Biol Cybern (2006) 95:135–141
DOI 10.1007/s00422-006-0075-7

ORIGINAL PAPER

Biological
Cybernetics

Comments on 'Is partial coherence a viable technique for identifying generators of neural oscillations?'

Why the term 'Gersch Causality' is inappropriate: Common neural structure inference pitfalls.

Luiz A. Baccalá · Koichi Sameshima

Received: 28 March 2005 / Accepted: 20 March 2006 / Published online: 20 May 2006

Abstract To aid prospective neural connectivity inference analysts and hoping to preclude misconception spread, we exploit the didatic value of some of the issues raised by Albo et al. (Biol Cybern 90: 318–326, 2004) who claim that signal-to-noise ratio (SNR) values can lead to mistakes in structural inference when using partial coherence in connection to Gersch's 1970 method for spotting signal sources (Gersch in Math Biosci 14: 177– 196, 1972). We show theoretically that Gersch's method is able only to spot which measurement of some common underlying factor has the least amount of additive noise and that this has nothing to do with any reasonable notion of 'causality' as suggested by Albo et al. (Biol Cybern 90: 318–326, 2004). We also show that despite the inherent structural ambiguity of the model used by Albo et al. (Biol Cybern 90: 318–326, 2004) to back their claim, its data can nonetheless furnish the correct time precedence hierarchy between the activities in its measured structures, both when simple (correlation) and more sophisticated methods are used (partial directed coherence) (Baccala and Sameshima in Biol Cybern 84:463–474, 2001a) in a true depiction of time series causality.

L. A. Baccalá (✉)
Telecommunications and Control Engineering Department, Escola Politécnica, University of São Paulo, Av. Prof. Luciano Gualberto, Trav. 3, 158, São Paulo, SP 05508-900, Brazil
e-mail: baccala@lcs.poli.usp.br

K. Sameshima
Radiology Department, Medical School, University of São Paulo, Av. Dr. Arnaldo 455, São Paulo, SP 01246-903, Brazil

1 Introduction

Reading Albo et al. (2004) brings to mind a number of recent issues as the authors examine the use of partial coherence in the inference of connectivity from a notion put forward by Gersch (1972) in the early 1970s. The central idea is that partial coherence computation allows categorizing the relative roles played by neural structures when considered in triplets. This follows immediately from the fact that the partial coherence (PC) between a pair of time series describes their mutual linear relationship after subtracting the effect of the remaining member of the triplet. Thus, logically one would pinpoint the third structure as responsible for the joint activity of the pair before subtraction. This communality was termed 'Gersch causality' by Albo et al. (2004) (the sole reference to this in the whole of ISI's database to date) as opposed to the prior literature which, more correctly in our view, refers to this as 'Gersch' driving.

To investigate PC's faithfulness in connectivity inference, Albo et al. (2004) performed the numerical analysis of data from a model of nondispersive signal propagation subject to additive observation noise. The main conclusion was that inference accuracy using Gersch's method depends on the signal-to-(additive)noise ratios of the measured variables.

Our aims in this letter are: (1) to spell the theoretical rather than simulation based reason why partial coherence and Gersch's method fail in the case of the propagation model used in Albo et al. (2004); (2) to explain why it has worked elsewhere Gersch and Goddard (1970); (3) to argue for the avoidance of the term causality in connection to Gersch's proposal and (4) to provide the discussion of simple connectivity inference alternatives like partial directed coherence (Baccala and

Sameshima 2001a; and in Albo et al.'s 2004 simple case even correlation analysis).

We use basic linear estimation properties to prove that the phenomenon in Albo et al. (2004) results from the fact that the time series subject to the least amount of added noise allows the largest reduction in the partial coherence between the other series simply because it is the best estimate of the communality shared by the other time series (Sect. 2).

This property holds regardless of the time precedence between the time series which is the basis of any reasonable definition of causality as we point out in Sect. 3. In fact, we show that even though the model in Albo et al. (2004) is structurally ambiguous (Sect. 4), its time precedence structure can be captured correctly by methods like partial directed coherence or even by ordinary correlation; this is illustrated in (Sect. 5) by analysing simulated data. Our conclusions are collected in Sect. 6.

2 Basic linear estimation results

In a nutshell, the model in Albo et al. (2004) is composed of three delayed and noise corrupted observations of the same underlying time series. We can start by examining any two series under the premise that they share a common factor $z(t)$ subject to possible delays:

$$v(t) = z(t - \tau_v) + n_v(t), \tag{1}$$

$$w(t) = z(t - \tau_w) + n_w(t), \tag{2}$$

where $z(t), n_v(t)$ and $n_w(t)$ are zero-mean mutually independent time series, respectively, with power spectral densities $S_z(f), S_{n_v}(f) = S_{n_v}$ and $S_{n_w}(f) = S_{n_w}$ where we assumed additive noise whiteness (see Appendix). One is free to either estimate $v(t)$ in terms of $w(t)$, via a linear operator $H_{v|w}$, or to do it the other way round: use $H_{w|v}$ to estimate $w(t)$ in terms of $v(t)$. The best estimation option is the one with the least quadratic estimation error. One can compare the relative mean error powers MSE of the estimation alternatives via the ratio (a proof is given in Appendix):

$$\frac{\mathrm{MSE}_{v|w}}{\mathrm{MSE}_{w|v}} \sim \frac{1 + S_z(f)/S_{n_v}}{1 + S_z(f)/S_{n_w}}. \tag{3}$$

What this ratio means is that the estimation error is smallest if the time series with the least amount of additive noise is used for estimating the series with the most amount of additive noise. Because the smallest MSE corresponds to the ability to subtract (account for) the largest amount of communality, the time series with the least added noise provides a better estimate of the time series with the most added noise.

For a set $x_i(t)$ of three time series, the partial coherence between any pair i, j is defined as

$$\kappa_{ij}(f) = \frac{S_{x_{i|k}x_{j|k}}(f)}{\sqrt{S_{x_{i|k}}(f)S_{x_{j|k}}(f)}}, \tag{4}$$

i.e. it is just the ordinary coherence between

$$x_{m|k}(t) = x_m(t) - H_{m|k}x_k(t) \tag{5}$$

(for $m \in \{i,j\} \neq k$) which represents the residual left from $x_m(t)$ after the best possible linear (least-squares) estimate based on $x_k(t)$ via the linear transformation $H_{m|k}$ is subtracted. In fact, one can directly calculate (4) from the spectral densities and cross-spectra from $x_m(t)$ (Bendat and Piersol 1993) as pointed out in Albo et al. (2004) without resorting to prior $x_{m|k}(t)$ estimates.

This implies that any triplet of time series with the same underlying communality as in Albo et al.'s (2004) model have the largest amount of communality extracted when the time series with least amount of added noise is used to estimate the other two. In fact, Cases I and II addressed in Albo et al.'s (2004) are extreme, because irrespective of the delays involved, the time series with least amount of added noise (zero noise) perfectly extracts the communality from the other two leading to the results reported therein.

3 The issue of causality

The concept of causality rests on the notion that 'cause cannot follow consequence'. This implies temporal event ordering and leads to a strict notion of relative time precedence.

Definitely, Gersch's method is incapable of pinpointing temporality by examining the nullity of $|\kappa_{ij}(f)|^2$ because it is based on the best linear estimators $H_{i|k}$ and $H_{j|k}$. Despite this, Gersch's method has been used successfully (Gersch and Goddard 1970) when neural activity propagation was derived from a single source whose power decayed along its propagation path so that its later delayed versions were relatively more subject to additive noise in the correct time precedence order. Therefore, the use of the term 'Gersch' causality cannot in general be considered appropriate and its use only serves to cloud connectivity inference issues.

There are many ways in which time precedence can be appraised. One of them is via cross-correlation analysis (Bendat and Piersol 1993) even though a number of restrictions hold for its reliable use (Baccala and Sameshima 2001b). Another is via ideas that reflect

Granger causality (Granger 1969; Baccala et al. 1998) such as partial directed coherence (Baccala and Sameshima 2001a) and allied techniques (Saito and Harashima 1981; Schnider et al. 1989; Kaminski and Blinowska 1991). But before we briefly illustrate this via numerical means, we examine in more detail the nature of the model used by Albo et al. (2004).

4 The theoretical model

In the same notation as Albo et al. (2004), yet rewriting and numbering the model equations we obtain

$$x_0(t) = 0.8x_0(t-1) - 0.5x_0(t-2) + n_0(t), \tag{6}$$

where using

$$u_1(t) = u_2(t-3), \tag{7}$$
$$u_2(t) = x_0(t), \tag{8}$$
$$u_3(t) = u_2(t-5), \tag{9}$$

(see Fig. 1a) leads to the observed variables

$$x_1(t) = u_1(t) + n_1(t), \tag{10}$$
$$x_2(t) = u_2(t) + n_2(t), \tag{11}$$
$$x_3(t) = u_3(t) + n_3(t), \tag{12}$$

which are effectively used for connectivity inference. It is readily apparent that Eqs. (6–9) differ from Eqs. (10–12) as only the $x_i(t)$, $i = 1, 2, 3$ variables are accessible for measurement whereas $x_0(t)$ and $u_i(t)$ are internal variables. The only interesting dynamics is contained in Eq. (6) where $x_0(t)$ depends on its past values $x_0(t-1)$ and $x_0(t-2)$, the measured variables $x_1(t), x_2(t)$ and $x_3(t)$ involve no dynamics of their own and are nothing other than delayed and noise corrupted versions of $x_0(t)$ by means of $n_1(t)$, $n_2(t)$ and $n_3(t)$, respectively.

4.1 Inherent model ambiguity

It is immediately seen that the same exact observations $x_i(t)$ would be generated if the $u_i(t)$ connectivity were described in different ways. For example,

$$u_1(t) = u_2(t-3), \tag{13}$$
$$u_2(t) = x_0(t), \tag{14}$$
$$u_3(t) = u_1(t-2), \tag{15}$$

as shown in Fig. 1b. The most general connectivity structure possible (Fig. 1c) that generates the same $x_i(t)$ is given by

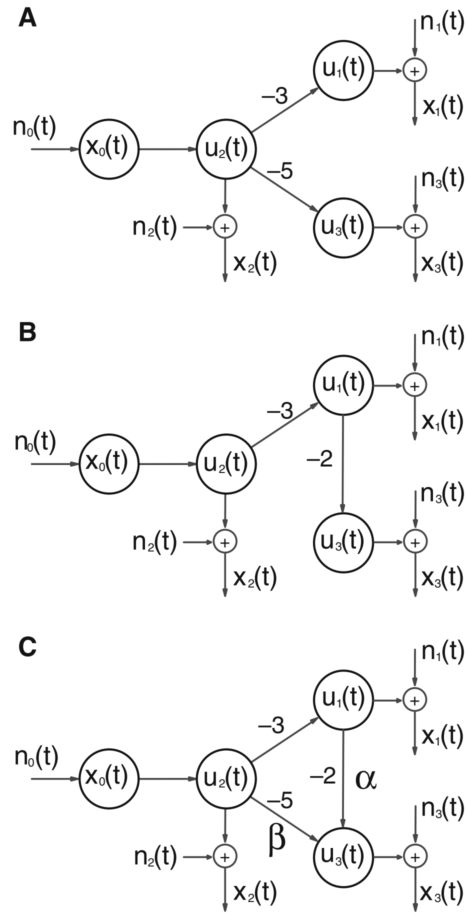

Fig. 1 Diagrams represent alternative internal model structures corresponding to Eqs. (7–9) (**a**) such as Eqs. (13–15) (**b**) or more generally Eqs. (16–18) (**c**). Diagrams **a** and **b** are special cases of **c** obtained by, respectively, setting $\alpha = 0$, $\beta = 1$ or $\alpha = 1$, $\beta = 0$. The negative numbers represent the signal delays involved

$$u_1(t) = u_2(t-3), \tag{16}$$
$$u_2(t) = x_0(t), \tag{17}$$
$$u_3(t) = \alpha u_1(t-2) + \beta u_2(t-5), \tag{18}$$

where $\alpha^2 + \beta^2 = 1$ to insure power flow preservation.

Despite this structural ambiguity, this model imposes a strict time precedence hierarchy among the observed signals: $x_2(t)$ precedes $x_1(t)$ which in turn precedes $x_3(t)$. The observed signals $x_i(t)$ $(i = 1, 2, 3)$ are devoid of any intrinsic feedback and there is neither power loss nor shape distortion in the original signal $x_0(t)$ as it propagates.

While the wiring details connecting the structures remain ambiguous, it is possible to correctly infer the actual time precedence between $x_i(t)$. For brevity, we next illustrate this numerically.

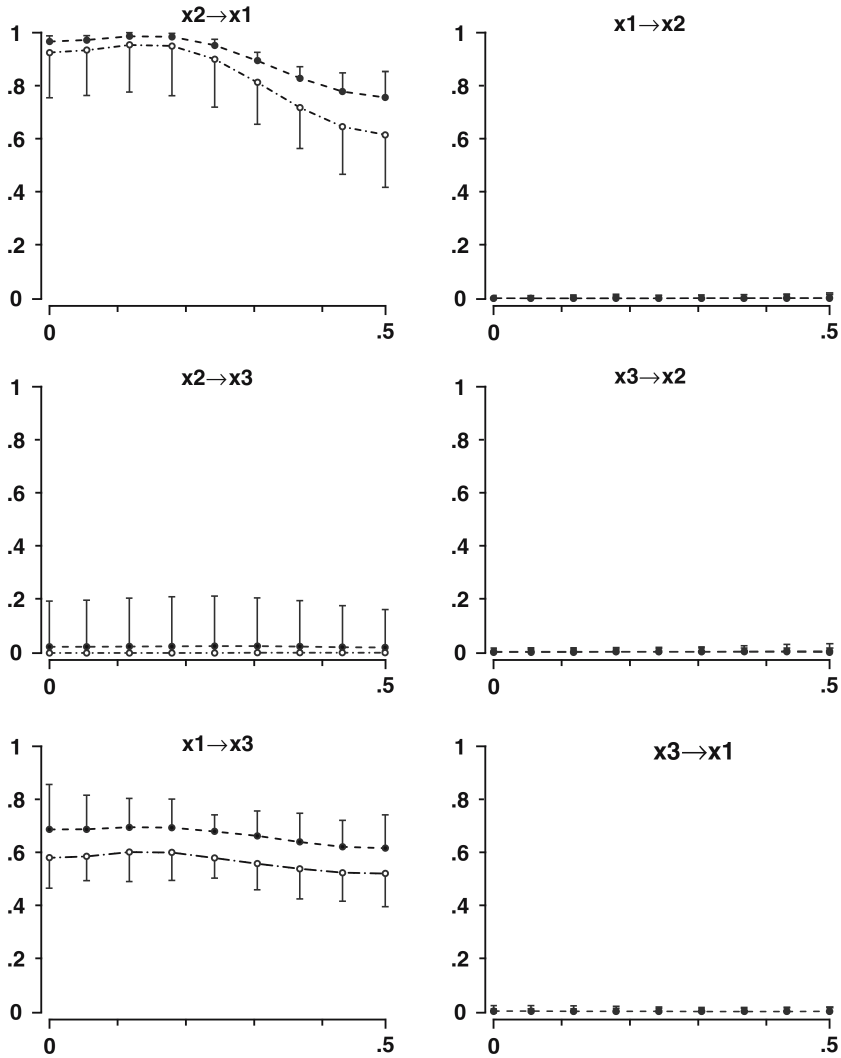

Fig. 2 PDC inference produces identical average connectivity results leading to the minumum lag structure of Fig. 1b because of the use of model order criteria that prevent overparametrization. The nullity of the right panel graphs correctly shows lack of signal feedback. *Half bars* indicate the observed standard deviations

5 Model time precedence inference

We generated data using the model under Cases I and II (Albo et al. 2004) (they differ in that the amount of noise added to delayed versions of $x_0(t)$ to generate $x_1(t)$ and $x_3(t)$ which are swapped according to the case). The numerical results shown represent averages over 60 time series realizations that are 256 points long.

5.1 Partial directed coherence analysis

Conceptually partial directed coherence (PDC) is closely related to partial coherence [3] and was introduced as its factorization which requires the construction of a parametric time series model representation. Average PDC results are shown in Fig. 2 where $x_1(t)$ precedes $x_3(t)$ and where $x_2(t)$ precedes $x_1(t)$ in both Cases I and II in accordance with Fig. 1b.

Finding the number of parameters the model should contain is essential to parametric modelling. This step is usually performed with the help of model order choice criteria (Brockwell and Davis 1991). These criteria essentially provide the minimum number of lags p into the past that need to be considered for adequate modelling whilst avoiding overparametrization whose excessive number of parameters would not only contribute little to fitting but would result in large model estimation uncertainty. The results in Fig. 2 represent averages of what are mostly $p = 3$ models as provided by Akaike's criterion (Lütkepohl 1993). These results are compatible with the structure in Fig. 1b whose maximum delay is of three lags. In fact, it is possible to arbitrarily fix $p = 5$ when constructing a parametric model as in Fig. 3; this leads to an average inferred connectivity similar to Fig. 1c in accordance with the model ambiguity discussed. In fact, setting $p = 5$ is a means of forcing the model to consider the full set of possibilities represented in Fig. 1c.

In summary, PDC is able to highlight the correct signal precedence relationship $x_2 \to x_1$, and $x_1 \to x_3$ and

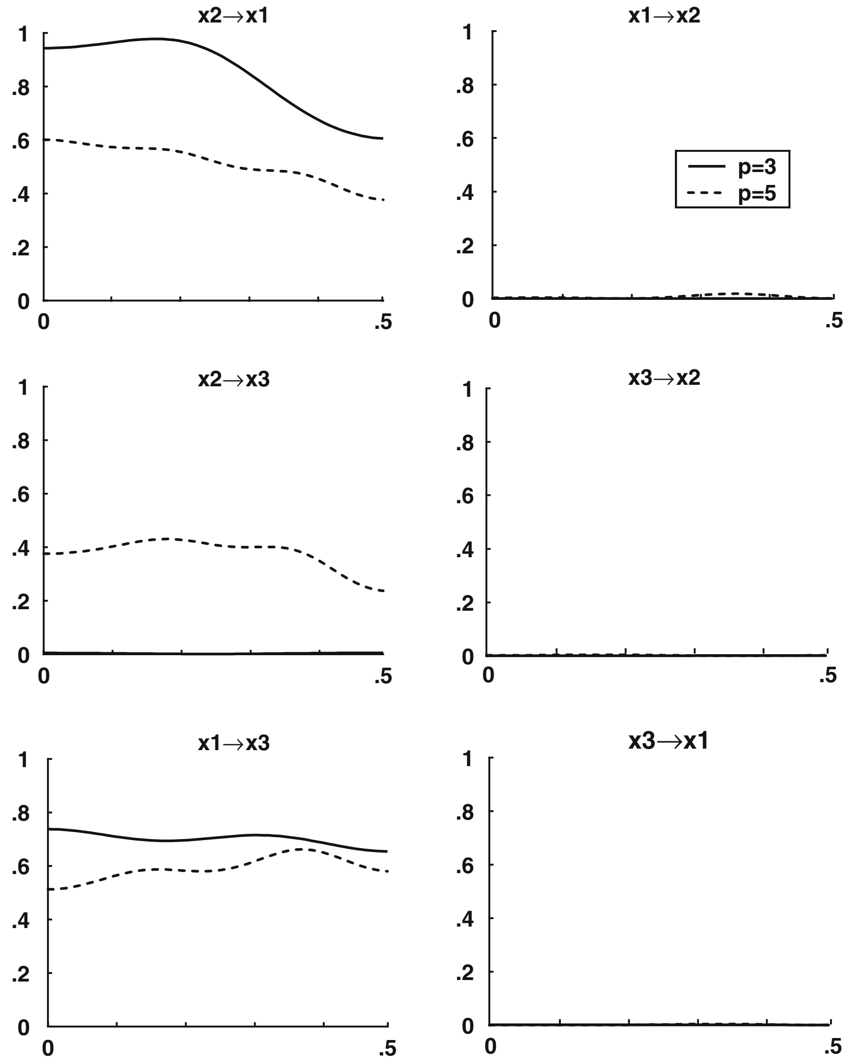

Fig. 3 Comparison between average PDCs computed for Case I model orders $p = 3$ and $p = 5$. Feedback absence is correctly detected as well on the right panels. Use of $p = 3$ produced Fig. 1b whereas Fig. 1c results from $p = 5$. Similar results hold for Case II

$x_2 \to x_3$ in all cases. Most importantly, however, absence of feedback is correctly inferred as well.

5.2 Correlation analysis

Finally, because of its simplicity and nonparametric character, we also use pairwise correlation (as one in fact should always do when analysing many time series). The autocorrelation of $x_2(t)$ and its cross-correlations to $x_1(t)$ and $x_3(t)$ peaking at lags 3 and 5, respectively, are contained in Fig. 4a for a single 256-point time series trial. The cross-correlation between $x_1(t)$ and $x_3(t)$ peaking at lag 2 is also depicted. To illustrate this, we used a high noise scenario when independent zero mean white noise $n_i(t)$ of equal unit variance was added to generate $x_i(t)$ from $x_0(t)$ according to the model equations. It is clear from Fig. 4a that the same time precedence between time series can be deduced as before.

To show that $x_1(t)$ and $x_3(t)$ are nothing more than just noise corrupted versions of $x_2(t)$, we have rescaled the cross-correlations functions by their maxima and shifted them to zero to compare them to the autocorrelation of $x_2(t)$. The results are shown in Fig. 4b: the autocorrelation is always below the cross-correlations. All graphs have the same overall shape, consistent with what one would expect from correlating $x_2(t)$ with a noise corrupted delayed version of itself in the absence of signal dispersion (signal propagation without shape distortion).

Therefore, one must conclude that auto/cross-correlation comparisons suffice to supply information for this model by showing that the observations are compatible with a propagated version of $x_2(t)$ viewed under noise corruption. Furthermore the same clear time precedence hierarchy is obtained as with PDC. We remark, however, that correlation methods have limitations when feedback is present or when there is substantial signal dispersion (see Baccala and Sameshima 2001 for a comparison between PDC and cross-correlation).

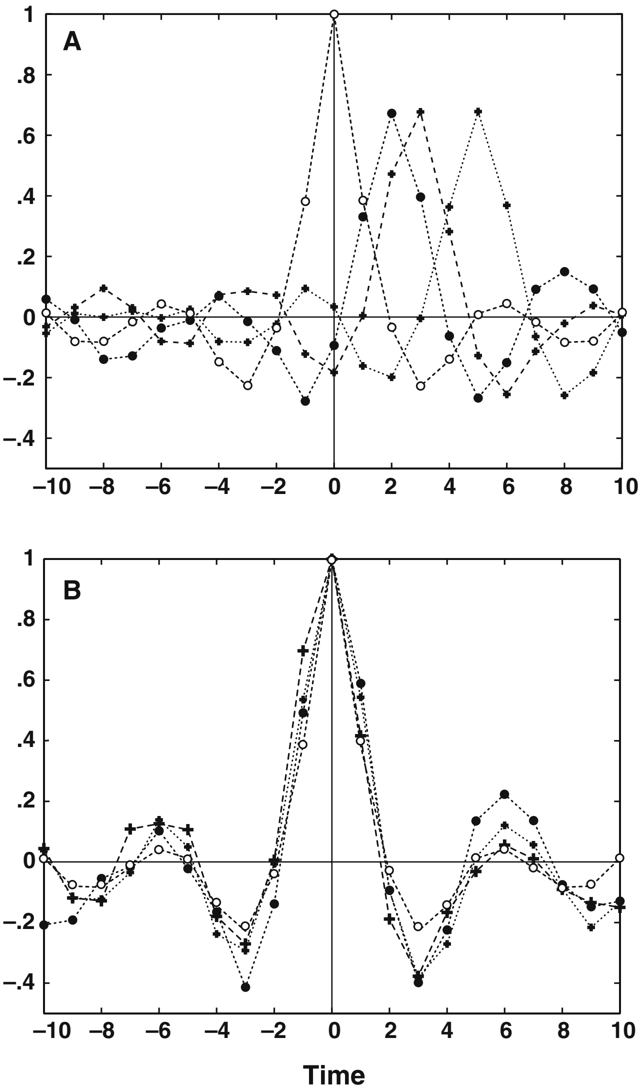

Fig. 4 **a** Single trial ($n = 256$ observed points) autocorrelation of $x_2(t)$ and its cross-correlations with the other time series $x_1(t)$ and $x_3(t)$ peaking at delay lags, repectively of 3 and 5, whereas the cross-correlation between $x_1(t)$ and $x_3(t)$ peaks at their relative delay of 2 sample lags which identify the correct lag between time series even under conditions harsher than Cases I and II in Albo et al. (2004) and **b** shape comparison between the latter after lag correction and amplitude normalization showing distorsionless propagation

6 Final comments and conclusions

We have shown that Gersch's method does not have anything to do with time precedence and hence does not qualify as a *causality* criterion. It does, however, allow one to identify the structure where the same communality is least affected by additive noise regardless of the time precedence between structures. In fact, this is the key to its success in past applications when the signal-to-noise ratio of delayed signal versions was degraded as the signal spread.

To find the correct time precedence, i.e. causal hierarchy between activities, calls for other methods. When little to no signal distortion and feedback are present, simple correlation methods suffice, whereas the presence of a complex feedback structure, may call for more sophisticated methods (Baccala and Sameshima 2001a,b) such as PDC, whose connectivity results must be interpreted cautiously by taking into account the existence of possibly inherently ambiguous connectivity data generating patterns. In practice, this means that known anatomical constraints must be used to address the effective connectivity between structures.

Appendix

Derivation of Eq. 3

From the principle of orthogonality (Haykin 1989) it follows that the optimum linear least squares estimator of an output $v(t)$ from an input $w(t)$ is given by

$$H_{v|w}(f) = \frac{S_{vw}(f)}{S_{ww}(f)} \tag{19}$$

when $H_{v|w}$ is represented in the frequency domain. However, using Eqs. (1) and (2) leads to

$$H_{v|w}(f) = \gamma_{n_w}(f)\ T_{\tau_v - \tau_w}, \tag{20}$$

where $T_{\tau_v - \tau_w}$ is a time shift operator[1] and where

$$\gamma_{n_w}(f) = \frac{S_z(f)}{S_z(f) + S_{n_w}(f)} \tag{21}$$

follows from the mutual independence of $z(t)$ and $n_w(t)$ whose power spectral densities are, respectively, given by $S_z(f)$ and $S_{n_w}(f)$.

To evaluate estimation accuracy we need to compute the mean squared estimation error

$$\mathrm{MSE}_{v|w} = \int S_{v|w}(f)\ \mathrm{d}f \tag{22}$$

where $S_{v|w}(f)$ is the power spectral associated with the estimation error $e_{v|w} = v - H_{v|w}w$. As one can easily show this is given by

[1]Depending on $\tau_v - \tau_w$'s signal, $H_{v|w}(f)$ may be either antecipative or nonantecipative, i.e. its output cannot respond before an input is changed. Sometimes, the notion of being nonantecipative is termed causality in system theoretic developments. Here to prevent confusion we explicitly avoid the term causality in this context.

$$S_{v|w}(f) = \gamma_{n_w}(f)\ S_{n_w}(f). \tag{23}$$

whereas

$$S_{w|v}(f) = \gamma_{n_v}(f)\ S_{n_v}(f) \tag{24}$$

is valid due to the symmetry between v and w.

Because the integrand in (22) is nonnegative and because we assume $n_w(t)$ and $n_v(t)$ are white, to compare the relative magnitudes of $\mathrm{MSE}_{v|w}$ to $\mathrm{MSE}_{w|v}$, it suffices to compare the ratio between (23) and (24). This produces (3) which has the same behaviour as the ratio of interest $\mathrm{MSE}_{v|w}/\mathrm{MSE}_{w|v}$.

References

Albo Z, Di Prisco GV, Chen YH, Rangarajan G, Truccolo W, Feng JF, Vertes RP, Ding MZ (2004) Is partial coherence a viable technique for identifying generators of neural oscillations? Biol Cybern 90:318–326

Baccala LA, Sameshima K (2001a) Partial directed coherence: a new concept in neural structure determination. Biol Cybern 84:463–474

Baccala LA, Sameshima K (2001b) Overcoming the limitations of correlation analysis for many simultaneously processed neural structures. Prog Brain Res 130:33–47

Baccala LA, Sameshima K, Ballester G, Valle AC, Timo-Iaria C (1998) Studying the interaction between brain structures via directed coherence and Granger causality. Appl Signal Process 5:40–48

Bendat JS, Piersol AG (1993) Engineering applications of correlation and spectral analysis, 2nd edn. Wiley, New York

Brockwell PJ, Davis RA (1991) Time series: theory and methods, 2nd edn. Springer, Berlin Heidelberg New York

Gersch W (1972) Causality or driving electrophysiological signal analysis. Math Biosci 14:177–196

Gersch W, Goddard GV (1970) Epileptic focus location–spectral analysis method. Science 169:701–702

Granger CWJ (1969) Investigating causal relations by econometric models and cross-spectral methods. Econometrica 37:424–438

Haykin SS (1989) Modern filters. Macmillan, New York

Kaminski MJ, Blinowska KJ (1991) A new method of the description of the information flow in the brain structures. Biol Cybern 65:203-210

Lütkepohl H (1993) Introduction to multiple time series analysis, 2nd edn. Springer Berlin Heidelberg New York

Saito Y, Harashima H (1981) Tracking of information within multichannel EEG record–causal analysis in EEG. In: Yamaguchi N, Fujisawa K (eds) Recent advances in EEG and EMG data processing. Elsevier, Amsterdam, pp 133–146

Schnider SM, Kwong RH, Lenz FA, Kwan HC (1989) Detection of feedback in the central nervous system using system identification techniques. Biol Cybern 60:203-212