Granger Causality-based Information Fusion Applied to Electrical Measurements from Power Transformers

J. Rodríguez-Rivero[c,1], J. Ramirez[a], F.J. Martinez-Murcia[a], F. Segovia[a], A. Ortiz[a], D. Salas[a], D. Castillo[a], I.A. Illan[a], C.G. Puntonet[b], C. Jimenez-Mesa[a], F.J. Leiva[c], S. Carillo[c], J. Suckling[e], J.M. Gorriz[a,e,*]

[a]*Dpt. Signal Theory, Networking and Communications, University of Granada, Spain*
[b]*Dpt. Computer Technology and Architecture, University of Granada, Spain*
[c]*Endesa, Madrid, Spain*
[d]*Department of Communications Engineering. University of Málaga, Spain*
[e]*Department of Psychiatry. University of Cambridge, UK*

Abstract

In the immediate future, with the increasing presence of electrical vehicles and the large increase in the use of renewable energies, it will be crucial that distribution power networks are managed, supervised and exploited in a similar way as the transmission power systems were in previous decades. To achieve this, the underlying infrastructure requires automated monitoring and digitization, including *smart-meters*, wide-band communication systems, electronic device based-local controllers, and the Internet of Things. All of these technologies demand a huge amount of data to be curated, processed, interpreted and fused with the aim of real-time predictive control and supervision of medium/low voltage transformer substations. Wiener-Granger causality, a statistical notion of causal inference based on Information Fusion could help in the prediction of electrical behaviour arising from common causal dependencies. Originally developed in econometrics, it has successfully been applied to several fields of research such as the neurosciences and is applicable to time series data whereby cause pre-

---

[*]Corresponding author
Email address: `gorriz@ugr.es` (J.M. Gorriz)
URL:
`https://www.endesa.com/es/proyectos/a201902-pastora-inteligencia-artificial-red-distribucion.html`
(S. Carillo)
[1]jacob.rodriguez@enel.com

cedes effect. In this paper, we demonstrate the potential of this methodology in the context of power measures for providing theoretical models of low/medium power transformers. Up to our knowledge, the proposed method in this context is the first attempt to build a data-driven power system model based on G-causality. In particular, we analysed directed functional connectivity of electrical measures providing a statistical description of observed responses, and identified the causal structure within data in an exploratory analysis. Pair-wise conditional G-causality of power transformers, their independent evolution in time, and the joint evolution in time and frequency are discussed and analysed in the experimental section.

1. Introduction

Contemporary medium and low voltage distribution networks are exploited under the assumption that they are designed to handle any kind of peak demand, avoiding rising network congestion, unacceptable voltage levels, or the influence of disruptive developments such as the introduction of the electrical vehicle, self-producers or powerwall systems. In order to accomplish an efficient predictive control of involved the electrical assets, the distribution systems have to incorporate much more intelligence than before, involving a whole spectrum of digital technologies: sensorization, smart meters, broadband communications, local electronic device controllers, IoT (Internet of Things), SCADAs (Supervision, Control and Acquisition of Data), energy management centers, advanced data processing software (data analytics), optimal control, intelligent reporting, and so on. The possibility to correlate and fuse information from electrical, image and other kind of data sources, such as the dissolved gas concentrations in transformer oil, with several simultaneous measurements, could be useful to distinguish the root cause of failures [1]. This is actually the main goal of the Monica and Pastora projects [2] which are being developed by ENEL and other

international companies, and aim to accurately determine the actual situation of low and medium-voltage distribution grids in real-time, preventing and resolving network failures, by fusing information from several SCADA measurements. The operating conditions of a power system, i.e. a power transformer (PT), at a given point in time can be determined if the network model and complex phasor voltages are known [3]. The power system may be operating in one of three possible states; that is, normal, emergency and restorative, as the operating conditions evolve in time. Corrective control measures based on state estimators [4, 5] or pattern recognition [1, 6, 7, 8] are continuously monitoring power systems through measurements acquired by the SCADA systems, maintaining the operating conditions in a nominal and secure state.

On the other hand, Wiener-Granger causality (G-causality) [9, 10, 11] is a statistical fusion information method with the aim of analyzing the flow of information between time series [12, 13, 14]. Originally conceptualized in [15] and firstly analysed, in terms of autoregressive (AR) modelling of stochastic processes, in [9], this popular method is based on the principles that (i) a cause occurs before its effect and (ii) knowledge of a cause improves prediction of its effect [12]. In a nutshell, a variable X is said to G-cause a variable Y if the past of X fused with the one of Y helps predict the future of Y more accurately than only using the past of Y. Moreover, G-causality aims to quantify directed functional connectivity by means of a statistical description of fused observed responses. On the contrary, methods for identifying effective connectivity aim to highlight "the simplest possible circuit diagram explaining observed responses" as shown in [16, 17, 18].

Time series forecasting assumes the use of an accurate model of systems, in this case power transformers, to predict future values of the signals based on previously observed values and/or other exogenous time series, which are incorporated to the main signal by combining them via early, intermediate or late fusion [26]. In this paper, we propose to our knowledge, the first supervisory data-driven system based on G-causality for early information fusion. The proposed method is able to monitor and predict the status of the power trans-

3

former by evaluating linear AR models of stochastic processes which identify
directed functional connectivity; that is, a statistical relationship among ob-
served variables that reflects (but not unambiguously) the underlying physical
mechanisms of complex systems. In attempting to provide the determination
of vector AutoRegressive (VAR) models for power transformers, we sought to
address the following questions: (1) How effective is a G-causality paradigm at
modelling SCADA measurements? (2) Which mechanisms or connections be-
tween variables do VAR models focus on during the normal states of a power
transformer? (3) How does the VAR model evolve over a long time period; for
example, one year? (4) Can the VAR model effectively describe the general sta-
tus of a power transformer beyond the particularities in the operational mode as
a part of a complete power system (transmission, sub-transmission, distribution
and generation systems)?

2. Background on G-causality

Given two jointly distributed vector-valued stochastic processes $\mathbf{X} = [X_1, X_2, \ldots]$; $\mathbf{Y} = [Y_1, Y_2, \ldots]$, $\mathbf{Y}$ is said to G-cause $\mathbf{X}$ if and only if $\mathbf{X}$, conditional on its own past, is dependent of the past of $\mathbf{Y}$; in other words, the past values of Y yield information about the current value of $\mathbf{X}$ beyond information already contained in its own past. This is the notion of "causality" that is used throughout this paper and in many others [23], the Wiener-Granger sense, and is not necessarily linked with other causal notions, such as the ones described in [19, 16, 20, 21, 22].

2.1. Multivariate VAR model

Given a realisation $\mathbf{u}$ of length m of a discrete-time stationary vector stochas-
tic process $\mathbf{U}_1, \mathbf{U}_2, \ldots$, where $\mathbf{u}_t$ is a time-dependent n-dimensional vector with
components $u_{1t}, u_{2t}, \ldots, u_{nt}$, a p-th order VAR model for the process can be
defined as:

$$\mathbf{U}_t = \sum_{k=1}^{p} \mathbf{A}_k \mathbf{U}_{t-k} + \epsilon_t \tag{1}$$

4

The $n \times n$ real-valued matrices $\mathbf{A}_k$ are the regression coefficients, and the iid n-dimensional stochastic process ϵ_t the error terms. The model parameters, i.e. the coefficients $\mathbf{A}_k$ and the $n \times n$ error covariance matrix $\Sigma = Cov(\epsilon_t)$, are assumed to be time-independent (stationarity). Once the predictive model in equation 1 is fitted to data (realizations) by the estimation of the regression parameters, the value of the process at time t can be computed based on the past values up to the lag $t - p$. These regression coefficients, that are usually considered the predictable structure of the data whilst the error terms form the unpredictable structure[23], are determined via the Yule-Walker equations on the autocovariance sequence.

2.2. Un/conditional G-causality in the time domain

Based on the VAR model described in the previous section, the time-domain unconditional G-causality is illustrated as follows: Given two jointly-distributed multivariate processes $\mathbf{U}_{1,t}, \mathbf{U}_{2,t}$, the G-causality from $\mathbf{U}_{1,t}$ to $\mathbf{U}_{2,t}$, written $F_{\mathbf{U}_{2,t} \to \mathbf{U}_{1,t}}$, is defined by quantifying the improvement in the prediction of $\mathbf{U}_{1,t}$ when the past of $\mathbf{U}_{2,t}$ is included by early information fusion in the VAR model, over and above the restricted model including only its own past. This definition of causality is concerned with the comparison of different linear regression models of data [24], that is, the (restricted) VAR (p) model shown in equation 1 evaluated on the process $\mathbf{U}_{1,t}$ with the following fused or extended model:

$$\mathbf{U}_{1,t} = \sum_{k=1}^{p} \mathbf{A}'_{1,k} \mathbf{U}_{1,t-k} + \sum_{k=1}^{p} \mathbf{A}'_{21,k} \mathbf{U}_{2,t-k} + \epsilon'_{1,t} \tag{2}$$

with the residuals covariance matrix defined as:

$$\Sigma(\epsilon'_{1,t}) \equiv \mathrm{Cov}\left(\epsilon'_{1,t} \right) \tag{3}$$

where the dependence of $\mathbf{U}_{1,t}$ on the past of $\mathbf{U}_{2,t}$, given its own past, is established in the coefficients $\mathbf{A}_{21,k}$. If there is no conditional dependence of $\mathbf{U}_{1,t}$ on the past of $\mathbf{U}_{2,t}$ then $\mathbf{A}_{21,k} = 0$ for $k = 1, \ldots, p$. To measure the degree to which the regression in equation 2 yields a better model than the restricted regression in equation 1, we make use of the Maximum likelihood (ML) theory

5

[25] for the analysis of parametric data modelling. In particular, a test statistic for the null hypothesis of no conditional dependence $\mathbf{U}_{1,t}$ on the past of $\mathbf{U}_{2,t}$ based on the logarithmic likelihood ratio test (LRT) is proposed as:

$$F_{\mathbf{U}_{2,t} \to \mathbf{U}_{1,t}} \equiv \ln \frac{|\Sigma_1|}{|\Sigma_1'|} \tag{4}$$

where Σ_1 and Σ_1' are the residual covariance matrices of the VAR models in equations 1 and 2, respectively.

The aforementioned unconditional G-causality statistic is limited in the description of statistical dependencies in the following sense: if there are joint dependencies between $\mathbf{U}_{1,t}$ and $\mathbf{U}_{2,t}$ and a third set of variables, i.e. $\{\mathbf{U}_{d,t}\}$ for $d = 3, \ldots, D$, then spurious causalities may be reported [23]. Fortunately, they may be eliminated by "conditioning out' the common dependencies available in the data. In this case, the null hypothesis to be tested by the LRT is still the one previously defined and the causality $\mathbf{U}_{2,t} \to \mathbf{U}_{1,t}$ conditioned on $\{\mathbf{U}_{d,t}\}$, which we write $F_{\mathbf{U}_{2,t} \to \mathbf{U}_{1,t} | \mathbf{U}_{d,t}}$, is again as in equation 4, but with the inclusion of $\{\mathbf{U}_{d,t}\}$ in both regressions (restricted and fused regressions), accounting for its joint effect; that is, we redefine:

$$\hat{\mathbf{U}}_{1,t} = \begin{pmatrix} \mathbf{U}_{1,t} \\ \{\mathbf{U}_{d,t}\} \end{pmatrix} \tag{5}$$

2.3. G-causality in the frequency domain

The spectral decomposition of G-causality is a powerful tool that provides additional information in the frequency domain [10]. Via the Yule-Walker equations and using the definition of the cross-power spectral density (cpsd) of the two-jointly distributed multivariate processes $\mathbf{U}_{1,t}, \mathbf{U}_{2,t}$, we define the (unconditional) spectral G-causality as:

$$f_{\mathbf{U}_{2,t} \to \mathbf{U}_{1,t}} \equiv \ln \frac{|\mathbf{S}_1(w)|}{|\mathbf{S}_1(w) - \mathbf{H}_{21}(w)\Sigma_2 \mathbf{H}_{21}^*(w)|} \tag{6}$$

where $\mathbf{S}_1(w) \longleftrightarrow \mathrm{Cov}(\mathbf{U}_{1,t}, \mathbf{U}_{1,t-k})$ is the cpsd of $\mathbf{U}_{1,t}$; that is, the Fourier transform of the $n \times n$ matrix covariance sequence, $\mathbf{H}_{21}(w) = \left(\mathbf{I} - \sum_{k=1}^{p} \mathbf{A}_{21} e^{-jkw}\right)^{-1}$ and $\Sigma_2 = \mathrm{Cov}(\epsilon_{2,t})$. Again we evaluated, but in the frequency domain, the "best model" between the restricted and the fused VAR (p) regressions.

6

3. Materials and preprocessing for WSS processes

The database used in this paper was acquired during 2018 during the execution of the above mentioned projects [2] which were developed by ENEL (Endesa) and collaborators. The set consisted of electrical variables measured from a collection of 17 power transformers, a total of 42 variables were recorded mainly at a sample rate of 12 samples/hour for the whole calendar year of 2018. Subsequently, the data was resynchronized at this sample rate by linear interpolation. Most of these signals include the 'Phase Imbalance', 'Active Energy Exported', 'Active Energy Imported', 'Capacitive Reactive Energy Exported', 'Capacitive Reactive Energy Imported', 'Inductive Reactive Energy Exported', 'Inductive Reactive Energy Imported', 'Intensity R', 'Intensity S', 'Intensity T' [2], 'Active Power R', 'Active Power S', 'Active Power T', 'Active Power', 'Reactive Power R', 'Reactive Power S', 'Reactive Power T', 'Reactive Power', 'Room Temperature', 'Transformer Temperature', 'Tension R', 'Tension S', and 'Tension T' that were used to monitor the operation of the power transformer, as well as fault diagnosis and rapid intervention. The temperature of the insulating oil in the power transformer is referred to Temperature (T) (an indirect measure of the PT temperature) and is one of principal electrical variables to be monitored and controlled in order to prevent system failure within this time series prediction problem.

3.1. Preprocessing

The crucial condition for a consistent G-causality analysis is that all the variables must be covariance stationary (CS, or wide-sense stationarity). Thus, all the preprocessing was devoted to obtaining time-independent first and second statistical moments (mean and variance) of each variable. Multivariate VAR (MVAR) models would be invalid and may contain so-called "spurious regression" results [23]. Good practice when attempting to achieve this assumption is to apply one or more of the following preprocessing steps [27]: (i) linear detrend,

---

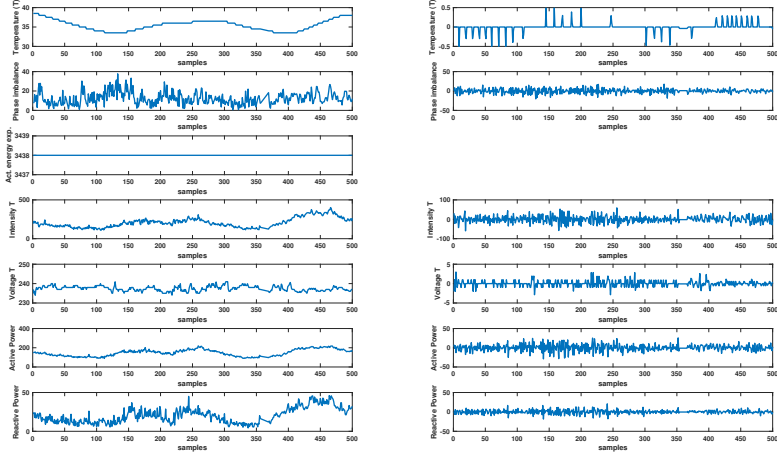[2]R,S,T stands for the three current phases

7

Figure 1: The effect of preprocessing applied to the set of raw variables in seeking to obtain covariance stationary processes (wide sense stationary). Observe the effect of differencing the raw signal (second column), i.e. the constant variable (third row on the left) is not considered in the following analyses.

(ii) removal of temporal mean and division by temporal standard deviation, (iii) for multi-trial data, removal of ensemble mean and division by ensemble standard deviation, and (iv) differencing and/or windowing as necessary to achieve CS. In particular, the variables analysed in this scenario required the application of i), z-zcore computation for artefact rejection (replacing outliers by the ensemble mean), null or negligible variables/trials removal after differencing and the application of iv) by the definition of a number of observation of 500 per trial and at least 40 trials at each analysed period (up to 4 segments in the whole yearly time period, see figure 1). To achieve such data collection and the following ensemble comparisons, we selected only the variables that were jointly available in all the power transformers (see figure 2). This led us to discard several variables from collections and a complete set of measures (power transformer #8) where only three variables fulfilled the above mentioned CS preconditions.
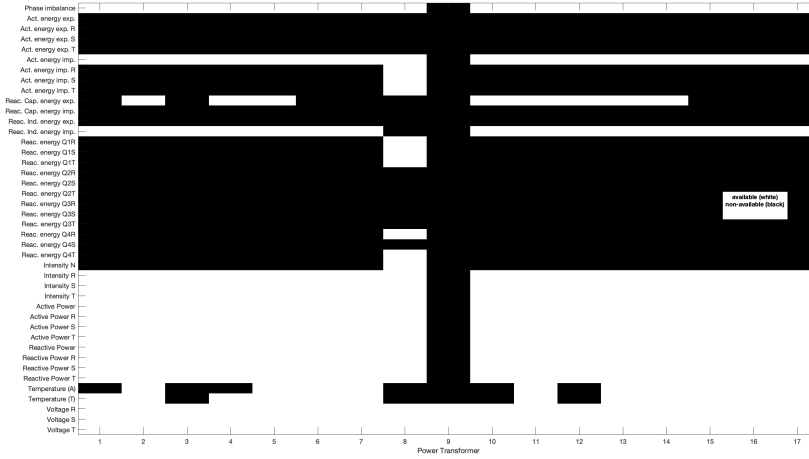
8

Figure 2: Variable selection process applied to the set of PTs. The variables displayed in black were discarded for the subsequent G-casuality based analysis.

4. Experimental Analysis

We carried out a complete set of experiments based on multiple equivalent representations of a VAR model (regression parameters, autocovariance sequence, cpsd, etc. of underlying processes) for computing multivariate G-causality from time series data in the time and frequency domains. We employed a modified version of the MVGC Matlab toolbox [23] which was mainly designed with application to empirical neuroscience data, although G-causal inference is a statistical framework which has been productively applied in many areas when the assumptions underpinning the method are satisfied.

First, we needed to determine the number of lags to be included in the estimation of multivariate VAR models, i.e. the model order. To this purpose, a criterion that balances the variance accounted by the model against the number of coefficients to be estimated may be selected based on the Akaike information criterion (AIC)or the Bayesian information criterion (BIC). Usually, a selection of a small p value produces a poor representation of the data, whereas a large p value can lead to problems of model estimation. By computing both criteria,
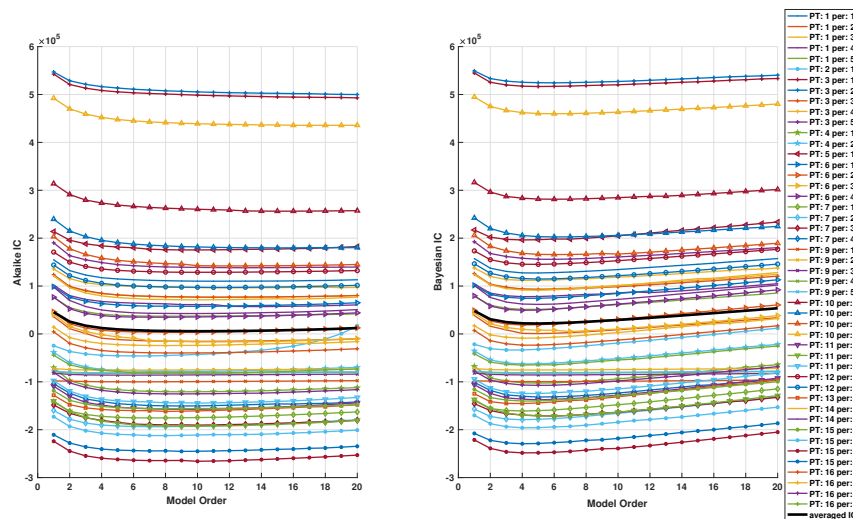
9

Figure 3: On the left, the AIC criterion vs model order for the complete set (50) of PTs and operation periods. On the right the BIC criterion. Note in black solid lines the mean value of both criteria.

we surprisingly found (see figures 3 and 4 that all the PTs (from 1 to 17) are modelled by almost the same number of lags (around 10 for AIC and 6 for BIC) during different periods (*per*) and operating with different conditions. To be conservative, we selected the first criterion for the rest of experiments.

4.1. How effective is a G-causality paradigm at modelling SCADA measurements?

The promising behaviour observed in the determination of the model order is not enough to state that the G-causality framework is an effective tool for PT time series modelling, although it could be considered as a good starting point. Given a model order, first we need to estimate the corresponding VAR model parameters, i.e. Σ and $\mathbf{A}_k$ in equation 1, for the previously selected model order. By fitting them using the standard OLS method, we calculated the VAR parameters from the autocovariance sequence for both, the full and the reduced regressions, in equation 2, and finally, checked the resulting causalities for significance at a given level ($\alpha = 0.05$) in equation 4, using Geweke's χ^2 test
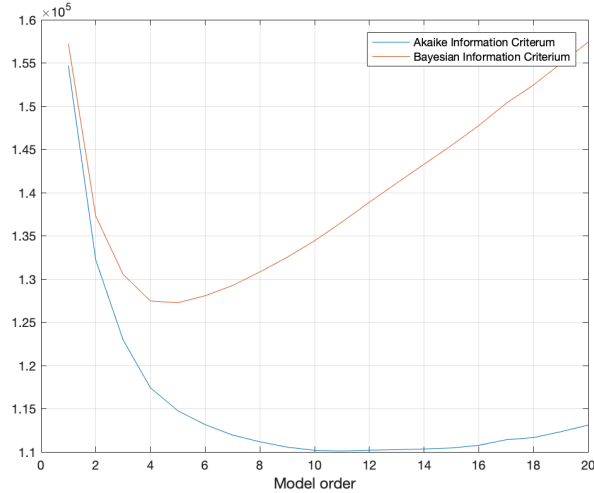
10

Figure 4: Comparison of the AIC and BIC versus model order for the determination of the model order using the PT #1.

[11] and a multiple hypothesis test correction based on the false discovery rate [30]. As an example, in figure 5 we show the error sequence, ϵ_t, obtained from the fitted parameters for the analysed variables in PT #1, and the resulting tests for significance. All the simulations showed a VAR spectral radius less than 1 (stability), positive-definite residuals covariance matrices, and a set of significant strongly connected variables which were found in almost all the PTs. After the computation of LRTs in time and frequency domains in equations 4 and 6 we statistically inferred properties of the underlying populations by the standard large-sample theory [28]. In particular we applied the F-test for small populations to establish the statistical significance of the estimated causality against the null hypothesis (no causality). Finally, under limited samples sizes the theoretical asymptotic distributions may not be sufficiently accurate, thus a non-parametric empirical based approach, such as the non-parametric bootstrap [29] was tested. In figure 6 we show all the aforementioned analyses that established the statistical significance of the results at a given critical value $(\alpha = 0.05)$.
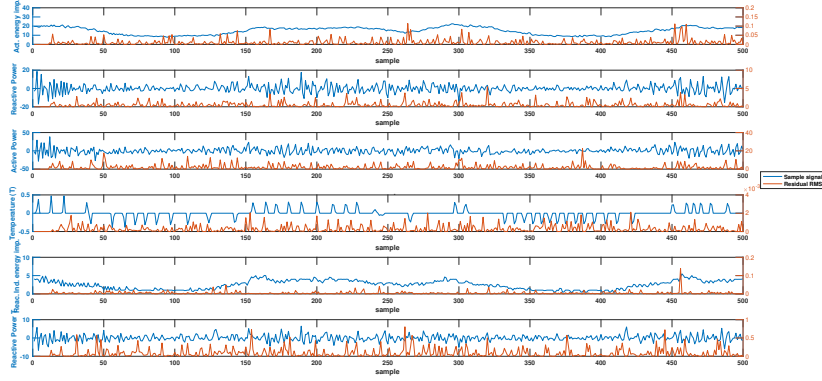
11

Figure 5: MSE sequence computed from the set of trials and a sample sequence within the ensemble using the PT #1. Only the variables included in the strongest connections are shown.
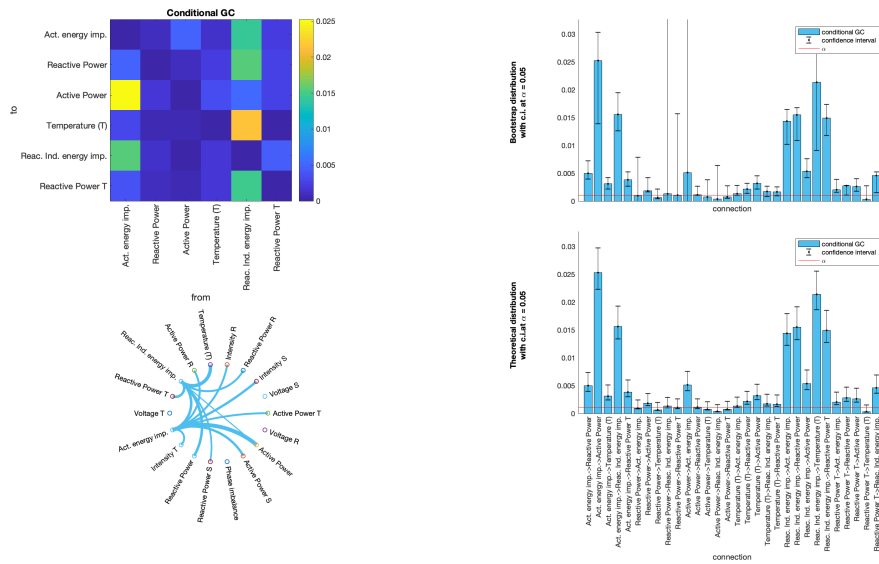

Figure 6: Statistical inference of the conditional G-causality. We show the F matrix for relevant variables, the circular graph for strongest connections and two methods for assessing the significance of the results at a critical significant value equal to $\alpha = 0.05$.

4.2. Which mechanisms or connections between variables do VAR models focus on?

In addition to the individual assessment of the PTs, we analysed the ensemble connectivity maps to highlight which connections were common in separate analyses (throughout 51 periods). Then, we computed the cumulative conditional pair-wise G causalities and generated the circular graphs accordingly. In addition, we calculated the occurrences (on the left) of the corresponding (weak or strong) connections and represented the 5 strongest ones (on the right) by evaluating the connection matrices F in figure 7. As shown in the latter figure, there are persistent weak connections which are removed from the cumulative matrix in the circular graph due to the presence of stronger mechanisms, which suggests the possibility of improving the prediction of some variables by including exogenous information, i.e. Intensity S could help in the prediction of Temperature (T)[3], crucial for detecting PT failures, or the Phase Imbalance (an important variable to be assessed for preventing costumer complaints) is G-caused by Reactive Power R and Active Power T [4]. More important, we find other significant weak connections (in comparison with the aforementioned *trivial* ones), beyond the theoretical thermal and electrical model of the PT, that arise from the same analysis as shown in the circular graphs at the bottom of the same figure. Therefore, this empirical technique revealed the presence of mutual co-occurrences from individual analyses and partially answers the questions raised in the introduction section about the effectiveness of the VAR model to describe the general status of a PT, beyond its instantaneous behaviour. However, this exploratory procedure could be biased due to the presence of noise depending on the particularities of operation in each PT, therefore an ensemble analysis was also mandated to highlight which generative mechanisms, extracted from stationary and conditional analyses were producing the observed data.

---

[3]Indeed, the current explains the PT temperature except for a particular lag that depends on the thermal inertia of the PT

[4]Indeed, phase imbalance is obtained from the analysed power labels R, S, T
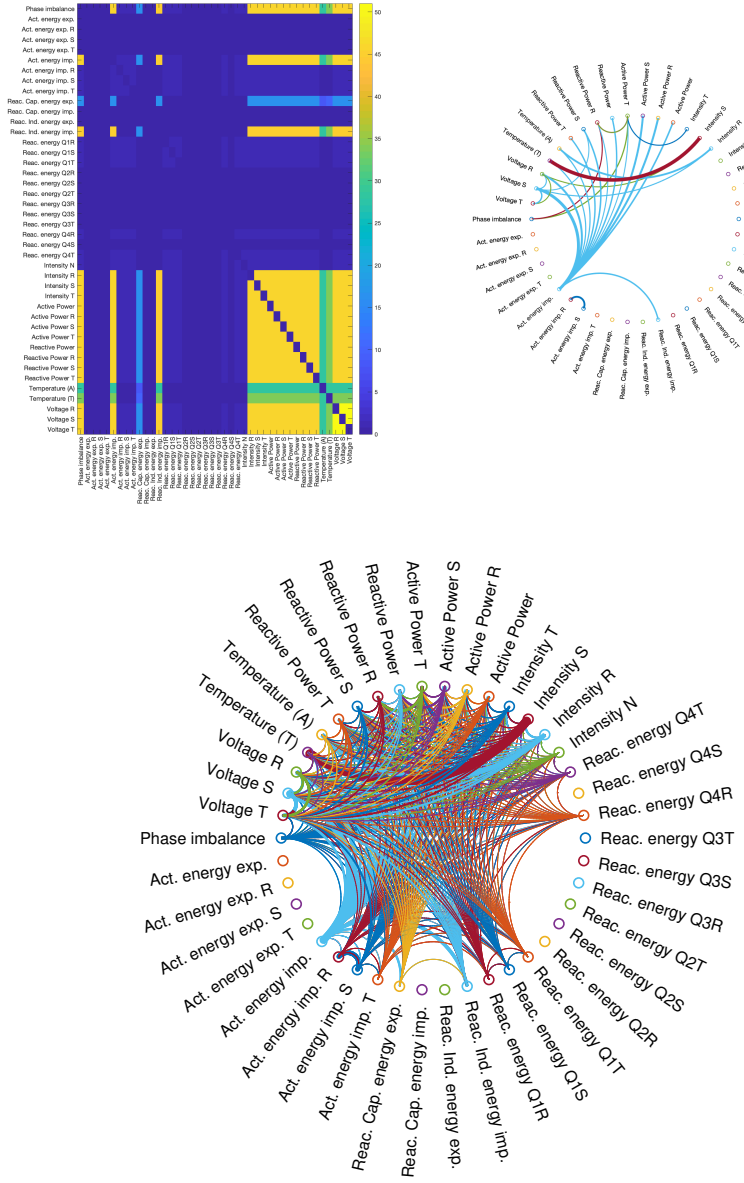
Figure 7: Cumulative conditional G causality matrix with connectivity graphs for all the PTs. We show the number of occurrences in the matrix on the left and the connectivity graph on the right. Observe how, for relevant variables, non-persistent connections such as Intensity S → Temperature T provide the strongest connections in the connectivity matrix. At the bottom, all the connections extracted from the GC-analysis.

14

4.3. The evolution of the PT VAR model in time

The main goal of any control operator is to monitor the operation of the PT system in its normal secure state, as the operating conditions change during daily operation [3]. G-causality allows us to manage the PT by the identification of its operating conditions in terms of the statistically-significant strong connections between relevant variables. Subsequently, necessary preventive actions could be taken in case the system state is found to be abnormal in the analysed periods. In figure 8, the evolution of the significant F components is depicted for PT #1 during 5 periods in 2018. Again, other individual analyses allows us to explain the global operation of the set of PTs in terms of G-connectivity, although it could be slightly biased due to the presence of PT-dependent noise. The complete PT-wise evolution of the entire dataset is depicted in figure 9 as a general circular-graph analysis based on G-causality. Despite that with increasing period number the number of processed variables decreases (see in example period #5 in the bottom of the figure), the remaining variables effectively model the casual structure of the underlying mechanisms, as shown in the bottom figure on the right. This PT-wise analysis in time reveals a set of variables that G-cause other relevant ones, such as Temperature T or Phase Imbalance. Thus, they should be considered for improving time series prediction and control.

4.4. A general status for the PT VAR model

In this section we assume that the variables acquired from different PTs are drawn from the same unknown distribution, thus modelling the same power system. In particular, we consider the period #1 of the whole set and preprocess the data in the same way as the previous sections to avoid typical problems in this scenario such as colinearity, non-stationary, and long term memory (non-vanishing autocorrelation). With all these preprocessing steps in mind and fulfilling the above mentioned conditions, we may extract from this ensemble analysis only general features characterising the set of PTs working in different operating conditions.
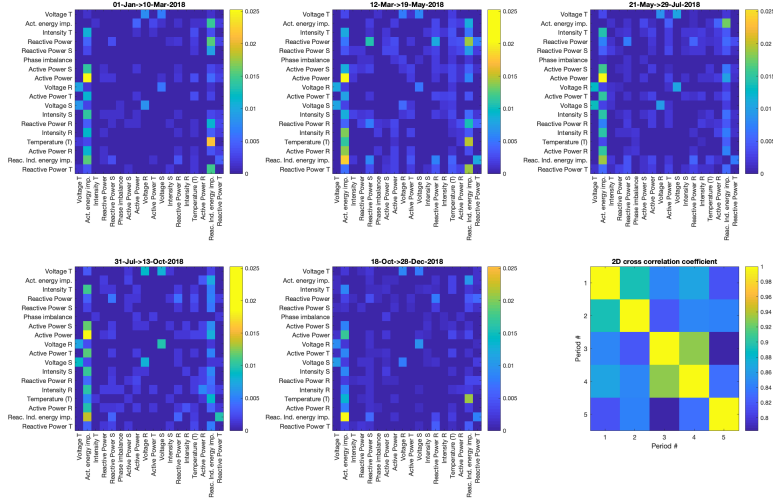
Figure 8: Temporal evolution of the connectivity matrices in 5 operation periods for PT #1. We show the F matrix for relevant variables (F-test at $\alpha = 0.05$), and the 2D-cross correlation coefficient, showing the strong relation between the analyzed periods.

First, we selected the variables fulfilling the aforementioned conditions in the period from January to March in 2018. This resulted in 16 variables out of 42 from the set of SCADA measurements. We acquired 500 observations in more than 600 available trials in that period. Second, following steps of the previous section we derived the time-domain pairwise-conditional causalities and checked the significant test to obtain the results displayed in figure 10. Not surprisingly, a similar analysis in the frequency domain carried out on the averaged DFT coefficients (the f frequency matrix) in the range $[0, f_s/2]$, where $fs = 1/300$ Hz provided almost the same connectivity matrix [10] (see middle column in figure 10). In the last row on the right, we plot the frequency G-causality for the two strongest conditional connections, that is, Reactive Power R and Phase Imbalance, which are found on average in the first period (January-March) of the whole set of transformers, as a general feature characterizing the operation conditions for all PTs.
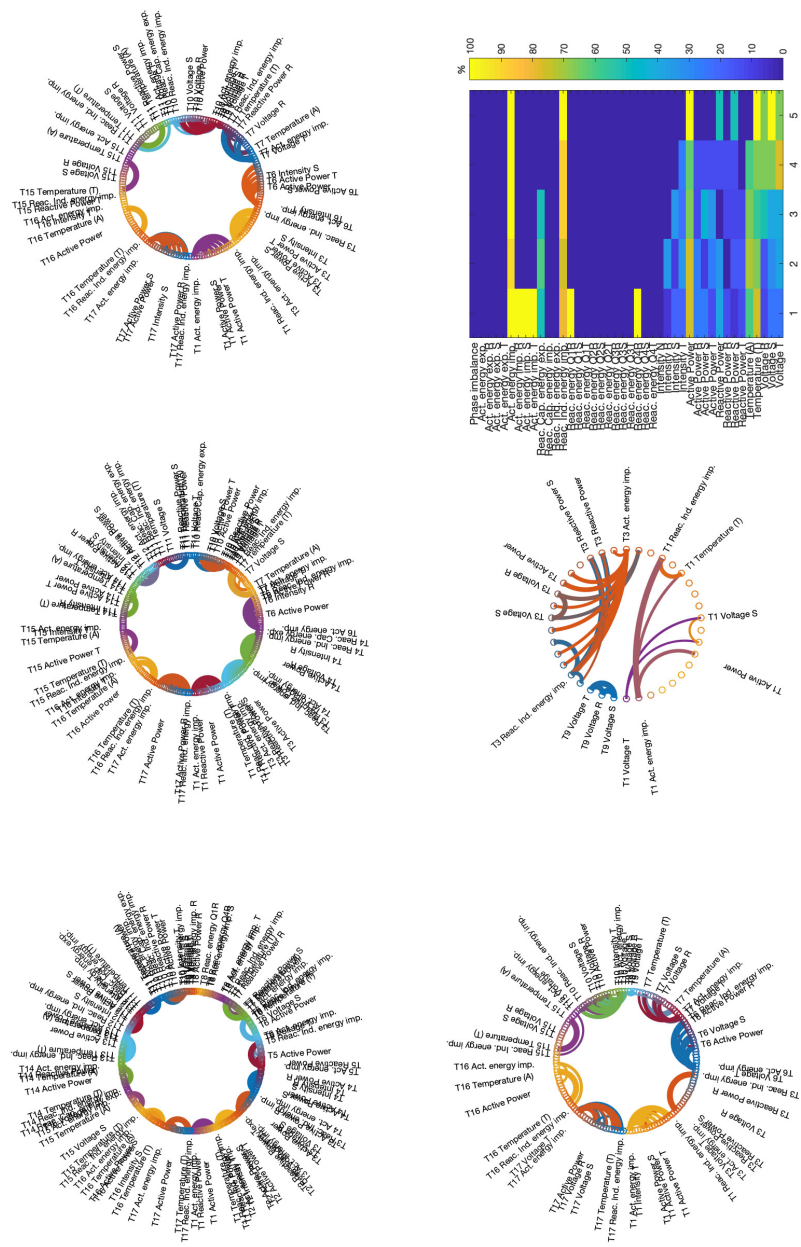
16

Figure 9: Temporal evolution of the connectivity matrices in 5 operation periods. We show the F matrix for relevant variables (F-test at $\alpha = 0.05$), and the % of relevant connections presented in the periods, showing that the casual connections are almost independent of the analysed periods.
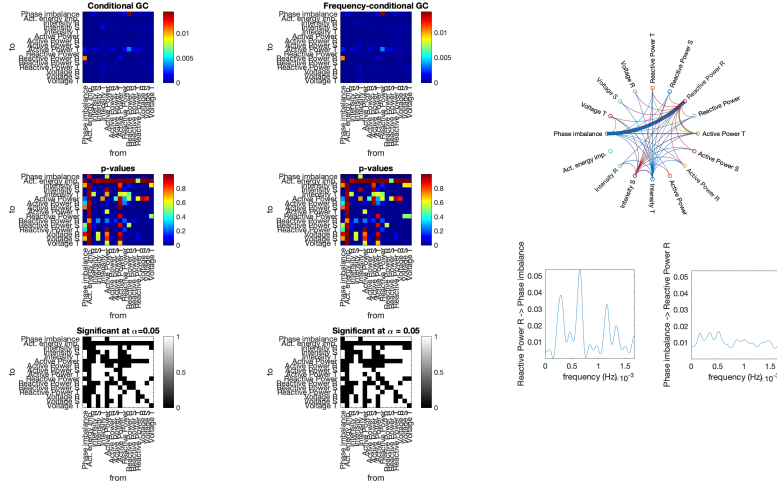
Figure 10: Pairwise G-causality in time and frequency domains and circular graph of connectivity for period #1. We highlight the strongest connections of the relevant variables derived from the bottom figure on the left.

4.5. Preliminary results on time series prediction

The proposed methodology monitors and predicts the status of the power transformer by evaluating linear AR models of stochastic processes. This is automatically completed by considering the resulting Granger-causality models, determined off-line, as an early fusion method within time series prediction systems, similar to those based on non-linear AR neural networks with exogenous inputs (NARX) in a closed-loop configuration [32], or the more general Recurrent Neural Networks (RNN) including Long Short-Term Memory LSTM) Networks [31]. We considered 6 significant variables within the analysed period for $PT\#1$ (as shown in Figure 5) as inputs to a two-delay element NARX network consisting of a 10-neuron single MLP-based hidden layer. First, we fitted the whole system using 18k data samples and cross validation (70% for training, 15% for validation and 15% for testing). Then, the remaining 2k samples in the selected period were used to check the ability of the closed-loop network configuration for forecasting unseen new data; see figure 11. This kind of simulation
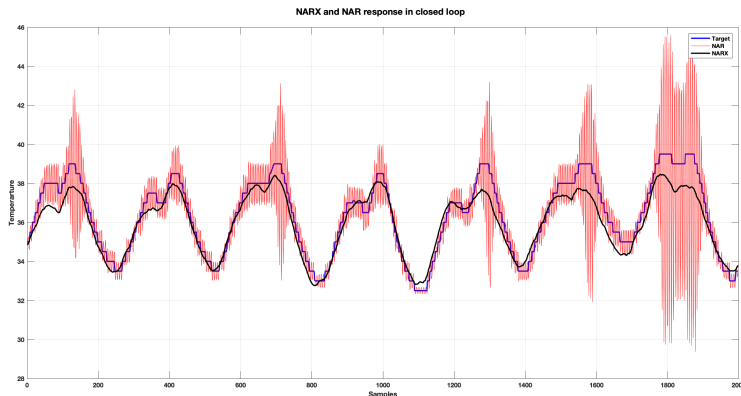
18

Figure 11: Closed-Loop performance of a NARX-based forecasting system and comparison with the same configuration of a NAR network with a single input (Temperature). Note the improvement of the proposed configuration, in terms of error between target and prediction, when exogenous times series are selected by the GC model.

provides an effective way of evaluating the relevance of the G-causal notion in terms of prediction ability, since in a closed-loop configuration the system predicts the target signal without using its previous values, rather with delayed exogenous variables in combination with the previous output value. Moreover, these recurrent architectures allow us to process sequences of inputs at a given time, thus making it possible to discover trends towards anomalous temperature values.

5. Conclusion

Beyond single voltage measurements and isolated case studies for back-office monitoring and predicting failures of conspicuous components, e.g. bushings in PTs [1], we propose a general methodology based on G-causality to explore the conditional correlations between variables which could be integrated in comprehensive monitoring systems. Inspection of the proposed circular graph-based analyses reveals the generative mechanisms underlying the observed data. The proposed analysis is two-fold: a static analysis considering the set of PTs and

19

then averaging; and a time-dependent analysis describing the evolution of the connectivity matrix in time and assessing the ratio of the relevant/processed variables for all the evaluated periods. All these analyses were conducted under a strongly-grounded statistical framework to check for significance at a given level of confidence.

The big differences found between ensemble and PT-wise analysis invites us to improve the acquisition process to obtain *stable* periods including WSS processes with the same number of variables. Although the results have good confidence from a statistical point of view, this clearly affects the conclusions of the present work since there are missing variables in the last acquisition periods which could be reporting spurious causalities where there is no direct causal influence. Moreover, the analysis includes some shifts in the operating temporal window due to the presence of acquisition errors or almost constant measurements, which are removed from the analysis at the differencing step, i.e. $x(t) = x(t) - x(t-1)$, for obtaining covariance stationary variables.

Key to the proposed methodology is the use of SCADA measurements, employed throughout the paper and previously acquired within the Pastora and Monica projects [2]. This study focused on the modelling of power substations, including PT and the power elements (i.e. LV circuits) to which they are connected. We analysed up to 42 electrical variables that are claimed to accurately model PT operation. These variables were measured at the input and output of the PT, and it could be argued that other circuit elements should be incorporated into the model. However, this is generally not the case in the extant literature. Recalling that G-causality does not correspond with physical notions of causality [18], only pair-wise effects between input and output variables were considered, whilst controlling for all other variables. Clearly, this is not a novel concept in signal processing and modelling; indeed, there are several examples in different research fields where this approach provides additional information about the underlying processes, including: spectral smoothing, functional interpolation, speech synthesis, and information-based approaches. This exploratory analysis moves beyond the physical notion of causality determining the opera-

20

tion of the PT by examining the forward and backward variables (i.e. prediction errors) and their G-connections or mechanisms, ensuring that they satisfy the assumptions underpinning the inference method.

350   Finally, it is worth mentioning that the connectivity measure in terms of the LRTs, as defined herein, may be interpreted as a predictive model. In fact, the generalised variance of a regression model in equations 4 and 6 may be viewed as a quantification of the reduction in the prediction error, when the past of the process U_2 is included in the explanatory variables of a VAR model for U_1.

355   Thus, whenever a strong G-causality mechanism is found, a reduction in the error of the prediction model is achieved. This could be useful when considering complex prediction models and the set of variables to be included, as exogeneous variables, to predict the target or desired signal.

    Once these exploratory analyses have been conducted on this *moderate-sized*

360   dataset, future investigations will investigate the problem of pattern recognition. We will collect larger datasets to avoid the typical limitations of the papers in this field, that is, the use of insufficient test data in cross-validation experiments causing over-fitting and worse generalization performance on the holdout sample set. In this way, we plan to apply this novel methodology to publicly available

365   datasets to boost the performance of well-known, simple and representative statistical learning algorithms [8].

Gómez Expósito for his helpful advice and comments.

References

[1] Bartlomiej Dolata, Sebastian Coenen. Online Condition Monitoring Becomes Standard Configuration of Transformers - Practical Application for Optimized Operation, Maintenance and to Avoid Failures. e-ARWtr2016 transformers. Advanced Research Workshop on Transformers (2)3 -5 October 2016. La Toja Island– Spain

[2] https://www.endesadistribucion.es/en/innovacion-nuevas-tecnologias/proyecto-monica-red-distribucion.html.

[3] Ali Abur, Antonio Gómez Expósito. Power System State Estimation. Theory and Implementation. 1st Edition. 2004 CRC Press. ISBN 9780203913673

[4] A Gomez-Exposito, A Abur, A de la Villa Jaen, C Gomez-Quiles. A multi-level state estimation paradigm for smart grids Proceedings of the IEEE 99 (6), 952-976. 2011.

[5] P Zarco, AG Exposito. Power system parameter estimation: a survey IEEE Transactions on power systems 15 (1), 216-222. 2000.

[6] R. M. A. Velásquez and J. V. M. Lara, "Expert system for power transformer diagnosis," 2017 IEEE XXIV International Conference on Electronics, Electrical Engineering and Computing (INTERCON), Cusco, 2017, pp. 1-4.

[7] A. Peimankar, S.J. Weddell,T. Jalal and A.C. Lapthorn Evolutionary multi-objective fault diagnosis of power transformers. Swarm and Evolutionary Computation Volume 36, October 2017, Pages 62-75.

[8] Piotr Mirowski,Yann Lecun. Statistical Machine Learning and Dissolved Gas Analysis: A Review October 2012. IEEE Transactions on Power Delivery 27(4) DOI: 10.1109/TPWRD.2012.2197868

[9] Granger, C.W.J., 1969. Investigating causal relations by econometric models and cross-spectral methods. Econometrica 37, 424–438.

[10] Geweke, J., 1982. Measurement of linear dependence and feedback between multiple time series. J. Am. Stat. Assoc. 77, 304–313.

[11] Geweke, J., 1984. Measures of conditional linear dependence and feedback between time series. J. Am. Stat. Assoc. 79, 907–915.

[12] Bressler, S., Seth, A., 2011. Wiener-Granger causality: A well established methodology. Neuroimage 58, 323–329.

[13] John R. Freeman Granger Causality and the Times Series Analysis of Political Relationships Vol. 27, No. 2 (May, 1983), pp. 327-358 American Journal of Political Science.

[14] Patrick A. Stokes and Patrick L. Purdon. A study of problems encountered in Granger causality analysis from a neuroscience perspective. PNAS August 22, 2017 114 (34) E7063-E7072; https://doi.org/10.1073/pnas.1

[15] Wiener, N., 1956. The theory of prediction, in: Beckenbach, E.F. (Ed.), Modern Mathematics for Engineers. McGraw Hill, New York, pp. 165–190.

[16] Valdes-Sosa, P.A., Roebroeck, A., Daunizeau, J., Friston, K., 2011. Effective connectivity: Influence, causality and biophysical modeling. Neuroimage 58, 339–361.

[17] Aertsen, A., Preißl, H., 1991. Dynamics of activity and connectivity in physiological neuronal networks, in: Schuster, H. (Ed.), Nonlinear Dynamics and Neuronal Networks. VCH Publishers Inc., New York, pp. 281–302.

[18] Friston, K.J., Harrison, L., Penny, W., 2003. Dynamic causal modelling. Neuroimage 19, 1273–1302.

[19] Pearl, J., 2009. Causality: Models, Reasoning and Inference (2nd ed.). Cambridge University Press, Princeton, New York.

[20] Friston, K.J., 2011. Functional and effective connectivity: A review. Brain Connectivity 1, 13–36.

[21] Roebroeck, A., Formisano, E., Goebel, R., 2009. The identification of interacting networks in the brain using fMRI: Model selection, causality and deconvolution. Neuroimage .

[22] Roebroeck, A., Seth, A., Valdes-Sosa, P., 2010. Causal time series analysis of functional magnetic resonance imaging data. Journal of Machine Learning Research 12, 65–94.

[23] Anil K. Seth. A MATLAB toolbox for Granger causal connectivity analysis. Journal of Neuroscience Methods 186 (2010) 262-273.

[24] Adam B. Barrett, Lionell Barrett, Anil K. Seth. Multivariate Granger causality and generalized variance Physical Letter Review E 81, 041907.2010

[25] Edwards, A.W.F., 1992. Likelihood (Expanded Edition). Johns Hopkins University Press, Baltimore

[26] Juan Manuel Górriz Sáez, Carlos García Puntonet, Moisés Salmerón, Juan José González de la Rosa. A new model for time-series forecasting using radial basis functions and exogenous data. Neural Computing and Applications 13(2): 101-111 (2004)

[27] Ding M, Chen Y, Bressler S. Granger causality: basic theory and application to neuroscience. In: Schelter S, Winterhalder M, Timmer J, editors. Handbook of time series analysis. Wienheim:Wiley; 2006. p. 438–60.

[28] Lionel Barnett and Anil K.Seth. The MVGC multivariate Granger causality toolbox: A new approach to Granger-causal inference. Journal of Neuroscience Methods Volume 223, 15 February 2014, Pages 50-68.

[29] B. Efron Bootstrap Methods: Another Look at the Jackknife Ann. Statist. Volume 7, Number 1 (1979), 1-26.

[30] Y. Benjamini and D. Yekutieli, "The control of the false discovery rate in multiple testing under dependency", Ann. Stat, 29(4), 2001.

[31] F.J. Martinez, Javier Ramirez, Fermin Segovia, Andres Ortiz, Susana Carrillo, Javier Leiva, Jacob Rodriguez-Rivero and Juan Manuel Gorriz Prediction of Transformer Temperature for Energy Distribution Smart Grids Using Recursive Neural Networks. International Conference on Time Series and Forecasting. Granada, Spain. I.S.B.N: 978-84-17970-79-6.

[32] Javier Ramirez, Francisco J. Martinez Murcia, Fermin Segovia, Susana Carrillo, Javier Leiva, Jacob Rodriguez-Rivero and Juan M. Gorriz. Power transformer monitoring based on a non-linear autoregressive neural network model with exogenous inputs. International Conference on Time Series and Forecasting. Granada, Spain. I.S.B.N: 978-84-17970-79-6.