

# Data-based Causality Detection from a System Identification Perspective

Vinicius M. Marques, Celso J. Munaro, Sirish L. Shah

**Abstract**—The problem of detecting causality, from routine operating data, is reviewed from a system identification perspective. It is shown that even simple examples from the literature under Granger causality analysis do not have adequate model fit. As an alternative, this study uses the system identification platform to capture causality from process data. For example, the model inadequacy test is considered an important reason to reject a given causal relationship. The rich framework of system identification techniques and the choice of models to deal with exogenous variables and nonlinearities are shown to be an extremely suitable foundation to detect casual relationships. The utility of the proposed approach is illustrated by several benchmark examples including the analysis of routine operating data in an industrial case study.

**Index Terms**—Cause and effect relationship, system identification, correlations.

## I. INTRODUCTION

The investigation of cause and effect relationships in signals from industrial processes is useful to confirm and discover known and unknown material and signal flow paths and subsequently use this information to find the root cause of process faults and upsets such as plant-wide oscillations.

Granger causality analysis is a well-known method used mainly in the areas of econometrics and neurosciences [1]. This technique has also been applied recently to find the root cause of plant-wide oscillations from an industrial data set [2]. Oscillation detection methods isolate all variables with a common oscillation frequency of concern and the root cause is found using methods based on Granger Causality analysis. Methods based on Transfer Entropy [3] have also been used for the same purpose. A direct transfer entropy concept has also been proposed to detect whether there is a direct information from one variable to another [4].

Granger's original idea [1] was to conclude if one variable causes the other by comparing the prediction error of a linear regression model including only past values

of this same variable and comparing it with a linear regression model using past values of both variables. These results have been recently reviewed and applied in the area of neurosciences [5].

Although Granger's method is quite consistent with respect to causality aspects, modeling issues deserve attention when inferring cause and effect analysis in the definitive stage. It is important to check the adequacy of models that relate the variables of interest, since the causality inferences are only valid if the models adequately capture interactions between variables in the data [6].

The variance of the sum of squared errors is a simple measure that can be used to assess model adequacy. The model 'consistency' was proposed to check this adequacy [7] using the correlation vector of the real data and the correlation vector of simulated data generated via the MultiVariate AutoRegressive (MVAR) model. In contrast with residual analysis as used in the context of system identification, the consistency check in Granger causality does not give guidance on how to improve the model.

The inclusion of all measured variables in the MVAR model is a possible cause of an inadequate model [8]. If a variable that is not related to others is included in the model, correlation between the residuals and this variable may occur, and the error or the residual of the model will most likely be larger. The main critique of conventional Granger causality approach is that if the model structure is incorrect then the residuals are not likely to give evidence of causality between the signals considered in the models. The purpose of this paper is to highlight this drawback and suggest an alternative approach.

When variables from industrial processes are considered, the modeling issues become even more challenging, due to problems such as the presence of unmeasured disturbances or extraneous variables and nonlinearities.

In this paper, Granger causality analysis is reviewed from a system identification perspective and a new method is proposed. Additionally, the inadequacy of a model using the correlation vectors is considered as important evidence to reject interaction between variables. To overcome the confounding effect of extraneous variables, different model structures (such as ARX, ARMAX or BJ) can be used to handle the presence of such

V. M. Marques and C. J. Munaro are with the Department of Electrical Engineering, Federal University of Espírito Santo, Vitória, Brazil (e-mail: viniciusmoumar@yahoo.com.br; munaro@ele.ufes.br). Currently, C. J. Munaro is on sabbatical leave at the University of Alberta, Dept of Chemical and Materials Engineering with financial support of CNPq.

S. L. Shah is with the Department of Chemical and Materials Engineering, University of Alberta, Edmonton, Canada (e-mail: sirish.shah@ualberta.ca).

exogenous variables. The proposed method is extended to the multivariate case and allows one to detect the topology among variables including direct and indirect connectivity.

This paper is organized as follows. The theoretical basis of Granger Causality Analysis is introduced and discussion of data quality, validation, and preprocessing are presented in Section II. The use of system identification techniques for causality quantification is explained in Section III. Evaluation of the proposed method on a simulated as well as an industrial data set is presented in Section IV, followed by concluding remarks in Section V.

## II. GRANGER CAUSALITY QUANTIFICATION

We briefly review here the salient points of Granger Causality Analysis (GCA) and several aspects related to the models required to quantify the interaction between variables of interest.

Consider three time series,  $x_1(t)$ ,  $x_2(t)$  and  $x_3(t)$ , of length  $N$ , and assuming that they are jointly stationary, their joint representations can be described by:

$$x_m(t) = \sum_{j=1}^k A_{m1}x_1(t-j) + \sum_{j=1}^k A_{m2}x_2(t-j) + \sum_{j=1}^k A_{m3}x_3(t-j) + \epsilon_m(t) \quad (1)$$

where  $m = 1, 2, 3$  and  $k$  is the order of the model and the noise terms  $\epsilon_m$  are independent and identically distributed (iid).

The term 'lag' will be used for the correlation and cross-correlation computation

To measure the causal influence of  $x_2(t)$  on  $x_1(t)$ , a MVAR model for  $x_1(t)$  without including past values of  $x_2(t)$  is obtained,

$$x_1(t) = \sum_{j=1}^k A_{11,j}x_1(t-j) + \sum_{j=1}^k A_{13,j}x_3(t-j) + \epsilon_{12}(t) \quad (2)$$

and the measure of causality is obtained by comparing the variance of error  $\epsilon_{12}(t)$ , as obtained when the time series of all three variables is included, with the variance of the error  $\epsilon_{12}(t)$  in which the time series  $x_2(t)$  is not included in the MVAR model. This causal influence can be quantified [5] by

$$F_{2 \rightarrow 1} = \ln \frac{\text{var}(\epsilon_{12}(t))}{\text{var}(\epsilon_1(t))}. \quad (3)$$

In total four models are required to test causality between a pair of variables (for example,  $x_1$  and  $x_2$ ): two models (1) and (2) to obtain  $F_{2 \rightarrow 1}$  and two models to obtain  $F_{1 \rightarrow 2}$ .

To confirm the results, additional tests are evaluated to measure the quality of the models and the significance of the values obtained.

Before quantifying causality using Granger analysis, the data are detrended and mean centred. Differencing is also applied if data is nonstationary (see [6] for details).

The four MVAR models required to quantify causality between two variables are usually validated by checking if they were able to capture the correlation structure in the data, using a measure called model 'consistency' given in [7]<sup>1</sup>. One way of improving consistency is to increase the model order. In general, the order is chosen based on the Akaike information criterion (AIC) and the Bayesian information criterion (BIC) [9]. The other possibility is to include or to exclude variables from the model, since it is well-known that this is one of the reasons for model inadequacy.

Consistency can be calculated for one variable or for all variables. In the later case an average measure of models is obtained. This measure is not very useful for improving the model.

The Durbin-Watson test is also applied to check the whiteness of residuals. In real data, the presence of exogenous/latent variables can cause problems in this test. Partial GCA was proposed in this situation [10], assuming that the exogenous/latent signal affects all variables.

The Granger causality magnitude is statistically significant if the coefficients in the corresponding  $A_{ij}$  elements of the MVAR model are jointly significantly different from zero. The F-test on the null hypotheses to check if elements  $A_{ij}$  are zero is used for this purpose, with corrections for multiple comparisons [6].

## III. CAUSALITY VIA SYSTEM IDENTIFICATION TECHNIQUES

For the time being, let us consider two variables  $x_1(t)$  and  $x_2(t)$  for which causality has to be determined. As in GCA let us consider the following 2 models

$$A_1(q)x_1(t) = B_1(q)x_2(t - \theta_1) + \epsilon_1(t) \quad (4a)$$

$$A_2(q)x_2(t) = B_2(q)x_1(t - \theta_2) + \epsilon_2(t) \quad (4b)$$

where  $A_1(q)$ ,  $B_1(q)$ ,  $A_2(q)$ ,  $B_2(q)$  are polynomials with parameters to be estimated and  $\epsilon_1(t)$  and  $\epsilon_2(t)$  are the residuals. In the system identification (SI) context, (4a) is an ARX model with  $x_1(t)$  as output and  $x_2(t)$  as input. The input and output are exchanged for model (4b). The time delays  $\theta_1$  and  $\theta_2$  and the model orders are assumed to be known apriori [11].

Now, let us assume that  $x_2(t)$  is the input that is the cause of output  $x_1(t)$ . If the model order and the time delay  $\theta_1$  are properly chosen, a good model can be obtained. The quality of the model (4a) can be measured by the fitness value given by [11]:

$$fit = 100 \times \left( 1 - \frac{\|y - \hat{y}\|}{\|y - \bar{y}\|} \right) \quad (5)$$

where  $y$  is the data output,  $\hat{y}$  is the one step ahead prediction of  $y$  and  $\bar{y}$  is the mean value of  $y$ . The one-step ahead prediction is generally a simple criteria to satisfy, meaning that even a poor model will yield a good estimate. One could also obtain the k-step ahead

<sup>1</sup>Due to lack of space the reader is referred to references [6] and [7] for the definition of consistency.

prediction. This would be a hard test to satisfy especially when  $k$  is large and would require a good model from a data set that is not affected by extraneous variables. For models with fit values close to 100%, no statistical tests are required, since the error is negligible. If this is not the case, the residual analysis can be performed to improve the model. The whiteness of residuals is checked by examining its autocorrelation. This is equivalent to the Durbin-Watson test used in GCA. If this test fails, the use of other model structures such as ARMAX or BJ can be considered, which is not available for the GCA approach.

In this paper the fit values are computed for one step ahead prediction.

The second test is the cross-correlation between the residuals and the input. The plot of the cross-correlation as a function of the lags considering the confidence limits gives valuable insight into the adequacy of the model structure. For example, the correlation of  $x_2(t + \tau)$  with  $\epsilon_1(t)$  for negative  $\tau$  is an indication of signal  $x_1(t)$  being a feedback signal to  $x_2(t)$  [11].

Therefore, if  $x_2(t)$  is the input for  $x_1(t)$ , a model whose residual tests is satisfactory can be found. If this is the case, what happens if the parameters of model (4b) are estimated, that is  $x_1$  is the input with  $x_2$  as output? Since  $x_1(t)$  and  $x_2(t)$  are related, and if the sample time is commensurate with the time constants, one should be able to find a model with good fit value. However, the residual analysis will indicate correlation between the residuals  $\epsilon_1(t)$  and  $x_2(t)$ , due to the feedback from  $x_2(t)$  to  $x_1(t)$ .

Based on our assumption of causality, this is correct and the causality from  $x_2(t)$  to  $x_1(t)$  is confirmed. A lower fit value is also possible for the model fitted with input(s) and output(s) exchanged.

The other possible situations are:

- Feedback: the cross-correlation between residuals and inputs for negative lags will appear for both models and with good fit values for both.
- No causality: no models will be validated and also the fit values will be low.

One arrives at a conclusion after models in both directions are fitted to data.

A question that needs to be considered is the threshold value of fit that can be considered good. This value is dependent on the data, and is affected by factors such as noise variance and sample time. For all experiments conducted in this work, a fit value was considered valid when higher than 25% of the maximum fit value.

#### A. Extension to the multivariate case

It is possible that two variables may be connected directly or indirectly through an intermediate variable. In this case the previous analysis would only provide a 'yes' or 'no' answer to connectivity. If the answer is 'yes', one would still be unsure whether that connection is direct or indirect. The approach used here is similar

to that used in [5], where no further improvement of the predictions using a given variable was an indication that it was mediated by other variables, so that it could be excluded. The same approach is used here using the fit value instead of the variance of the residuals. To illustrate this idea consider a case involving 3 variables that may be connected directly or indirectly as shown in Figure 1.

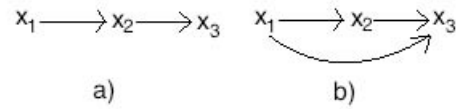


Fig. 1. Example to show direct and indirect interaction paths.

The proposed method would indicate that  $x_1$  causes  $x_2$  and that  $x_2$  causes  $x_3$  in both cases. It could also indicate that  $x_1$  causes  $x_3$ . To find if  $x_1$  causes  $x_3$  directly or indirectly, a second model using  $x_1$  and  $x_2$  as inputs and  $x_3$  as output should be evaluated. This model is then compared to models from  $x_1$  to  $x_3$  and  $x_2$  to  $x_3$ , using the fit values. If the fit value does not increase, one would then conclude that all input information for  $x_3$  comes from  $x_2$ .

Therefore, after the pair wise analysis, all variables that are affected by more than one variable can be checked for direct or indirect interaction in this manner.

The problem of finding topology can be also addressed using subset selection methods. Several methods are compared in [12], and they could be also used here after causality has been determined.

#### B. Comparing Granger Causality and System Identification Approaches

Both methodologies are based on fitting models to data in order to infer causality, but they differ in how models are obtained and used. The method proposed here, that will be referred as the SI method, provides more flexibility in obtaining a good model. While the GCA method requires good models, the SI approach simply rejects models that are not validated, concluding that the variables do not interact. Table I summarizes the main features of these approaches. The purpose of the SI method as used here is to simply determine causality and not to obtain the 'best' model to relate the cause(s) to the effect(s).

### IV. CASE STUDIES

In these examples, GCA is performed using the GCCA toolbox [6]. All other analysis are performed using functions from the System Identification Toolbox in Matlab.

#### A. Example 1

Here we consider the benchmark example from [5]. In this example, two models are analysed to show the possibility of detecting direct and indirect causality:  $x_2$  affects  $x_3$  directly and  $x_1$  indirectly, and  $x_3$  affects  $x_1$

TABLE I  
COMPARING GCA AND SI APPROACH

	GCA	SI
Improvement of model quality	-Increase model order	-Increase model order -Change the time delay and model structure
Model validation	-Auto-correlation of residuals -Model Consistency -Squared sum of residuals	- Residual analysis using auto-correlation and cross-correlation -Squared sum of residuals
Interaction quantification	-Variance of the residual	-Fit value
Decision about causality	-Direction of interaction that gives the largest relative magnitude	-Fit value and cross-correlation analysis
Direct/indirect effect	-Conditional GC	-Pair-wise analysis followed by a multi-variate analysis

directly (according to the topology in Figure 1.a). In this work, the variables  $X$ ,  $Y$ ,  $Z$  as used in the original paper [5] were substituted for  $x_1$ ,  $x_2$ ,  $x_3$ , respectively.

TABLE II  
EXAMPLE 1, MODEL 1: GC AND CONSISTENCY VALUES

GC values	$x_1$	$x_2$	$x_3$
$x_1$	-	<b>0.0098</b>	<b>0.1562</b>
$x_2$	0.0013	-	0.0032
$x_3$	0.0068	<b>1.0781</b>	-
Consistency	$x_1$	$x_2$	$x_3$
$x_1$	-	87.938	85.986
$x_2$	77.594	-	77.551
$x_3$	91.706	75.035	-

The interactions using GCA are shown in Table II for models of order 4. The bold values are those that are significant (considering the F-test>1 and p-value<0.1). The variables listed in row one (namely  $x_1$ ,  $x_2$  and  $x_3$ ) are the cause variables and the corresponding effect variables appear in the first column. Thus,  $x_2$  and  $x_3$  cause  $x_1$  and  $x_2$  causes  $x_3$ . Table II also shows the consistency for each MVAR model. Although several values of consistency are smaller than 80% (as a threshold proposed in [6]), the average value (85.7316) is higher than 80%. The smaller values correspond to MVAR model for  $x_2$ . The residual analysis of the MVAR models gives a valuable insight to understand the rationale for causality analysis (Figure 2).

The plots in the diagonal are the auto-correlation of residuals for the three MVAR variables considering the past values of all variables. The red line represents the 99%-confidence interval, showing that the residuals are uncorrelated. Cross-correlation appears between a variable and the residual error of the MVAR model if this variable does not contribute to reduce this error. In Fig. 2, this situation is shown in plots (2,1) and (2,3) for variables  $x_1$  and  $x_3$  and the MVAR model of  $x_2$ , and in plot (3,1) for the variable  $x_1$  and the MVAR model

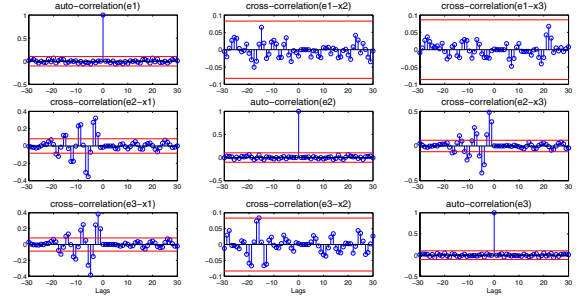


Fig. 2. Residual analysis for example 1, model 1 using GC.

of  $x_3$ .

The SI method is then applied using an ARX model of order 4, with the choice of the time delay that gives the best model quality (fit value) for each pair. The result is shown in Table III. The bold values correspond to interactions that need to be considered to obtain adequate models. The values in parenthesis correspond to models with good fit values whose correlation of residuals with each input was outside confidence interval for negative lags (Figure 3). For example, plot (2,1) clearly shows that correlation for negative lags happens when  $x_1$  is used as input to  $x_2$ . The same applies to pairs  $x_1$ - $x_3$  and  $x_2$ - $x_3$  (see plots (3,1) and (2,3)). The higher the fit value the stronger the interaction. The conclusion is that  $x_2$  affects both  $x_3$  and  $x_1$ , but its effect is stronger on  $x_3$ , and that  $x_3$  affects  $x_1$ . What remains is to check if  $x_2$  affects  $x_1$  directly or through  $x_3$ . The model using  $x_2$  and  $x_3$  as input was obtained and its fit value was 60.4. Therefore,  $x_2$  does not add more information than that contained in  $x_3$  to explain  $x_1$ , and we conclude that  $x_2$  only affects  $x_3$ .

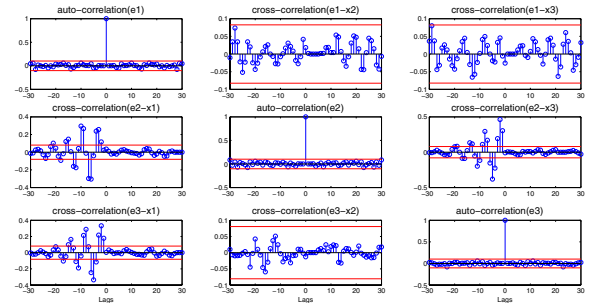


Fig. 3. Residual analysis for example 1, model 1 using SI. In this case residuals were computed for ARX models.

A second model is related to the situation corresponding to Figure 1.b, containing both direct and indirect causal influences from  $x_2(t)$  to  $x_1(t)$ .

In [5], it is shown that conditional causality is able to find that  $x_2$  affects  $x_3$  directly or indirectly using the two models. The method proposed here is now applied to the second model, using ARX models with same order as before (order=4). The result is shown in Table III.

Again,  $x_2$  and  $x_3$  affect  $x_1$  and  $x_2$  affects  $x_3$ . To check if  $x_2$  affects  $x_1$  directly, an ARX model of order 4 with the same delays that yields fit values from Table III is fitted using  $x_2$  and  $x_3$  as inputs and  $x_1$  as output, and its fit value was 73.2. Since this value is greater than 71 (Table III), we conclude that both  $x_2$  and  $x_3$  directly affect  $x_1$ .

TABLE III  
EXAMPLE 1: FIT VALUES

Model 1	$x_1$	$x_2$	$x_3$
$x_1$	-	<b>58.20</b>	<b>60.66</b>
$x_2$	(47.58)	-	(47.65)
$x_3$	(45.27)	<b>68.39</b>	-
Model 2	$x_1$	$x_2$	$x_3$
$x_1$	-	<b>71</b>	<b>71</b>
$x_2$	(45)	0	(45)
$x_3$	(45)	<b>66</b>	0

### B. Example 2

This example uses routine operating data from three control loops from a thermoelectric power plant (Figure 4). The level loop from a steam drum (LIC400) adjusts the set point for the water flow loop (FIC408). This flow comes from a reservoir (deaerator) whose level is given by LIC430. The signals were collected over 14000 seconds with  $T_s = 5s$  (Figure 5, left column). To see the interaction among these signals, a filtered data set with a second order Butterworth filter of frequency  $0.05Hz/Hz$  is plotted for the first to 2000s only (Figure 5, right column). Clearly, LIC400 affects FIC408: when the level decreases the flow is increased. Although the flow FIC408 comes from the tank with level LIC430, it is difficult to establish a relationship between them. This may happen because the level LIC430 depends on other variables and also the controller tuning on this level loop is detuned or very slow, as should be the case. The signals were detrended and differenced for use in the algorithms.

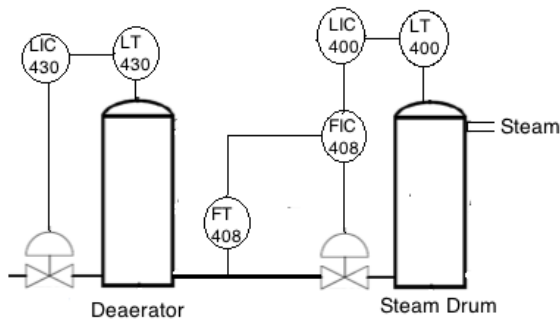


Fig. 4. Control strategy for the boiler loops in a power plant.

1) *Analysis via GCA*: The order used was 13, according to AIC (BIC indicated an order of 6). The model consistencies were 12.5, 50.0 and 10.6 for the three MVAR models, respectively. The average model consistency was

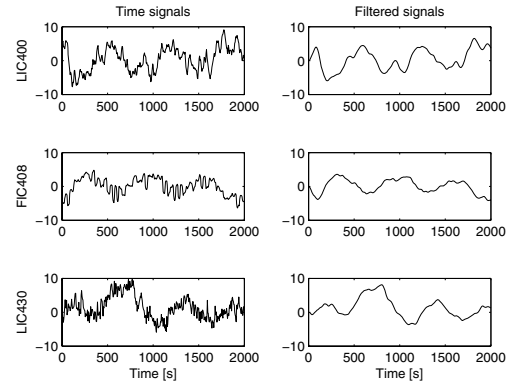


Fig. 5. Routine operating data from a thermoelectric power plant: Time (left column) and filtered (right column) signals.

14.2. The Durbin-Watson test returned values of [0.9945 0.6017 0.9562] for the three loops, respectively. Increasing the order to 40, the Durbin-Watson test returned values of [0.9923 0.9957 0.9138] for the three loops, but the consistency value increased only to 19.7, indicating that for this example increasing the order only improves the model consistency slightly. The residual analysis shown in Figure 6 illustrates the correlation between residuals and inputs for MVAR model for order=40.

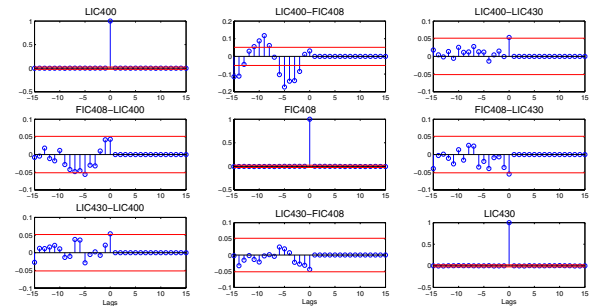


Fig. 6. Residual analysis for example 2 using GCA. Diagonal plots: auto-correlation of residue; off-diagonal plots: cross-correlation between residual and input, in this order.

Table IV shows in bold the GCA values that are significant (for  $F\text{-test} > 1$  and  $p\text{-value} < 0.1$ ) using order 40. They show that LIC400 clearly causes FIC408. The metrics also show that FIC408 causes LIC430, albeit with a very low value, but there is no criterion to reject this interaction.

A difficulty in such situations is to decide a threshold value to discard the interaction.

2) *Analysis via system identification*: An 8th order model was used, and the time delay was chosen to increase the model quality (fit value). An ARX model was fitted but the residuals for most models were auto-correlated. Also, correlation between residuals and input variables was also present for positive lags, indicating that the model was inadequate. An ARMAX model was

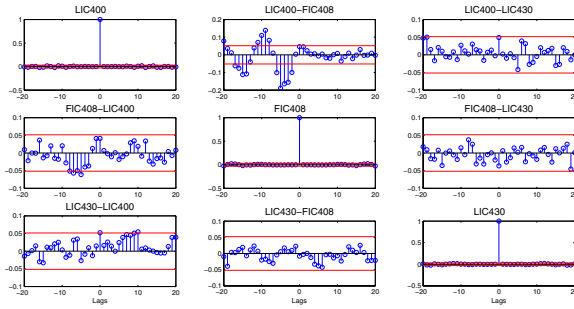


Fig. 7. Residual analysis for example 2 using an ARMAX model in SI: Diagonal plots: auto-correlation of residue; off-diagonal plots: cross-correlation residue - input, in this order.

then fitted using the same order, resulting in uncorrelated residuals and correlation only for negative lags. The results are shown in Table IV.

TABLE IV  
EXAMPLE 2: INDUSTRY DATA

GCA values	LIC400	FIC408	LIC430
LIC400	-	<b>0.0243</b>	0.0162
FIC408	<b>0.2475</b>	-	0.0140
LIC430	0.0215	<b>0.0254</b>	-
Fit values	LIC400	FIC408	LIC430
LIC400	-	(8.1)	7.1
FIC408	<b>33.7</b>	0	25.0
LIC430	6.3	6.8	0

The effect of LIC400 in FIC408 is clear. No relation is found between LIC430 and the other variables. The correlation between residuals and inputs is shown in Figure 7, where it is clear that correlation exists (for negative lags) only between FIC408 and the residuals for the model that has LIC400 as output. A model using LIC400 and LIC430 as inputs and FIC408 as output was also fitted, using the same order and delays used for the results of Table IV, and the fit value was 33.7, which indicates that LIC430 does not help explain FIC408.

## V. CONCLUDING REMARKS

The Granger causality analysis is reviewed under a system identification framework. The main weakness of the approach appears to be its intrinsic necessity of including all variables to obtain model parameters, yielding inadequate models with larger residual errors especially when variables that are not related to the model output are also included. On the other hand when system identification techniques are used to fit models to pairwise variables, and if the models are deemed to be inadequate as per the correlation tests, then one can quickly infer that the variables used do not interact. These results are extended to the multivariate case using multiple input models.

ARX models provide good results in most examples, but more complex model structures can be used to handle situations such as auto-correlated residuals, presence

of exogenous variables, nonlinearities and so on. The residual analysis clearly shows that it is prudent to use the rich framework of system identification and the associated validation methods to confirm interaction or relationship between variables when data from industry is considered.

It is clear that GCA is model structure agnostic and this may not bode well when one is interested in investigating causality from process systems data where there are inevitably unmeasured disturbances plus feedback relationships. It is for this reason that the model adequacy tests are of paramount importance and it is these tests that allow us to confirm or reject topology relationships.

In summary, GCA provides some causality information but not enough to confirm or refute interaction in all cases. On the other hand, system identification through residual analysis and its ability to select different model structure is able to unmask the interaction between the variables.

## ACKNOWLEDGMENTS

The authors are grateful to CNPQ and thank Arcelor-Mittal Tubarão for the data and the information provided for the 2<sup>nd</sup> example in this paper.

## REFERENCES

- [1] C. W. J. Granger, "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica*, vol. 37, no. 3, pp. 424–438, August 1969.
- [2] T. Yuan and Qin, "Root cause diagnosis of plant-wide oscillations using granger causality," *8th IFAC Symposium on Advanced Control of Chemical Processes*, Singapore 2012.
- [3] M. Bauer, J. W. Cox, M. H. Caveness, J. J. Downs, and N. F. Thornhill, "Finding the direction of disturbance propagation in a chemical process using transfer entropy," *IEEE Trans. Cont. Syst. Tech.*, vol. 15, no. 1, 2007.
- [4] P. Duan, F. Yang, T. Chen, and S. L. Shah, "Direct causality detection via the transfer entropy approach," *In press with IEEE Transactions on Control Systems Technology*, 2012.
- [5] M. Ding, Y. Chen, and S. L. Bressler, "Granger causality: Basic theory and application to neuroscience," *Elsevier Science*, 2008.
- [6] A. Seth, "A matlab toolbox for granger causal connectivity analysis," *Journal of Neuroscience Methods*, vol. 186, pp. 262–273, 2010.
- [7] M. Ding, S. Bressler, W. Yang, and H. Liang, "Short-window spectral analysis of cortical event-related potentials by adaptive multivariate autoregressive modeling: data preprocessing, model validation, and variability assessment," *Biological cybernetics*, vol. 83, pp. 35–45, 2000.
- [8] C. Granger and P. Newbold, "Spurious regressions in econometrics," *Journal of Econometrics*, vol. 2, pp. 111–120, 1974.
- [9] G. Schwartz, "Estimating the dimension of a model," *The annals of statistics*, vol. 6, no. 2, pp. 461–464, 1978.
- [10] S. Guo, A. Seth, K. Kendrick, C. Zhou, and J. Feng, "Partial granger causality: eliminating exogenous inputs and latent variables," *Journal of Neuroscience Methods*, vol. 172, pp. 79–93, 2008.
- [11] L. Ljung, *System Identification: Theory for the User.*, 2nd ed. Prentice-Hall, 1999.
- [12] N. Hsu, H. Hung, and Y. Chang, "Subset selection for vector autoregressive processes using lasso," *Computational Statistics and Data Analysis*, vol. 52, pp. 3645–3657, 2008.