DETECTION AND DIAGNOSIS OF PLANT-WIDE OSCILLATIONS USING GRANGER CAUSALITY

Vitor de Oliveira Thomaz
Espirito Santo Operations Unit
Petróleo Brasileiros S.A.
Av Nossa Senhora da Penha, 1688, Barro Vermelho
Vitória, Espirito Santo, Brazil
e-mail: vothomaz@petrobras.com.br

Celso José Munaro
Electrical Engineering Department
Federal University of Espirito Santo
Av. Fernando Ferrari, 514, Goiabeiras
Vitória, Espirito Santo, Brazil
e-mail: munaro@ele.ufes.br

ABSTRACT
This paper addresses the problem of detecting multiple oscillations and tracking down their root cause. The oscillations are detected by two methods based on spectral analysis, and the power related to the oscillations is used to identify the signals more likely to be the root cause. Since this method of finding the root cause is prone to fail when multiple frequencies are detected, Granger causality and subset selection methods are proposed. After a presentation of the methods, examples are provided using both simulated and industrial data. The clustering of signals containing the same frequency of oscillations followed by the methods to indicate the root cause is shown to be a good approach for rapid troubleshooting using only historical data.

KEY WORDS
Cause–effect analysis, fault diagnosis, spectral analysis

1 Introduction
Oscillation is a common disturbance of a control process. It may be due to poorly tuned controllers, the non-linearity of the actuators, natural phenomena, etc. These oscillations may propagate plant wide. Plant-wide oscillations cause poor control performance, off-spec product, low product quality, and high production costs. Research to develop techniques for detecting and diagnosing oscillations is motivated by the need to improve the productivity and process reliability, and to decrease the variability.

In modern plants there is a growing need for energy integration to improve the efficiency of energy use, optimize performance, and reduce on-site equipment. But achieving the goal of better energy integration for increased efficiency results in highly coupled systems. This coupling facilitates the propagation of oscillations and complicates detecting the root cause.

The problem of oscillation detection has been extensively studied and various techniques have been proposed. Different approaches to oscillation detection are presented by [1]. Among the approaches presented, we highlight the time domain methods, as proposed by [2]; frequency domain methods that make extensive use of the power spectral density of the autocorrelation function (ACF) of the signal, as proposed by [3]; the analysis of the spectral envelope, as proposed by [4]; as well as principal components analysis (PCA) proposed by [5]. In this paper, the techniques of ACF and Spectral Envelope will be used for detecting oscillations.

The problem of root-cause detection is an important problem that has been extensively studied and various techniques has been proposed, such as [6] and [7]. In both proposals, prior knowledge of the cause–effect relationships among the plant loops is required. A strategy using Granger Causality Analysis (GCA) is presented in [8] to identify the root cause of an oscillation using only historical data. A great advantage of using GCA is that no previous knowledge of the cause–effect relationship among the signals is necessary.

Eventually, the topology behind the causality is also required, and this information is not provided by GCA. Several methods to detect connectivity were compared in [9], and one of them will be used here to determine the topology after the causality is known.

In Section 2 of this paper, the techniques for oscillation detection based on spectral analysis will be presented. In Section 2.2, the method for the identification of Granger Causality is presented. In Section 2.3, a method to find the topology is presented. In Section 3, two case studies based on simulated data as well as data from an industrial plant will be presented, and in Section 4, observations and conclusions will be made.

2 Methods for detecting oscillations and root cause analysis

2.1 Methods for detecting oscillations
The method proposed in [3], called the ACF method, allows detecting multiple frequencies and provides an approach for testing the regularity of the oscillation period. This technique uses the zero-crossings of the autocovariance function to detect multiple oscillations, ob-
tained using the inverse Fourier transform of a selected band of frequencies of the power spectral density. The band of frequencies provides a candidate for the oscillatory signal, and if it corresponds to an oscillatory signal, its energy can be measured by integrating the power over the band of frequencies. The energy measured this way is and indication of the strength of the oscillation and will be used to select the signals that are likely candidates to be the root cause of the oscillation. An interesting feature of this method is its ability, due its use of the auto-covariance function, to eliminate noise.

The other method for detecting oscillations used here is the Spectral Envelope Method [4], which is a frequency domain technique to explore the periodicity of categorical time series. The method assigns numerical values to each of the categories using a statistical hypothesis test, followed by a spectral analysis of the resulting discrete-valued time series. An index called the oscillation contribution index (OCI) is proposed to isolate the key variables as the potential root cause candidates of the common oscillation.

### 2.2 Granger Causality Analysis

The method of Granger Causality Analysis was proposed in [10] for the identification of causal relationships between econometric models and it is based on the concept introduced in [11]. Granger’s proposal is based on the principle that the prediction of a signal \( X_1(t) \) can be improved by incorporating the past data of a second series \( X_2(t) \), so that \( X_2(t) \) has a causal influence on \( X_1(t) \) when compared with a model including only the data of \( X_1(t) \). Consider two temporal series \( X_1(t) \) and \( X_2(t) \) of two stochastic processes whose respective autoregressive models can be written as follows.

\[
X_1(k) = \sum_{j=1}^{m} A_{11,j}X_1(k-j) + \sum_{j=1}^{k} A_{12,j}X_2(k-j) + \epsilon_{1|2}(k)
\]

(1)

\[
X_2(k) = \sum_{j=1}^{m} A_{21,j}X_1(k-j) + \sum_{j=1}^{k} A_{22,j}X_2(k-j) + \epsilon_{2|1}(k)
\]

(2)

Here, the \( A_{ij} \) are the coefficients of the AR model, \( \epsilon_{ij}(k) \) is the prediction model error, where \( i, j = \{1, 2\} \) and \( m \) is the model order that defines the lag number to be included in the AR model. Excluding the cross relationships of each term \( (X_1) \) and \( (X_2) \) and rewriting a single variable AR model of each term, we have

\[
X_1(k) = \sum_{j=1}^{m} B_{11,j}X_1(k-j) + \epsilon_1(k)
\]

(3)

\[
X_2(k) = \sum_{j=1}^{m} A_{22,j}X_2(k-j) + \epsilon_2(k)
\]

(4)

In this way, if the variance of \( \epsilon_{1|2} \) is less than \( \epsilon_1 \) in statistical terms, this means that the prediction of \( X_1(k) \) is more accurate than the original model when the original values from \( X_2(k) \) are included. Thus, there is an influence of \( X_2(k) \) on \( X_1(k) \), and this can be quantified in the time domain as:

\[
F_{j\to i} = \ln \left( \frac{\text{var}(\epsilon_i)}{\text{var}(\epsilon_{ij})} \right)
\]

(5)

Here, \( \epsilon_i \) is obtained by regression of only \( X_1(k) \) and the \( \epsilon_{ij} \) is obtained when a complete model as in (1) and (2) is used.

Intuitively, the generalization of the bi-variable case to the multivariable case is simple and direct: in an \( n \)-variable system, it is that \( X_j \) causes \( X_i \) if the knowledge of \( X_j \) helps to predict \( X_i \) when all the other variables are included in the model. A matrix of the covariance matrix of the complete remainder model can be written as:

\[
\sum = \begin{bmatrix}
\sum_{i1} & \sum_{i2} & \cdots & \sum_{in} \\
\sum_{21} & \sum_{22} & \cdots & \vdots \\
\vdots & \vdots & \ddots & \vdots \\
\sum_{n1} & \sum_{n2} & \cdots & \sum_{nn}
\end{bmatrix}
\]

(6)

If the prediction variable \( X_j \) in time is excluded from the complete model, then the limited model is obtained. The remainder matrix is defined as:

\[
\rho = \begin{bmatrix}
\rho_{11} & \cdots & \rho_{1(j-1)} & \rho_{1(j+1)} & \cdots & \rho_{1n} \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
\rho_{(j-1)1} & \cdots & \rho_{(j-1)n} & \rho_{(j+1)1} & \cdots & \rho_{(j+1)n} \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
\rho_{n1} & \cdots & \rho_{n(j-1)} & \rho_{n(j-1)1} & \cdots & \rho_{nn}
\end{bmatrix}
\]

(7)

It has size \( (n - 1) \times (n - 1) \).

In this way, the Granger Causality of the variable \( j \) for the variable \( i \) conditioned by the other variable is measured by:

\[
F_{j\to i|n} = \ln \left( \frac{\rho_{ij}}{\sum_{ii}} \right)
\]

(8)

An adequate definition of the order of the model is a critical factor for the correct identification of the model to be used. A model with inadequate order will present an elevated prediction error. A correct definition of the order of the model can be made by two methods: the Akaike Information Criterion (AIC), or the Bayesian Information Criterion (BIC), which are presented in [12].

### 2.3 Methods to determine the topology

The causality is detected in general no matter whether the variables are connected directly or indirectly through
an intermediate variable. In [9], the topology behind the causality was sought by checking whether no further improvement of the predictions were achieved using a given variable, so that it could be excluded. However, as discussed in [13], the use of all variables to fit the model, even those that are not related to the output, may result in poor models, with unreliable conclusions. This problem may be addressed using the so called subset selection methods, designed to find the connectivity between the variables for high order systems. In [9], the methods Top–down, Bottom–up, and Lasso are discussed, and using different information-based criteria, eight hybrid methods are compared. A variation of the Top–down method is proposed and analyzed here. This method considers a model with all inputs (variables), and uses a given information-based criterion to decide on which inputs can be eliminated. The variables kept in the model are said to be connected to the output variable. One drawback of this method is its dependence on the search path, which is overcome in this paper by using all combinations of the inputs. Only the variables whose causality to the output were detected are used, avoiding the problem discussed above for the strategy proposed in [6].

3 Applications

The methodology proposed will be illustrated through its application to a simulated as well as an industrial data set. The application in the simulation model will be shown in 3.1 and that for the industrial model in Section 3.2.

3.1 Simulated system

Consider a system with three variables given by

\[
\begin{align*}
\dot{x}(k) &= 0.8x(k - 1) - 0.5x(k - 2) + 0.4z(k - 1) + \gamma y(k - 2) + \rho_3(k) \\
\dot{y}(k) &= 0.9y(k - 1) - 0.8y(k - 2) + \rho_1(k) \\
\dot{z}(k) &= 0.5z(k - 1) - 0.2z(k - 2) + 0.5y(k - 1) + \rho_2(k)
\end{align*}
\]

(9)

Here,

\[
\begin{align*}
\rho_1(k) &= \xi_1(k) + \alpha sin(k) \\
\rho_2(k) &= \sqrt{0.2}\xi_2(k) + 2\sin(2k) \\
\rho_3(k) &= \sqrt{0.3}\xi_3(k),
\end{align*}
\]

as used in [12], where \(\xi_n(k)\), with \(n = 1, 2, 3\), represents an independent white noise with unitary variance. Oscillatory inputs are added to illustrate the method of detecting oscillations, and the parameter \(\alpha\) is used to show the effect of different powers in the frequencies of the oscillation in these methods. Also, the parameter \(\gamma\) is used as in [12] to yield direct and indirect interactions among the variables, to test the methods of finding the topology behind the causality.

The simulation was carried out for \(\alpha = 0.5\) and \(\alpha = 4\) and the results are shown in Figure 1. The corresponding plots of the power spectral density and the spectral envelope used for the methods to detect the oscillations are shown in Figure 2. When the power of the signal with frequency 1 rad/s is low (\(\alpha = 4\)), the power of the corresponding bandwidth shown in Figure 2a becomes negligible.

Table 1 summarizes the results from both methods of detecting oscillations. When \(\alpha = 0.5\) only one frequency of oscillation (2 rad/s) is detected, while both frequencies of oscillations are detected when \(\alpha = 0.5\). The analysis of PSD and spectral envelope (Figure 2), used in these methods, shows clearly that frequencies with relatively small energy are eclipsed by those frequencies of higher energy.

![Figure 1. Time series (a) \(\alpha = 0.5\) and (b) \(\alpha = 4\)](image)

![Figure 2. (a) PSD to \(\alpha = 0.5\), (b) Spectral envelope to \(\alpha = 0.5\), (c) PSD to \(\alpha = 4\) and (d) Spectral envelope to \(\alpha = 4\)](image)

For the identification of the root cause by ACF analysis, it is assumed that signals with higher energy are probably the cause of the oscillations. Using the spectral envelope method, the OCI value is taken to indicate the probable cause of the oscillation. Evaluating the values of the energy and OCI shown in Table 1, one can note that the signals with greater energy are also those with higher OCI values, since they both come from spectral analysis. For \(\alpha = 0.5\), the root cause for the oscillation at the frequency 2 rad/s is indicated as being the variable \(z\). When \(\alpha = 4\), both methods indicate the variable \(x\) as the root cause, although the root cause is the same in both situations (the variable...
Table 1. Energy obtained by ACF and OCI

<table>
<thead>
<tr>
<th>Signal</th>
<th>Frequencies</th>
<th>1 rad/s</th>
<th>2 rad/s</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Energ</td>
<td>OCI</td>
<td>Energ</td>
</tr>
<tr>
<td>z</td>
<td>-</td>
<td>-</td>
<td>0.9928</td>
</tr>
<tr>
<td>y</td>
<td>-</td>
<td>-</td>
<td>0</td>
</tr>
<tr>
<td>x</td>
<td>-</td>
<td>-</td>
<td>0.9821</td>
</tr>
</tbody>
</table>

Table 2. Quantification of Granger Causality

| α = 4, γ = 0 |
|---|---|---|
| z | 1.5910 | 0.0099 |
| y | 0.0401 | 0.0213 |
| x | 0.3233 | 0.0100 |

| α = 4, γ = 0 |
|---|---|---|
| z | 1.3729 | 0.0145 |
| y | 0.0195 | 0.0631 |
| x | 0.3381 | 0.2305 |

Table 3. Topology

<table>
<thead>
<tr>
<th>γ = 0</th>
<th>γ = 0.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>z,y</td>
<td>-1.2535</td>
</tr>
<tr>
<td>y</td>
<td>-1.0101</td>
</tr>
<tr>
<td>z</td>
<td>-1.2661</td>
</tr>
</tbody>
</table>

In this case, the variable $z$ contains two frequencies of oscillation, and the energy of its spectrum is divided between them, yielding this spurious result for the root cause. The frequency of oscillation 1 rad/s was detected only for $\alpha = 4$. And the root cause was correctly indicated as being the variable $y$.

The root cause is now sought using a Granger causality analysis (GCA). The previous application of methods of detecting oscillations is very helpful since it clusters all the variables that are related in the frequency spectrum. Similarly, $\alpha = 0.5$ and $\alpha = 4$ are assigned and a GCA is performed, with results shown in Table 2: the significant values are in bold. The signals in the column are the cause and those in the rows are the effect, i.e., $F_{col \rightarrow lin}$. The model order was 14, obtained via the AIC criterion. The signals were previously checked to be covariance stationary [14].

For cases in which $\alpha = 0.5$ and $\alpha = 4$, the GCA analysis does not change. Evaluating the value of the causality quantification for $z$, we confirm that $F_{z \rightarrow y} < F_{y \rightarrow z}$, therefore one can conclude that $y$ causes $z$. Also, $F_{z \rightarrow x} > F_{x \rightarrow z}$, so that $z$ causes $x$. Upon evaluating the variable $y$, it is observed that $F_{y \rightarrow x} > F_{x \rightarrow y}$, leading to the conclusion that the variable $y$ causes $x$.

The GCA was performed on the three signals no matter the frequency of oscillation that was detected, since these analyses are independent. However, it is clear that the correct root causes were correctly found for different values of $\alpha$ and $\gamma$.

The GCA provides information about the cause–effect relationship between the variables. However, in some cases the connections behind this causality should be also known. To explore different topologies, the value of $\gamma$ was taken to be 0 and 0.2, and $\alpha = 4$, in two simulations. This parameter changes the effect of the variable $y$ on the variable $x$, which can be direct ($\gamma = 0.2$) or indirect ($\gamma = 0$). The result about causality is the same for both values (Table 1). The top–down method is a well-known method for subset selection. However, the result is dependent on the search path. For cases when the number of variables is not large, the combinations of possible inputs for a given output allows one to choose the model with the best AIC criterion. The variables whose contribution was mediated by other variables tend to be eliminated, making clear the topology of the variable under analysis. For the present example (Table 3), when the variable $x$ is taken as the output, it is clear that both $z$ and $y$ must be used as inputs when $\gamma = 0.2$, but only the variable $z$ is needed when $\gamma = 0$, since in this case the effect of the variable $y$ is mediated entirely by the variable $z$.

### 3.2 Industrial plant data

In this case, the proposed methodology is applied using data from three loops of a thermoelectric plant, Figure 3. The output of the loop level of the Steam Drum (LIC-01) determines the set-point of the flow loop of the feed water (FIC-01). The feed water comes from the reservoir (Deaerator), whose level is controlled by loop control LIC-02. The signals are collected for 14000 seconds with a 5-second sampling time ($T_s = 5$), Figure 4.

The frequency 0.0061 rad/s was detected in only one loop by the ACF method, being the root cause itself (Table 4). The frequency of oscillation 0.0102 rad/s was detected in three loops using the ACF and in two loops using the spectral envelope. Both methods indicate that loop FIC-01 is the probable root cause, since the value of its energy and OCI are the highest.

Table 5 shows the quantification matrix for the GCA, the numbers in bold are the significant values. Considering the column for LIC-01, it is observed that
Table 4. Energy obtained by ACF and OCI

<table>
<thead>
<tr>
<th>Signal</th>
<th>Frequencies</th>
<th>Energy</th>
<th>OCI</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIC-01</td>
<td>0.0061 rad/s</td>
<td>0.1394</td>
<td>-</td>
</tr>
<tr>
<td>FIC-01</td>
<td>0.0102 rad/s</td>
<td>0.8176</td>
<td>0</td>
</tr>
<tr>
<td>LIC-02</td>
<td>0.0061 rad/s</td>
<td>0.3786</td>
<td>0</td>
</tr>
<tr>
<td>FIC-01</td>
<td>0.0102 rad/s</td>
<td>0.9306</td>
<td>30.073</td>
</tr>
<tr>
<td>LIC-01</td>
<td>0.0061 rad/s</td>
<td>0.8918</td>
<td>29.3662</td>
</tr>
</tbody>
</table>

Table 5. Quantification of Granger Causality

<table>
<thead>
<tr>
<th>Signal</th>
<th>LIC-01</th>
<th>FIC-01</th>
<th>LIC-02</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIC-01</td>
<td>0</td>
<td>0.0103</td>
<td>0.0071</td>
</tr>
<tr>
<td>FIC-01</td>
<td>0.1558</td>
<td>0</td>
<td>0.0106</td>
</tr>
<tr>
<td>LIC-02</td>
<td>0.0055</td>
<td>0.0195</td>
<td>0</td>
</tr>
</tbody>
</table>

$F_{LIC-01 \rightarrow FIC-01} > F_{FIC-01 \rightarrow LIC-01}$ so that is possible to affirm that LIC-01 causes the FIC-01. For the column of FIC-01, $F_{FIC-01 \rightarrow LIC-02} > F_{LIC-02 \rightarrow FIC-01}$, so it possible to affirm that FIC-01 causes LIC-02. Since the values $F_{LIC-01 \rightarrow LIC-02}$ and $F_{LIC-02 \rightarrow LIC-01}$ are not significant, it is not possible to confirm a causal relationship between loops LIC-01 and LIC-02.

Based on the results of the Granger causality analysis, it is possible to confirm that the loop LIC-01 has an influence on the loop FIC-01, which is in agreement with their nature since the two loops are in cascade, where LIC-01 is the set-point of FIC-01, this type of cause/effect relationship being natural for this topology. Similarly, the flow variation observed in FIC-01 implies variation in the level of Deaerator, which is shown by GCA. The causality from FIC-01 to LIC-02 is also confirmed, though with a low value. This is expected, since variations in the flow (FIC-01) cause variations in the level (LIC-02). Although one could expect some relation between LIC-01 and LIC-02, it is not confirmed by GCA, eventually because it is small to be detected. Based on GCA, the root cause can be clearly indicated as being LIC-01. The result is not in accordance with provided by the values of energy or OCI (Table 4), but is far more reliable and confirmed by knowledge about the process. The topology analysis is meaningless in this case, and is not carried out.

4 Conclusion

Methods for detecting oscillations and for analysing their root cause were presented and applied to simulations and real data. The oscillations were detected using one method based on the autocorrelation function and another method based on the spectral envelope, with similar results for both simulations and real data. The power related to the frequencies of the spectrum related to the oscillations was used to indicate the source of the oscillations. The results were not correct for the case of multiple frequencies of oscillations. On the other hand, the use of Granger causality analysis and the subset selection method correctly indicated the sources of the oscillations and also the connectivity behind the causality relations, both for simulations and real data. Prior knowledge about the models and the process under analysis confirmed the results provided by the algorithms. Based on this analysis, a suitable approach is to detect those signals with the same frequency of oscillation using one of the two methods, find the signals that are related and under the same disturbance, apply the Granger causality analysis to find the causality relations and the root cause, and finally to apply a subset selection method to provide the topology of the connections.

References


Figure 5. (a) PSD to industrial data and (b) Spectral envelope to industrial data


