Multivariate Autoregressive Model with Instantaneous Effects to Improve Brain Connectivity Estimation

Silvia Erla\textsuperscript{ab}, Luca Faes\textsuperscript{b}, Enzo Tranquillini\textsuperscript{b}, Daniele Orrico\textsuperscript{c}, Giandomenico Nollo\textsuperscript{b}

\textsuperscript{a}CIMEC, University of Trento, Trento, Italy
\textsuperscript{b}Dept. of Physics, University of Trento, Trento, Italy
\textsuperscript{c}Dept. of Neurological Pathology, S. Chiara Hospital, Trento, Italy

Correspondence: Silvia Erla, University of Trento, Via delle Regole 101, 38060 Mattarello, Trento, Italy.
E-mail: silvia.erla@email.unitn.it, phone +39 0461 882771, fax +39 0461 883659

Abstract. Evaluation of brain connectivity in the frequency domain is based on prior multivariate autoregressive (MVAR) model identification from multichannel neurological time series. The MVAR model commonly used in neuroscience applications accounts only for lagged effects among the time series and forsakes instantaneous effects. However, zero-lag interactions are likely to occur among simultaneously recorded neural signals, and the impact of their exclusion on connectivity measures has not been investigated yet. In this study we propose the use of an extended MVAR model including instantaneous effects, and compare its performance to that of the traditional MVAR approach using the Partial Directed Coherence (PDC). We show by simulations that, in presence of zero-lag correlations, the PDC derived from traditional MVAR modeling may produce misleading frequency domain connectivity evaluation, and that in such situations the correct connectivity pattern is recovered using the extended MVAR model. Then we provide examples of multichannel EEG recordings in which instantaneous effects are found to be far from negligible, and thus extended MVAR modeling seems more suitable to elucidate direction and strength of the interactions among EEG rhythms.

Keywords: Frequency Domain, Granger Causality, Model Identification, Multichannel EEG, Multivariate Autoregressive Models, Partial Directed Coherence, Zero-lag Correlation

1. Introduction

The assessment of brain connectivity is one of the major issues that are faced nowadays in the field of experimental neuroscience [Horwitz, 2003]. This fact is attested, from the signal processing point of view, by the recent development of methods aimed at the estimation of the causal influences among different brain areas in humans [Baccalà and Sameshima, 2001; Kaminiski et al., 2001; Astolfi et al., 2006]. Among them, the partial directed coherence (PDC) [Baccalà and Sameshima, 2001] is emerging as one of the most widely used tools to estimate brain connectivity, thanks to the fact that it provides a direct frequency domain description of the concept of Granger causality [Granger, 1969].

As well as with many other causality measures, the evaluation of PDC relies on fitting the available multichannel data set with a multivariate autoregressive (MVAR) model that describes, for each time series, the linear contribution coming from its past samples and the past samples of all other time series. The MVAR model implemented in all neuroscience applications is strictly causal, in the sense that only lagged effects are modeled and instantaneous (i.e., not lagged) effects among the time series are not described by any model coefficients. Neglecting instantaneous effects in MVAR models implies that any zero-lag correlation among the time series is translated into a correlation among the model residuals [Korhonen et al., 1996]. This prevents the use of tools such as the coherence or the directed coherence requiring uncorrelation of the model residuals to be accessible. Moreover, in presence of significant instantaneous correlations the impact of assessing causality by explicitly forsaking the cross-covariance matrix of the model residuals, as done in PDC evaluation, is at present unknown.

In the present study, the effects of excluding instantaneous effects from MVAR models on PDC evaluation are investigated. Furthermore, we propose the utilization of an extended MVAR model combining instantaneous and lagged effects to improve the frequency domain evaluation of Granger
causality. The PDC derived from traditional and extended MVAR models are compared using simulations reproducing known conditions of interactions even at zero lag, as well as examples of multichannel EEG recorded from normal subjects in the resting eyes closed condition.

2. Methods

1.1. Multivariate autoregressive model without instantaneous effects

The set of $M$ simultaneously observed zero-mean time series $X(n)=[x_1(n),...,x_M(n)]^T$ is assumed to be adequately described by the MVAR model

$$X(n) = \sum_{k=1}^{p} A(k) X(n-k) + U(n),$$

(1)

where $p$ is the model order, $A(k), k=1,\ldots,p$, are $M \times M$ matrices containing the elements $A_{ij}(k)$ that describe the linear interaction at lag $k$ from $x_j(n-k)$ to $x_i(n)$ ($i,j=1,\ldots,M$), and $U(n)=[u_1(n),...,u_M(n)]^T$ is a vector of zero-mean white noise processes with covariance matrix $\Sigma_U$. Note that, with this strictly causal model representation ($k>0$), zero-delay cross-correlations among the observed series $x_i$ cannot be described by the model coefficients and are thus explained by correlations among the input noises $u_i, k=1,\ldots,M$; as a result, the noise covariance $\Sigma_U$ is generally not diagonal.

The MVAR model (1) can be interpreted in terms of Granger’s definition of causality [Granger, 1969] which states that $x_i$ causes $x_j$ if the knowledge of $x_i(n-1), x_i(n-2),\ldots, x_i(n-p)$ leads to a better prediction of $x_j(n)$. In terms of MVAR coefficients, $x_i$ causes $x_j$ if at least one off-diagonal element $A_{ij}(k)$ of the matrices $A(k)$ is significantly different from zero. Note that, with this causality definition, only the past of a time series is considered to influence another series and instantaneous effects are not accounted.

1.2. Multivariate autoregressive model with instantaneous effects

The MVAR model (1) can be extended to also consider instantaneous effects from $x_j$ to $x_i$ as a possible source of interaction by introducing the model

$$X(n) = \sum_{k=0}^{p} B(k) X(n-k) + W(n),$$

(2)

in which instantaneous effects are included in the form of the nonzero matrix $B(0)$. In (2), the input noise $W(n)=[w_1(n),...,w_M(n)]^T$ is a vector of zero-mean uncorrelated processes with diagonal covariance matrix $\Sigma_W$. Here uncorrelation among the noise inputs is assured by the presence of the instantaneous effects; indeed, it can be shown that the model (1) is transformed into model (2) by left-multiplying (1) with the matrix $L^{-1}=I-B(0)$, where $I$ is the $M \times M$ identity matrix and $L$ results from the Cholesky decomposition [Kay, 1988] of the noise covariance matrix $\Sigma_W$: $\Sigma_W L L^T$. In this way we have $B(k)=L^{-1}A(k), k=1,\ldots,p$, and the noise covariance of $W(n)=L^{-1}U(n)$ becomes the diagonal matrix $\Sigma_W=A$. As a result of the Cholesky decomposition $L$ is a lower triangular matrix, and thus $B(0)$ becomes lower triangular with null diagonal. To fulfil this constraint in practical applications, the observed time series within the vector $X$ have to be ordered in a way such that instantaneous effects are allowed from $x_j$ to $x_i$ ($B_{ij}(0)\neq 0$ for each $j<i$) but not from $x_i$ to $x_j$ ($B_{ji}(0)=0$ for each $j<i$).

According to Granger’s definition of causality [Granger, 1969], we can state that $x_i$ causes $x_j$ when at least one off-diagonal element of the matrices $B(k)$ is significantly different from zero: $B_{ij}(k)\neq 0, k=1,\ldots,p$. Since the Granger definition considers only past samples when evaluating the prediction improvements, we omit $B(0)$ from the check. Nevertheless, we stress that the presence of instantaneous effects ($B(0)\neq 0$) makes all $B(k)$ different from $A(k)$, even with $k>0$, and thus different Granger causality patterns may be found depending on whether instantaneous effects are included or not into the MVAR model fitted to the available data.

1.3. Partial Directed Coherence from MVAR models without and with instantaneous effects

Partial Directed Coherence (PDC) was introduced in [Baccalà and Sameshima, 2001] as a frequency domain descriptor of the directed linear relation between pairs of time series $x_i(n)$ and $x_j(n)$, when observed in conjunction with a set of other time series. PDC is traditionally computed from the MVAR model without instantaneous effects (1), leading to the definition:
\[
\beta_{ij}(f) = \frac{A_{ij}(f)}{\sum_{m=1}^{M} |A_{mj}(f)|^2},
\]
where \( f \) is the normalized frequency in the range \([0, 0.5]\) and
\[
A_{ij}(f) = \delta_{ij} - \sum_{k=1}^{P} A_{ij}(k)e^{-j2\pi f k},
\]
for \( \delta_{ij} = 0 \) if \( i \neq j \) and \( \delta_{ij} = 1 \) if \( i = j \) and \( j = \sqrt{-1} \), is the \((i,j)\) element of the inverse transfer matrix of the MVAR model in the frequency domain. The PDC offers a frequency domain representation of the concept of Granger causality: indeed we can state that, with \( i \neq j \), \( x_j \) causes \( x_i \) at frequency \( f \) when \( A_{ij}(f) \), and thus \( \beta_{ij}(f) \), differs significantly from zero.

By analogy with the above procedure, we now provide a definition for the PDC based on the model with instantaneous effects formulated in (2). First, we note that the \((i,j)\) element of the inverse transfer matrix of the MVAR model (2) in the frequency domain becomes:
\[
\tilde{B}_{ij}(f) = \delta_{ij} - B_{ij}(0) - \sum_{k=1}^{P} B_{ij}(k)e^{-j2\pi f k}.
\]
Since Granger causality accounts only for lagged effects, we exclude \( B_{ij}(0) \) from the following computations setting \( \tilde{B}_{ij}(f) = \tilde{B}_{ij}(f) + B_{ij}(0) \). Then we define a new PDC function, thereinafter denoted as iPDC to indicate that it is derived from the MVAR model with instantaneous effects:
\[
\tilde{\beta}_{ij}(f) = \frac{1}{\lambda_i} \frac{\tilde{B}_{ij}(f)}{\sum_{m=1}^{M} \frac{1}{\lambda_m^2} |\tilde{B}_{mj}(f)|^2},
\]
where \( \lambda_i^2 \) is the variance of noise \( w_i \) in (2). Besides of the extension to instantaneous effects, the iPDC function defined in (6) includes also the factors \( 1/\lambda_i \) weighting each spectral element \( \tilde{B}_{ij}(f) \). This is in agreement with the generalization recently suggested by the proposers of PDC, which leads to a variance stabilization of the frequency domain representation of Granger causality via PDC [Baccalà et al., 2007]. Nevertheless, we remark that the iPDC defined in (6) is different from the generalized PDC proposed in [Baccalà et al., 2007], as the coefficients \( B_{ij}(k) \) are different from \( A_{ij}(k) \) whenever significant instantaneous effects are present among the investigated time series.

3. Simulations

In this section we compare the ability of PDC and iPDC to describe causal connectivity patterns imposed on simulated time series with both instantaneous and lagged interactions. We remark that in absence of instantaneous effects \( B(0)=0 \) the MVAR model (2) reduces to model (1), and thus PDC and iPDC are equivalent connectivity descriptors. Here we consider the MVAR process \((M=5\) channels, order \( p=3\)) with the connectivity pattern shown in Fig. 1a and described by the following equations:

\[
\begin{align*}
x_1(n) &= 1.58x_1(n-1) - 0.81x_1(n-2) + w_1(n) \\
x_2(n) &= 0.9x_1(n) - 0.01x_2(n-2) - 0.6x_3(n-3) + w_2(n) \\
x_3(n) &= 0.3x_2(n-1) + 0.8x_3(n-1) + 0.3x_2(n-2) - 0.25x_3(n-2) + 0.3x_2(n-3) + w_3(n), \\
x_4(n) &= 0.9x_3(n) - 0.6x_2(n-1) + 0.3x_2(n-3) + w_4(n) \\
x_5(n) &= -0.3x_4(n-1) + 0.9x_1(n-2) - 0.3x_4(n-2) + 0.6x_3(n-3) + w_5(n)
\end{align*}
\]

where the input processes \( w_1, \ldots, w_5 \) are considered as uncorrelated white noise processes of unit variance \( \left(\Sigma_w=\Lambda=I\right) \). The coefficients of the corresponding MVAR process without instantaneous effects can be obtained from those of the process (7) as: \( A(k)=L B(k), k=1,\ldots,3 \) (with \( L=\left[I-B(0)\right]^{-1} \)), while its nondiagonal noise covariance matrix is: \( \Sigma_u=LA\Lambda^T \).
The squared PDC and iPDC functions obtained for the two models taking the squared modulus of Eqs. (3) and (6) are shown in Fig. 1. It is evident that the inclusion of instantaneous effects into the MVAR model improves the evaluation of directional lagged relationships between pairs of time series, and thus the detection of Granger causality in the frequency domain. While along some arcs the two functions overlap, analysis of the plots, in which PDC and iPDC display different trends, evidences the better performance of the new approach. In fact, PDC fails to assess some of the imposed directional influences. For instance, the nonzero PDC values from $x_1$ to $x_2$ lead to the wrong interpretation that there are lagged effects, while they are actually instantaneous. Moreover, PDC suggest the existence of a direct connection from $x_1$ to $x_4$ that is actually not present (neither instantaneous nor lagged). These wrong interpretations induce also an underestimation of the expected PDC from $x_1$ to $x_5$, which remains low as a consequence of the normalization condition. On the contrary, the iPDC always provides a correct representation of the connectivity pattern: it is indeed uniformly zero from $x_1$ to $x_2$ and from $x_1$ to $x_4$, and reaches unity from $x_1$ to $x_5$ correctly indicating that all the lagged influences originating from $x_1$ at the frequency of its main spectral peak are directed towards $x_5$. Another disagreement emerges considering the link between $x_4$ and $x_5$, that is erroneously regarded as closed loop by PDC and correctly regarded as unidirectional lagged interaction from $x_4$ to $x_5$ using iPDC.

**Figure 1.** (a) Connectivity pattern imposed for the simulated multivariate process, with instantaneous and delayed interactions represented by dashed and solid arrows, respectively. (b) Matrix layout plots of the spectral functions. Diagonal plots ($i=j$): power spectrum of $x_i$; off-diagonal plots: $|\tilde{\chi}_{ij}(f)|^2$ (black solid) and $|\chi_{ij}(f)|^2$ (red dashed).

**4. EEG application**

EEG recordings were acquired (10-20 System, 256 Hz sampling rate, Fpz common reference) from ten healthy subjects (22-35 yrs) in the relaxed awake state with eyes closed. For each subject, the 19 acquired unipolar signals were bandpass filtered (FFT filter, 0.3-40 Hz), cleaned up of ocular and ECG artifacts by means of Independent Component Analysis where necessary, and then downsampled to 128 Hz. After selection of 8 s stationary windows, five bipolar signals, representative of different cortical areas ($x_1$: occipital; $x_2$: left temporal; $x_3$: right temporal; $x_4$: central; $x_5$: frontal) were obtained as the difference between pairs of adjacent unipolar signals (respectively: O1-O2; T3-T5; T4-T6; Fz-Cz; Fp1-Fp2). Starting from these signals ($M=5$ series, $N=1024$ samples), a strictly causal MVAR model (Eq. (1)) of order $p=5$ was identified, and the corresponding model with instantaneous effects was estimated as described in Sect. 1.2. In all subjects, the presence of non negligible zero-lag interactions among bipolar EEG signals was documented by the significant correlations found for the residuals in model (1), i.e. $\Sigma_U$ was not diagonal.

Fig. 2 shows a representative example of the spectral functions obtained for one of the subjects. Considering the alpha frequency band (8-13 Hz), the dominant connectivity direction is from back to front; both PDC and iPDC appear indeed higher in the directions from $x_j$ to $x_i$ than from $x_i$ to $x_j$, with $j<i$. However, iPDC seems to highlight the information flow better than PDC, as documented by its higher
values in some of the back-to-front connections (e.g., $x_1 \rightarrow x_5$, $x_2 \rightarrow x_5$, $x_3 \rightarrow x_5$, $x_4 \rightarrow x_5$), and lower values in some front-to-back directions (e.g., $x_2 \rightarrow x_1$, $x_3 \rightarrow x_1$). This finding was confirmed in the statistical analysis of all ten considered subjects. The analysis, summarized in Fig. 3, shows the distribution over subjects of the PDC and iPDC values averaged in the alpha band of the frequency spectrum. The results indicate a clear predominance of the back-to-front propagation using both PDC and iPDC, as can be seen from the connectivity pattern in Fig. 3b. Moreover, the squared modulus of the iPDC function resulted in statistically significantly higher values than the traditional PDC function in almost all the back-to-front causal directions (Fig. 3a). Hence, the iPDC function seems a better descriptor than the PDC of the expected neurophysiological behavior, i.e. the propagation of the alpha rhythm from the occipital to the frontal brain areas [Kaminski et al., 1995].

**Figure 2.** Matrix layout plots of the spectral functions obtained from bipolar EEG recordings of a healthy subject. Diagonal plots ($i=j$): power spectrum of $x_i$; off-diagonal plots: $|\pi_i(f)|^2$ (black solid) and $|\tilde{\chi}(f)|^2$ (red dashed). Vertical lines delimit the alpha band of the frequency spectrum.

**Figure 3.** (a) Distributions of the squared PDC from $x_i$ to $x_j$ ($|\pi_i(f)|^2$, circles) and the squared iPDC from $x_i$ to $x_j$ ($|\tilde{\chi}(f)|^2$, triangles), expressed as mean ± standard error over the 10 considered subjects, evaluated in the alpha band (mean values in the 8-13 Hz frequency range) for each interconnection $ij$; * $p<0.05$ PDC vs. iPDC, Student t-test for paired data. (b) Average connectivity pattern over the 10 subjects obtained from the PDC and iPDC values in (a) according to arrow thickness (0-0.1: no arrow; 0.1-0.2: dotted arrow; 0.2-0.3: solid arrow; >0.3: solid thick arrow).
5. Conclusions

In this study we presented an extension of the MVAR modeling approach that also accounts for instantaneous effects, which led to a new formulation of PDC, here denoted as iPDC. The new approach is based on the fact that introducing coefficients which describe instantaneous effects into a MVAR model also changes the coefficients describing the lagged effects. Hence, different inferences about Granger causality in the frequency domain are made using PDC and iPDC. Simulations showed that iPDC goes a step beyond the traditional PDC thanks to its ability to disentangle lagged effects from a model including also instantaneous effects. In particular, we found that neglecting instantaneous effects may lead to wrong estimates of the autoregressive coefficients, and thus of PDC values, if the investigated multichannel data set exhibits zero-lag cross correlations. We observed that this is the case in a small EEG database, showing also that iPDC values describe the back-to-front propagation of the alpha activity better than PDC. Thus, although it needs further investigation in real data applications, the proposed refinement seems useful to improve the reliability of frequency domain connectivity identification.

References


