 1995.

[2] K. J. Åström, "Assessment of achievable performance of simple feedback loops," in *Int. J. Adapt. Control Signal Process.*, vol. 5, 1991, pp. 3–19.

[3] F. G. Shinskey, "The three faces of control valves," *Control Eng.*, Jul. 2000.

[4] S. F. Graebe, G. C. Goodwin, and G. Elsley, "Control design and implementation in continuous steel casting," *IEEE Control Syst. Mag.*, vol. 15, no. 4, pp. 64–71, Aug. 1995.

[5] D. B. Ender, "Process control performance: Not as good as you think," *Control Eng.*, pp. 180–190, Sep. 1993.

[6] L. Desborough and R. Miller, "Increasing customer value of industrial control performance monitoring—Honeywell's experience," in *Proc. AIChE Symp. Ser.*, vol. 98, 2002, pp. 153–186.

[7] M. A. Paulonis and J. W. Cox, "A practical approach for large-scale controller performance assessment, diagnosis, and improvement," *J. Process Control*, vol. 13, pp. 155–168, 2003.

[8] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis*, Cambridge, U.K.: Cambridge Univ. Press, 1997.

[9] N. F. Thornhill, S. L. Shah, and B. Huang, "Detection of distributed oscillations and root-cause diagnosis," in *Proc. CHEMFAS4*, Korea, Jun. 7-8, 2001, pp. 167–172.

[10] N. F. Thornhill, J. W. Cox, and M. Paulonis, "Diagnosis of plant-wide oscillation through data-driven analysis and process understanding," *Control Eng. Pract.*, vol. 11, pp. 1481–1490, 2003.

[11] J. Theiler, S. Eubank, A. Longtin, B. Galdrikian, B. Farmer, and J. D. Farmer, "Testing for nonlinearity in time-series—The method of surrogate data," *Physica D*, vol. 58, pp. 77–94, 1992.

[12] M. C. Casdagli, L. D. Iasemidis, J. C. Sackellares, S. N. Roper, R. L. Gilmore, and R. S. Savit, "Characterizing nonlinearity in invasive EEG recordings from temporal lobe epilepsy," *Physica D*, vol. 99, pp. 381–399, 1996.

[13] J. Timmer, U. Schwarz, H. U. Voss, I. Wardinski, T. Belloni, G. Hasinger, M. van der Klis, and J. Kurths, "Linear and nonlinear time series analysis of the black hole candidate Cygnus X-1," *Phys. Rev. E*, vol. 61, pp. 1342–1352, 2000.

[14] T. S. Rao and M. M. Gabr, "A test for linearity and stationarity of time series," *J. Time Ser. Anal.*, vol. 1, pp. 145–158, 1980.

[15] M. J. Hinich, "Testing for gaussianity and linearity of a stationary time series," *J. Time Ser. Anal.*, vol. 3, pp. 169–176, 1982.

[16] C. L. Nikias and A. P. Petropulu, *Higher-Order Spectra: A Nonlinear Signal Processing Framework*. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[17] M. A. A. S. Choudhury, S. L. Shah, and N. F. Thornhill, "Diagnosis of poor control loop performance using higher order statistics," *Automatica*, vol. 40, pp. 1719–1728, 2004.

[18] H. E. Emara-Shabaik, J. Bomberger, and D. E. Seborg, "Cumulant/bispectrum model structure identification applied to a ph neutralization process," in *Proc. Inst. Elect. Eng. Control Conf.*, Exeter, U.K., 1996, pp. 1046–1051.

[19] X. Zang and J. Howell, "Discrimination between bad tuning and nonlinearity induced oscillations through bispectral analysis," in *Proc. SICE Annu. Conf.*, Fukui, Japan, Aug. 6–8, 2003.

[20] N. F. Thornhill and T. Hägglund, "Detection and diagnosis of oscillation in control loops," *Control Eng. Pract.*, vol. 5, pp. 1343–1354, 1997.

[21] T. J. Harris, C. T. Seppala, P. J. Jofreit, and B. W. Surgenor, "Plant-wide feedback control performance assessment using an expert system framework," *Control Eng. Pract.*, vol. 9, pp. 1297–1303, 1996.

[22] N. F. Thornhill, M. Oettinger, and P. Fedenczuk, "Performance assessment and diagnosis of refinery control loops," in *Proc. AIChE Symp. Ser.*, vol. 94, 1998, pp. 373–379.

[23] M. Ruel and J. Gerry, "Quebec quandary solved by Fourier transform," *Intech*, pp. 53–55, Aug. 1998.

[24] C. Aldrich and M. Barkhuizen, "Process system identification strategies based on the use of singular spectrum analysis," *Minerals Eng.*, vol. 16, pp. 815–826, 2003.

[25] J. P. Barnard, C. Aldrich, and M. Gerber, "Identification of dynamic process systems with surrogate data methods," *AIChE J.*, vol. 47, pp. 2064–2075, 2001.

[26] C. J. Stam, J. P. M. Pijn, and W. S. Pritchard, "Reliable detection of nonlinearity in experimental time series with strong periodic components," *Physica D*, vol. 112, pp. 361–380, 1998.

[27] M. Small and C. K. Tse, "Applying the method of surrogate data to cyclic time series," *Physica D*, vol. 164, pp. 187–201, 2002.

[28] G. Sugihara and R. M. May, "Nonlinear forecasting as a way of distinguishing chaos from measurement error in time-series," *Nature*, vol. 344, pp. 734–741, 1990.

[29] T. Schreiber, "Interdisciplinary application of nonlinear time series methods," *Phys. Rep.-Rev. Sect. Phys. Lett.*, vol. 308, pp. 2–64, 1999.

[30] C. Rhodes and M. Morari, "Determining the model order of nonlinear input/output systems," *AIChE J.*, vol. 44, pp. 151–163, 1998.

[31] Control Loop Performance—ProcessDoctor (2004, Feb. 15). [Online]. Available: http://www. matrikon.com/products/processdoc/

[32] Plant Triage (2004, Feb. 15). [Online]. Available: http://www.expertune.com/planttriage.html

[33] C. Rhodes and M. Morari, "The false nearest neighbors algorithm: An overview," *Comput. Chem. Eng.*, vol. 21, pp. S1149–S1154, 1997.

[34] R. Hegger, H. Kantz, and T. Schreiber, "TISEAN 2.1 Surrogates Manual, Periodicity Artefacts,", http://www.mpipks-dresden.mpg.de/~tisean/TISEAN_2.1/index.html, Feb. 15, 2004. Retrieved.

[35] N. F. Thornhill, S. L. Shah, B. Huang, and A. Vishnubhotla, "Spectral principal component analysis of dynamic process data," *Control Eng. Pract.*, vol. 10, pp. 833–846, 2002.