Linear And Nonlinear Methods For Dynamical System Correlation Analysis

Adochiei F.C.
Faculty of Electrical Engineering
University POLITEHNICA of Bucharest
Bucharest, Romania
felix.adochiei@upb.ro

Edu I ,Jula N.
Department of Aircraft Integrated Systems and Mechanics
Military Technical Academy
Bucharest, Romania
edu_ioana_raluca@yahoo.com, nicolae.jula@gmail.com

***Abstract* – In recent decades different methods have been developed to analyse correlations of complex dynamic systems. In the field of medical tests, there is a growing interest in how the new information can be gained on the interactions between the main physiological regulating mechanisms both in healthy and diseased individuals. Recent advances in dynamic linear, nonlinear and information theory, allowed the study of information flows between causal and non-causal time series using multivariate analysis. To investigate the correlations in and between the main control systems, as well as for the quantification of the interactions between these complex systems a variety of linear and non-linear methods have been proposed up to the present time. In this paper will be presented the most used approaches, both linear and nonlinear, to quantify the direct or indirect correlations and directionality of these interactions (master-slave relationship).**

***Keywords* – dynamic systems.**

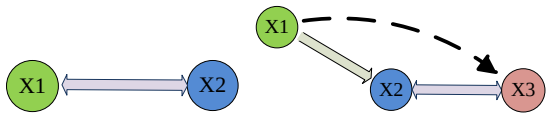
I. INTRODUCTION

Autonomous control systems are considered complex dynamic systems characterized by interactions, both linear and nonlinear, of component subsystems [1]. These systems and subsystems physiological interactions can be described as a closed loop feedback mechanisms (reverse reaction) and feed forward (forward reaction) mechanisms. There are important issues that must be taken into account in the analysis of physiological control systems time series:

- These complex systems interact both directly and indirectly. It is suggested the use of multidimensional approach instead of two-dimensional,
- Physiological time series (e.g., electrocardiogram, systolic and diastolic blood flow, pletismogram, respiratory, respiratory flow) are sometimes noisy, non-stationary and stationary short periods,
- Evaluation of correlation and causality can be achieved by applying the linear or nonlinear time series. While the nonlinear methods favour dynamic interactions of complex signals, linear methods favour individual behaviour biological signals in the frequency domain.

In order to investigate the relationships between these systems a variety of methods have been proposed. In capturing complex nonlinear interactions, occurring in physiological systems and subsystems, most approaches are based on the notion of Granger causality (GC) defined as follows: if a time series has a causal influence over another time series knowledge of the first past time series is useful to predict future values of the second time series [2]. The term refers to a causal relationship of cause and effect, an event causing another event.

Direct correlation between two time series x_1 and x_2 is understood that the interdependence between $x_1 \leftrightarrow x_2$ (from a time series to the other), while the indirect correlation is produced by means of another time series, shortly there is direct a correlation between: $x1 \leftrightarrow x_2$ and $x_3 \leftrightarrow x_2$ and indirect correlation between $x_1 \leftrightarrow x_3$, (correlation effects are mediated by one or more different time series). The definition of correlation generalizes the concept of causal correlation to describe the interactions, back and forth, between two time series [3] (Figure 1 ab).



(a) (b)

Fig. 1. - Suggestive representation of the types of direct and indirect correlations for (a) a bivariate case and (b) a case with multiple variables. (a) There is a direct correlation between $x_1 \leftrightarrow x_2$; (b) There are direct correlation between x_1 and x_2 and x_2 and x_3 and indirect correlation between x_1 and x_3 indirectly mediated by x_2 (correlation direction: unidirectional represented by "→" and "←" symbols and bidirectional by the "↔" symbol)

The most applied approach in literature to assess both direct and indirect correlations, in terms of levels or directions of correlation, can be grouped into five classes: Granger causality, Linear predictions, Entropy, Symbolisation and Phase synchronization.

II. GRANGER CAUSALITY

Norbert Wiener introduced a clear definition of the concept of causality between two time series in a statistical framework as follows: for two time series recorded simultaneously, a time series can be called causal another time series, where the latter series can be better predicted by using characteristics of the former [4, 5]. Today this concept is known as Granger causality (GC). *GC* between two

processes *X1(t)* and *X2(t)* is defined as: *X1(t)* has a causal influence on *X2(t), (X1(t) → X2(t))*, where knowledge of the past, both of *X1(t)* and of *X2(t)* reduces the error variance of the prediction of *X2(t* compared to the situation where it is known only the past of X2(t) (past and present causes future, but not vice versa) [2, 6]. Granger causality can be assessed both by means of linear and nonlinear methods. Among the most effective linear approaches that describe GC in frequency domain are Partial direct coherence (PDC) [7, 8] and Direct transfer function (Directed Transfer Function - DTF) [9] methods. These methods are based on the transfer matrix of an autoregressive model and assume stationary signals in the time interval to be investigated [10].

III. Linear Methods for Detecting Granger Causality

All the linear methods assess GC based on a multivariate autoregressive model (*MVAR*) parameters, and may be divided into methods for assessing GC in the time domain or methods for assessing GC in the frequency domain.

The method for detecting causal relationships between several linear time series is based on linear prediction theory. For a stationary time series *x(t)*, is taken into account the following autoregressive (AR) prediction of the current value of *x(t)* based on the latest *m* measurements:

$$x(t) = \sum_{j=1}^{m} a_j x(t-j) + \varepsilon_x(t) \qquad (2.1)$$

where εx (t) is the prediction error. If there is another stationary time series y (t), acquired simultaneously with x (t), then the following predictions of the values of x (t), both based on their previous values and based on past values of y (t) can be calculated according to:

$$x(t) = \sum_{j=1}^{m} a_j x(t-j) + \sum_{j=1}^{m} b_j y(t-j) + \varepsilon_{x|y}(t) \qquad (2.2)$$

A. Granger causality evaluation in the time domain

Geweke [11] was the first to propose a linear bivariate analysis of time series for Granger causality analysis, using the method called Linear Granger Causality (*LGC*) based on the error prediction variance of time series and associated with a statistical test for causality [6].

If are considered two stochastic processes, *X1* and *X2*, and LGC parameter values are normalized to the [0,1] range, then according to the method, if *X2* process has no causal influence on *X1* process, $LGC_{X_2 \to X_1}$ parameter value is 0, which means that that by knowing the past values of X2, X1 prediction does not improve, but when $LGC_{X_2 \to X_1} > 0$ the existence of a causal influences can be assumed. High values of $LGC_{X_1 \to X_2}$ indicate the presence of bidirectional correlations or feedback relationships between the two time series [12].

B. Granger causality evaluation in the frequency domain

The most important approaches for assessing the linear GC in the frequency domain are Partial Directed Coherence - PDC) (*PDC*) [13, 14] and Directed Transfer Function (*DTF*) [15]. These approaches are based on an autoregressive model (AR) adapted for these applications and assume stationary of the signals in the time interval under investigation [10] .

a) Partial Directed Coherence ***(PDC)***

PDC method is a parametric approach based on an *m*-dimensional *MVAR* model of order *p*. This method has the advantage of direct and indirect causal information transfer detection, because measures only direct effects/ influences of multivariate dynamic system signals. *PDC* method allows the quantification of the causal correlation level between two signals *Xi* and *Xi* depending on the *f* frequency. Values acquired using the *PDC* method were normalized between *0* and *1 v*alues and thus the causal correlation from *Xj* to *Xi* is surprised as a function of frequency *f*, yielding *PDC* parameter value equal to *0* when *Xj* does not influence *Xi* and *PDC* parameter value equal to *1* when all causal influences originating from *Xj* are directed towards *Xi* [8, 12, 16] [17].

b) Directed Transfer Function ***(DTF)***

DTF allows the determination of direct causal influences between two signals in relation to all other signals of the analysed system by applying a *MVAR* model using a transfer matrix to describe the transfer of causal information [9, 18]. By using the DFT method can be estimated both direct and indirect effects of a series of time to another. For this reason, a differentiation between direct and indirect causal interactions or both is not possible, leading to detect the presence of a greater number of interactions than are actually present [25].

IV. Nonlinear Methods for Detecting Granger Causality

Most limiting factor in nonlinear Granger causality analysis is the model selection because it should be adapted accordingly the dynamics of investigated bio-signals. From this point of view there are promising approaches in the literature, able to quantify the causality in the case of non-linear signals, such as methods based on the identification of linear and non-linear models, local and global.

Methods based on identifying global nonlinear models were proposed by Faes et al. [19]

They have introduced a new method for the detection of the correlation and causality between two time series based on nonlinear auto-regressive (*NAR*) models and the non-linear auto-regressive exogenous (*NARX*) models. To assess the *GC* through *NARX* models, Faes et al. examined the prediction mean square error and proposed a new parameter to which have been assigned values of 0 and 1, value of 0 being assigned to a completely predictable time series and the value of 1 to a series of completely unpredictable time series.

Causality or influence direction between two signals *x* and *y* can be investigated by reversing the roles of input and

output of the two series and by calculating normalized absolute and relative Predictability improvement (PI) parameters.

Another approach for *GC* analysis was introduced by Riedl et al. and is based on nonlinear additive autoregressive models (*NAARX*) with special inputs provided for bivariate time series [20].

Predictability improvement (*PI*) was measured by a cross-validation criterion that takes into account all the predictors used. To avoid false detections or false correlations, it was considered the existence of actual correlations when were detected more than 80% of the time series elements investigated, and otherwise were noted as artefact [20].

V. Nonlinear Predictions

Nonlinear prediction is based on the cross prediction and is similar to approaches based on predictability improvement (PI) in terms of the methodology that underlies them, but they differ because does not measure *GC*, but highlights another aspect (concept) of causality exploiting cross predictability asymmetry (*CP*) when used to detect interactions direction between the two series.

Farmer and Sidorowich [21] were the first who introduce the concept of local linear prediction called K nearest neighbour prediction, which was later integrated into different approaches. Based on this concept was developed the local linear prediction method, sometimes called linear approximation method of the nearest neighbour method.

Based on this approach, Fas et al. introduced bivariate linear prediction method for the identification of the causal interdependence between two series of time (*x, y*). This method has been associated with an approach of "cross - validation" for short-term time series analysis [22]. Fas et al. studied three different approaches of nonlinear mutual prediction, namely: Cross Prediction, Mixed Prediction and Predictability improvement methods for testing the capacity of these methods to evaluate the correlation levels and directionality for bivariate time series [23].

VI. Methods Based on Entropy

Entropy-based methods can also be used in the analysis of causality of complex system signals as follows: if are considered two stationary time series representing the records of *x* and *y* electrophysiological signals, of *n* and respectively *m* length, and for each time series is determined the probability of occurrence of each value, then to highlight the influence of signals on each other, may be used the formalism used in the discrete channel transmission analysis. If the two signals that are deemed to be the input and output of such a discrete channel and the values of signals are considered to represent the input and the output of channel, in order to highlight the influence of signals on each other, by using the calculated probabilities, can be determined the entropies of *x* and *y* signals. These entropies are: the input - output entropy *H(X, Y)*, conditional entropies *H (X|Y)* and *H (Y|X)* and the mutual information, *I(X, Y)*.

A. *Mutual information*

Mutual information measures the average information transmitted through the discrete transmission channel (between two discrete stationary time series, *x* and *y*). The estimation of mutual information between two discrete stationary time series, *x* and *y*, is based on the determination of *H(X)* and *H(y)* entropies and of joint entropy *H(x, y)*, where *p(xi)* and *p(yi)* are the probability distributions of *x* and *y* and *p (xi, yi)* is the joint probability distribution of both series.

B. *Conditional entropy*

Conditional entropy $H(X|Y)$ quantifies the level of correlation between two time series *(x, y)* and is a measure of the complexity of x in terms of *y*, the conditional entropy of *x* with respect to *y*.

C. *Transfer entropy*

Schreiber [24] proposed a theoretical approach based on information, called transfer entropy (*TE*), able to distinguish between two processes that send and receive information, to detect asymmetries of interaction and to capture the extent to which the influence of the dynamics of a process influence the conditional transition probabilities of another process.

TE measures using *GC* using the Predictability improvement methods and extends the concept of *SE* by considering transition probabilities in the detriment of static probabilities. In general, if there is no information flow from Y process to X process, then Y's status has no influence on the transition probabilities of X.

VII. Symbolisation

Symbolic methods allow a quantitative assessment, in detail, of the dynamics of short time series. Direct analysis of successive amplitudes of the signals is based on discrete states (symbols).

A. *Joint symbolic dynamics*

Joint symbolic dynamics method, was introduced by Baumert [25] and is based on bivariate dynamic processes analysis by means of symbols [26]. By using *JSD* method, physiological signal changes, on short term, can be captured and it is possible to assess the overall short-term correlations between complex physiologic dynamic systems. This approach has the advantage of being insensitive to non-stationary time series and is able to capture nonlinear bivariate correlations.

B. *Symbolic coupling traces*

An extension of *JSD* method, introduced by Wessel et al. [27], called Symbolic coupling traces method is based on structural models analysis and allows the detection of directionality (or bidirectionality) of delayed correlations in short bivariate time series.

Conclusions

Currently we are working on the development of new methodologies, algorithms and various software packages to quantify and analyse the relationships and correlations of

signals between the central nervous system and autonomic nervous system. The approaches presented in the paper are promising tools for detecting multivariate information flows and may provide additional prognostic information in the medical field by surpassing and complementing traditional linear techniques.

REFERENCES

[1] A. Voss, S. Schulz, R. Schroeder, M. Baumert, and P. Caminal, "Methods derived from nonlinear dynamics for analysing heart rate variability," *Philos Transact A Math Phys Eng Sci*, vol. 367, pp. 277-96, Jan 28 2009.

[2] C. W. J. Granger, "Investigating Causal Relations by Econometric Models and Cross-spectral Methods.," *Econometrica*, vol. 37, pp. 424-438., 1969.

[3] L. Faes and G. Nollo, *Multivariate frequency domain analysis of causal interactions in physiological time series*. Rijeka, Croatia, 2011.

[4] N. Wiener, "The theory of prediction," *Beckenbach, E.F. (Ed.), Modern Mathematics for Engineers.*, 1956.

[5] M. Kaminski, M. Ding, W. A. Truccolo, and S. L. Bressler, "Evaluating causal relations in neural systems: granger causality, directed transfer function and statistical assessment of significance," *Biol Cybern*, vol. 85, pp. 145-57, Aug 2001.

[6] B. Gourevitch, R. L. Bouquin-Jeannes, and G. Faucon, "Linear and nonlinear causality between signals: methods, examples and neurophysiological applications," *Biol Cybern*, vol. 95, pp. 349-69, Oct 2006.

[7] L. A. Baccala and K. Sameshima, "Overcoming the limitations of correlation analysis for many simultaneously processed neural structures," *Prog Brain Res*, vol. 130, pp. 33-47, 2001.

[8] L. A. Baccala and K. Sameshima, "Partial directed coherence: a new concept in neural structure determination.," *Biol. Cybern.*, vol. 84, pp. 463-474., 2001.

[9] M. J. Kaminski, Blinowska, K.J., "A new method of the description of the information flow in the brain structures.," *Biol. Cybern.* , vol. 65, pp. 203-210., 1991.

[10] W. Hesse, E. Moller, M. Arnold, and B. Schack, "The use of time-variant EEG Granger causality for inspecting directed interdependencies of neural assemblies," *J Neurosci Methods*, vol. 124, pp. 27-44, Mar 30 2003.

[11] J. Geweke, "Measurement of linear dependence and feedback between multiple time series," *Journal of the American Statistical Association*, vol. 77, pp. 304-313, 1982.

[12] E. Pereda, R. Q. Quiroga, and J. Bhattacharya, "Nonlinear multivariate analysis of neurophysiological signals," *Prog Neurobiol*, vol. 77, pp. 1-37, Sep-Oct 2005.

[13] L. A. Baccala and K. Sameshima, "Overcoming the limitations of correlation analysis for many simultaneously processed neural structures," *Prog Brain Res*, vol. 130,, pp. 33-47, 2001.

[14] L. A. Baccala and K. Sameshima, "Partial directed coherence: a new concept in neural structure determination.," *Biol. Cybern.*, vol. 84,, pp. 463-474., 2001.

[15] M. J. Kaminski, Blinowska, K.J., "A new method of the description of the information flow in the brain structures.," *Biol. Cybern.* , vol. 65,, pp. 203-210., 1991.

[16] L. Faes and G. Nollo, "Extended causal modeling to assess Partial Directed Coherence in multiple time series with significant instantaneous interactions," *Biol Cybern*, vol. 103, pp. 387-400, Nov 2010.

[17] B. Schelter, M. Winterhalder, M. Eichler, M. Peifer, B. Hellwig, B. Guschlbauer, C. H. Lucking, R. Dahlhaus, and J. Timmer, "Testing for directed influences among neural signals using partial directed coherence," *J Neurosci Methods*, vol. 152, pp. 210-9, Apr 15 2006.

[18] A. Korzeniewska, M. Manczak, M. Kaminski, K. J. Blinowska, and S. Kasicki, "Determination of information flow direction among brain structures by a modified directed transfer function (dDTF) method," *J Neurosci Methods*, vol. 125, pp. 195-207, May 30 2003.

[19] L. Faes, G. Nollo, and K. H. Chon, "Assessment of Granger causality by nonlinear model identification: application to short-term cardiovascular variability," *Ann Biomed Eng*, vol. 36, pp. 381-95, Mar 2008.

[20] M. Riedl, A. Suhrbier, H. Stepan, J. Kurths, and N. Wessel, "Short-term couplings of the cardiovascular system in pregnant women suffering from pre-eclampsia," *Philos Transact A Math Phys Eng Sci*, vol. 368, pp. 2237-50, May 13 2010.

[21] J. D. Farmer and J. J. Sidorowich, "Predicting chaotic time series," *Phys Rev Lett*, vol. 59, pp. 845-848, Aug 24 1987.

[22] L. Faes and G. Nollo, "Bivariate nonlinear prediction to quantify the strength of complex dynamical interactions in short-term cardiovascular variability," *Med Biol Eng Comput*, vol. 44, pp. 383-92, May 2006.

[23] L. Faes, A. Porta, and G. Nollo, "Mutual nonlinear prediction as a tool to evaluate coupling strength and directionality in bivariate time series: comparison among different strategies based on k nearest neighbors," *Phys Rev E Stat Nonlin Soft Matter Phys*, vol. 78, p. 026201, Aug 2008.

[24] T. Schreiber, "Measuring information transfer," *Phys Rev Lett*, vol. 85, pp. 461-4, Jul 10 2000.

[25] M. Baumert, T. Walther, J. Hopfe, H. Stepan, R. Faber, and A. Voss, "Joint symbolic dynamic analysis of beat-to-beat interactions of heart rate and systolic blood pressure in normal pregnancy," *Med Biol Eng Comput*, vol. 40, pp. 241-5, Mar 2002.

[26] A. Voss, J. Kurths, H. J. Kleiner, A. Witt, N. Wessel, P. Saparin, K. J. Osterziel, R. Schurath, and R. Dietz, "The application of methods of non-linear dynamics for the improved and predictive recognition of patients threatened by sudden cardiac death," *Cardiovasc Res*, vol. 31, pp. 419-33, Mar 1996.

[27] N. Wessel, A. Suhrbier, M. Riedl, N. Marwan, H. Malberg, G. Bretthauer, T. Penzel, and J. Kurths, "Detection of time-delayed interactions in biosignals using symbolic coupling traces," *EPL*, vol. 87, p. 10004, 2009.