1

Causality and graphical models in time series analysis

Rainer Dahlhaus and Michael Eichler
Universität Heidelberg

1 Introduction

Over the last years there has been growing interest in graphical models and in particular in those based on directed acyclic graphs as a general framework to describe and infer causal relations (Pearl 1995, 2000; Lauritzen 2000; Dawid 2000). This new graphical approach is related to other approaches to formalize the concept of causality such as Neyman and Rubin's potential-response model (Neyman 1935; Rubin 1974; Robins 1986) and path analysis or structural equation models (Wright 1921; Haavelmo 1943). The latter concept has been applied in particular by economists to describe the equilibrium distributions of systems which typically evolve over time.

In this chapter we take up the idea behind the dynamic interpretation of structural equations and discuss (among others) graphical models in which the flow of time is exploited for causal inference. Instead of the intervention based causality concept used by Pearl (1995) and others we apply a simpler notion of causality which is based on the obvious fact that an effect cannot precede its cause in time. Assuming that the variables are observed at different time points the noncausality relations in the system can be derived by examining the (partial) correlation between the variables at different time lags. In the same way we are able to identify the direction of directed edges in the graphical models thus avoiding the assumption of an a priori given ordering of the variables as e.g. in directed acyclic graphs for multivariate data. Since we have repeated measurements of the variables over time we are in a time series setting. More restrictively we focus in this chapter on the situation where we observe only one long multivariate, stationary time series (in contrast, for example, to the panel situation where we have multiple observations of possibly short series).

The concept of causality we use is the concept of Granger causality (Granger 1969) which exploits the natural time ordering to achieve a causal ordering of the variables. More precisely, one time series is said to be Granger causal for another series if we are better able to predict the latter series using all available information than if the information apart from the former series had been used. This concept of causality has been discussed extensively in the econometrics

literature, with several extensions and modifications. Eichler (1999, 2000) has used the definition of Granger causality to define causality graphs for time series. We discuss these graphs in Section 2 together with time series chain graphs and partial correlation graphs for time series (Dahlhaus 2000). In Section 3 we discuss Markov properties and in Section 4 statistical inference for these graphs. Section 5 contains two data examples. Finally Section 6 offers some concluding remarks. In particular we discuss the use of the presented graphs for causal inference and possible sources for wrong identification of causal effects.

2 Graphical models for multivariate time series.

Let $X = \{X_a(t), t \in \mathbb{Z}, a = 1, \ldots, d\}$ be a d-variate stationary process. Throughout the chapter we assume that X has positive definite spectral matrix $f(\lambda)$ with eigenvalues bounded and bounded away from zero uniformly for all $\lambda \in [-\pi, \pi]$. Let $V = \{1, \ldots, d\}$ be the set of indices. For any $A \subseteq V$ we define $X_A = \{X_A(t)\}$ as the multivariate subprocess given by the indices in A. Further $\overline{X}_A(t) = \{X_A(s), s < t\}$ denotes the past of the subprocess X_A at time t.

There exist several possibilities for defining graphical models for a time series X. We can distinguish between two classes of graphical models. In the first class the variable $X_a(t)$ at a specific time t is represented by a separate vertex in the graph. This leads to generalizations of classical graphical models such as the time series chain graph introduced in Definition 2.1. In the second class of graphical models the vertex set only consists of the components X_a of the series, which leads to a coarser modelling of the dependence structure of the series. As we will see below this leads to mixed graphs in which directed edges reflect Granger causality whereas the contemporaneous dependence structure is represented by undirected edges. These graphs are termed Granger causality graphs (Definition 2.4). In addition we have also partial correlation graphs for time series (Definition 2.6) which generalize classical concentration graphs to the time series situation.

Here we restrict ourselves to linear association and linear Granger noncausality, which formally can be expressed in terms of conditional orthogonality of closed linear subspaces in a Hilbert space of random variables (e.g. Eichler 2000, Appendix A.1). For random vectors X, Y, and Z, X and Y are conditionally orthogonal given Z, denoted by $X \perp Y \,|\, Z$, if X and Y are uncorrelated after the linear effects of Z have been removed. It is clear that this linear approach cannot capture the full causal relationships of processes which are partly of nonlinear nature. The need for this restriction results from problems in statistical inference. For example, there exist nonparametric methods for estimating linear dependencies whereas in the general case nonparametric inference seems hardly possible particularly in view of the curse of dimensionality. Even if parametric models are used, inference for nonlinear models is often impractical due to a large number of parameters and many applications therefore are restricted to linear models. We note that the graphical modelling approach can be extended to the nonlinear case by replacing conditional orthogonality by conditional independence which leads to the notion of strong Granger causality (Eichler 2000,

Sect. 5; Eichler 2001). For Gaussian processes the two meanings of the graphs of course are identical. Large parts of the results in this chapter also hold for these general graphs. For details we refer to the discussion in Section 6.

2.1 Time series chain graphs

The first approach for defining graphical time series models naturally leads to chain graphs. Lynggaard and Walther (1993) introduced dynamic interaction models for time series based on the classical LWF Markov property for chain graphs (cf. Lauritzen and Wermuth 1989; Frydenberg 1990). Here we discuss an alternative approach which defines the graph according to the AMP Markov property of Andersson, Madigan, and Perlman (2001). One reason for this is that there exists an intimate relation between the use of the AMP Markov property in time series chain graphs and the recursive structure of a large number of time series models which then immediately characterize the graph. Another reason is that the AMP Markov property is related to the notion of Granger causality which is used in the definition of Granger causality graphs. In particular the AMP Markov property allows to obtain the Granger causality graph from the time series chain graph by simple aggregation.

Definition 2.1 *(Time series chain graph) The time series chain graph (TSC-graph) of a stationary process X is the chain graph $G_{\mathrm{TS}} = (V_{\mathrm{TS}}, E_{\mathrm{TS}})$ with $V_{\mathrm{TS}} = V \times \mathbb{Z}$ and edge set E_{TS} such that*

$$(a, t-u) \longrightarrow (b,t) \notin E_{\mathrm{TS}} \;\Leftrightarrow\; u \leq 0 \text{ or } X_a(t-u) \perp X_b(t) \,|\, \overline{X}_V(t) \backslash \{X_a(t-u)\},$$
$$(a, t-u) \;\text{—}\; (b,t) \notin E_{\mathrm{TS}} \;\Leftrightarrow\; u \neq 0 \text{ or } X_a(t) \perp X_b(t) \,|\, \overline{X}_V(t) \cup \{X_{V\backslash\{a,b\}}(t)\}.$$

Since the process is stationary we have (a,t)—$(b,t) \notin E_{\mathrm{TS}}$ if and only if (a,s)—$(b,s) \notin E_{\mathrm{TS}}$ for all $s \in \mathbb{Z}$. The same shift invariance also holds for the directed edges. We further note that the above conditions guarantee that the process satisfies the pairwise AMP Markov property for G_{TS} (Andersson *et al.* 2000).

Example 2.2 (Vector autoregressive processes) Suppose X is a linear vector autoregressive (VAR) process

$$X(t) = A(1)X(t-1) + \ldots + A(p)X(t-p) + \varepsilon(t) \tag{2.1}$$

where the errors $\varepsilon(t)$ are independent and identically distributed with mean 0 and covariance matrix Σ. If $G_{\mathrm{TS}} = (V_{\mathrm{TS}}, E_{\mathrm{TS}})$ denotes the TSC-graph then it can be shown that

$$(a, t-u) \longrightarrow (b,t) \in E_{\mathrm{TS}} \;\Leftrightarrow\; u \in \{1, \ldots, p\} \text{ and } A_{ba}(u) \neq 0, \tag{2.2}$$

i.e. the directed edges in the graph reflect the recursive structure of the time series. Furthermore the undirected edges specify a covariance selection model (e.g. Dempster 1972) for the errors $\varepsilon(t)$ and we have with $K = \Sigma^{-1}$

$$(a,t) \;\text{—}\; (b,t) \in E_{\mathrm{TS}} \;\Leftrightarrow\; \varepsilon_a(t) \perp \varepsilon_b(t) \,|\, \varepsilon_{V\backslash\{a,b\}}(t) \;\Leftrightarrow\; K_{ab} \neq 0.$$

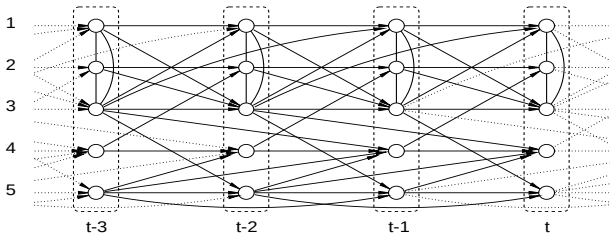

FIG. 1. TSC-graph G_{TS} for the VAR process in Example 2.2.

As an example we consider the five-dimensional VAR(2)-process with parameters

$$A(1)=\begin{pmatrix} \frac{3}{5} & 0 & \frac{1}{5} & 0 & 0 \\ 0 & \frac{3}{5} & 0 & -\frac{1}{5} & 0 \\ \frac{2}{5} & \frac{1}{3} & \frac{3}{5} & 0 & 0 \\ 0 & 0 & 0 & -\frac{1}{2} & \frac{1}{5} \\ 0 & 0 & \frac{1}{5} & 0 & \frac{2}{5} \end{pmatrix}, A(2)=\begin{pmatrix} 0 & 0 & -\frac{1}{5} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{5} & 0 & \frac{1}{3} \\ 0 & 0 & 0 & 0 & -\frac{1}{5} \end{pmatrix}, \Sigma=\begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{3} & 0 & 0 \\ \frac{1}{2} & 1 & -\frac{1}{3} & 0 & 0 \\ \frac{1}{3} & -\frac{1}{3} & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

The above conditions for the edges in E_{TS} lead to the TSC-graph in Fig. 1.

2.2 Granger causality graphs

We now define noncausality as we will use it throughout this chapter, and then Granger causality graphs (Eichler 2000). These are mixed graphs whose vertex set only consists of the components of the series. For the directed edges we use the notion of Granger causality (Granger 1969) and for the undirected edges we use the same definition as for the TSC-graphs. We mention that there has been a lively discussion in the econometrics literature on various definitions of causality; for overviews we refer to Geweke (1984) and Aigner and Zellner (1988).

Definition 2.3 *(Noncausality) X_a is noncausal for X_b relative to the process X_V, denoted by $X_a \nrightarrow X_b$ $[X_V]$, if*

$$X_b(t) \perp \overline{X}_a(t) \,|\, \overline{X}_{V\setminus\{a\}}(t).$$

Further X_a and X_b are partially contemporaneously uncorrelated relative to the process X_V, denoted by $X_a \nsim X_b$ $[X_V]$ if

$$X_a(t) \perp X_b(t) \,|\, \overline{X}(t), X_{V\setminus\{a,b\}}(t).$$

Definition 2.3 can be retained if X_a and X_b are replaced by multivariate subprocesses X_A and X_B, respectively.

In a multivariate setting there are different possible causality patterns such as direct and indirect causality, feedback, or spurious causality (cf. Hsiao 1982). These patterns can be formally described by the use of graphs.

Definition 2.4 *(Causality graph) The Granger causality graph of a stationary process X is the mixed graph $G_{\mathrm{C}} = (V, E_{\mathrm{C}})$ such that for all $a, b \in V$ with $a \neq b$*

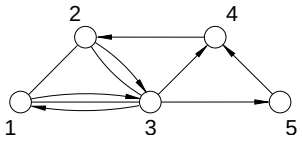

FIG. 2. Causality graph G_{C} for the VAR process in Example 2.2.

(i) $\quad a \longrightarrow b \notin E_{\mathrm{C}} \Leftrightarrow X_a \nrightarrow X_b \ [X_V]$,

(ii) $\quad a \mathrel{—} b \notin E_{\mathrm{C}} \Leftrightarrow X_a \nsim X_b \ [X_V]$.

For simplicity we will speak only of causality graphs instead of Granger causality graphs. We implicitly assume that each component also depends on its own past. This could be expressed by directed self-loops. Since the insertion of these loops does not change the separation properties for the graph we omit these loops for the sake of simplicity.

The relation between the TSC-graph $G_{\mathrm{TS}} = (V_{\mathrm{TS}}, E_{\mathrm{TS}})$ and the causality graph $G_{\mathrm{C}} = (V, E_{\mathrm{C}})$ is simple:

Proposition 2.5 *(Aggregation) Let G_{C} and G_{TS} be the causality graph and the TSC-graph, respectively, of a stationary process X. Then we have*

(i) $\quad a \longrightarrow b \notin E_{\mathrm{C}} \Leftrightarrow (a, t-u) \longrightarrow (b, t) \notin E_{\mathrm{TS}} \quad \forall u > 0 \quad \forall t \in \mathbb{Z}$,

(ii) $\quad a \mathrel{—} b \notin E_{\mathrm{C}} \Leftrightarrow (a, t) \mathrel{—} (b, t) \notin E_{\mathrm{TS}} \quad \forall t \in \mathbb{Z}$.

Proof (i) follows from a replicated application of the intersection property (C5) in Lauritzen (1996). Details are omitted. (ii) is straightforward. □

Example 2.2 (contd) From Proposition 2.5 and (2.2) it follows that $a \longrightarrow b \notin E_{\mathrm{C}}$ if and only if $A_{ba}(1) = A_{ba}(2) = 0$. The resulting causality graph of $X = \{X(t)\}$ is shown in Fig. 2. From this graph we can see for example that X_1 is noncausal for X_4 relative to the full process. More intuitively, the directed paths from 1 to 4 suggest that X_1 causes X_4 indirectly. This indirect cause seems to be mediated by X_3 since all paths from 1 to 4 intersect 3. In the next section we will see that such causality relations indeed can be derived formally from the graph.

2.3 Partial correlation graphs

We now introduce the partial correlation graph for a time series (Dahlhaus 2000). This undirected graph is the counterpart of a concentration graph for ordinary variables. The graph has a simple separation concept and allows additional conclusions about the dependence structure of the series.

Definition 2.6 *(Partial correlation graph) The partial correlation graph (PC-graph) $G_{\mathrm{PC}} = (V, E_{\mathrm{PC}})$ of a stationary process X is given by*

$$a \mathrel{—} b \notin E_{\mathrm{PC}} \Leftrightarrow X_a \perp X_b \,|\, X_{V \setminus \{a,b\}}.$$

There exists a nice characterization of PC-graphs in terms of the inverse of the spectral matrix of the process. Let $f_{ab}(\lambda)$ be the cross-spectrum between

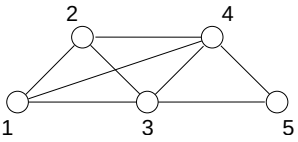

FIG. 3. Partial correlation graph G_{PC} for the VAR process in Example 2.2.

X_a and X_b, $f(\lambda) = \big(f_{ab}(\lambda)\big)_{a,b\in V}$ be the spectral matrix, and $g(\lambda) = f(\lambda)^{-1}$ its inverse. Then it can be shown (see Proposition 2.7) that

$$a \text{ — } b \notin E_{PC} \Leftrightarrow g_{ab}(\lambda) = 0 \text{ for all } \lambda \in [-\pi, \pi].$$

that is, we have a similar characterization as for concentration graphs. More important the absolute rescaled inverse $|d_{ab}(\lambda)| = |g_{ab}(\lambda)|/\big(g_{aa}(\lambda)g_{bb}(\lambda)\big)^{1/2}$ is a measure for the strength of the "connection" between X_a and X_b. To see this let $Y_{a|V\setminus\{a,b\}}$, $Y_{b|V\setminus\{a,b\}}$ the residual series of X_a, resp. X_b after the linear effects of the other components $X_{V\setminus\{a,b\}}$ have been removed. Then the partial cross spectrum $f_{ab|V\setminus\{a,b\}}(\lambda)$ is defined as the cross spectrum between the residual series $Y_{a|V\setminus\{a,b\}}$ and $Y_{b|V\setminus\{a,b\}}$, and

$$R_{ab|V\setminus\{a,b\}}(\lambda) = f_{ab|V\setminus\{a,b\}}(\lambda)/\big(f_{aa|V\setminus\{a,b\}}(\lambda)f_{bb|V\setminus\{a,b\}}(\lambda)\big)^{\frac{1}{2}} \tag{2.3}$$

is the partial spectral coherence (cf. Brillinger, 1981, Chap. 7 and 8), which is a kind of partial correlation between X_a and X_b "at frequency λ". Then we have

Proposition 2.7 *Under the regularity assumptions on X we have for $a \neq b$*

$$g_{aa}(\lambda) = 1/f_{aa|V\setminus\{a\}}(\lambda) \qquad \textit{and} \qquad d_{ab}(\lambda) = -R_{ab|V\setminus\{a,b\}}(\lambda).$$

In particular $d_{ab}(\lambda) = 0$ for all $\lambda \in [-\pi, \pi]$ if and only if a — $b \notin E_{PC}$.

For a proof see Dahlhaus (2000), Theorem 2.4.

Example 2.2 (contd) For general VAR processes of the form (2.1) we have $f(\lambda) = (2\pi)^{-1}A^{-1}(e^{-\mathrm{i}\lambda})\Sigma A^{-1}(e^{\mathrm{i}\lambda})'$ with $A(z) = I - A(1)z - \ldots - A(p)z^p$. Consequently $g(\lambda) = f(\lambda)^{-1} = 2\pi A(e^{\mathrm{i}\lambda})'\Sigma^{-1}A(e^{-\mathrm{i}\lambda})$ and the restriction $g_{ab} \equiv 0$ thus leads to the following parameter constraints

$$\sum_{u=0\vee h}^{p\wedge p+h} \sum_{j,k=1}^{d} K_{jk}A_{ja}(u)A_{kb}(u+h) = 0, \quad (h = -p, \ldots, p) \tag{2.4}$$

with $A(0) = I$ and $K = \Sigma^{-1}$. From these constraints we can derive the PC-graph for the VAR(2)-model from Example 2.2. Assuming that the terms in (2.4) are nonzero whenever one summand is nonzero we obtain an edge a — $b \in E_{\mathrm{PC}}$ whenever a and b are connected by a directed or an undirected edge in G_{C}. Additionally we find that for $a = 1$ and $b = 4$ the equation for $h = 1$ yields $K_{12}A_{24}(1) \neq 0$ which implies that 1 and 4 are also adjacent in G_{PC}. This

leads to the PC-graph in Figure 3. In general further conditional orthogonalities are possible under additional restrictions on the parameters. However, for this particular model an elaborate examination shows that this is not the case.

The condition $K_{12}A_{24}(1) \neq 0$ in the example implies that $1 \text{ — } 2 \longleftarrow 4$ is a subgraph of G_{C}. After insertion of the edge $1 \text{ — } 4$ this subgraph becomes complete, that is, any two vertices in the subgraph are adjacent. Thus the PC-graph of the process has been obtained from the causality graph by completing this subgraph and then converting all directed edges to undirected edges. This operation is called moralization. In the next section we will see that this relationship between the causality graph and the PC-graph holds for all processes satisfying the regularity assumptions in Section 2.

Summarizing the three graphs G_{TS}, G_{C}, and G_{PC} are related by two mappings which correspond to the operations of aggregation and moralization,

$$G_{\mathrm{TS}} \overset{\text{Prop. 2.5}}{\longmapsto} G_{\mathrm{C}} \overset{\text{Rem. 3.5}}{\longmapsto} G_{\mathrm{PC}}.$$

Thus the graphs in a certain sense visualize the dependence structure of the process at different resolution levels, the TSC-graph having the finest resolution.

3 Markov properties

The time series graphs introduced in the previous section visualize the pairwise interaction structure between the components of the process, i.e. they reflect the pairwise Markov properties of the process. The interpretation of the graphs is enhanced by global Markov properties which relate the separation properties of a graph to conditional orthogonality or causality relations between the components of the process. For graphs which contain directed edges there are two main approaches defining global Markov properties: One approach utilizes separation in undirected graphs by applying the operation of moralization to appropriate subgraphs while the other approach is based on path-oriented criteria like *d*-separation (Pearl, 1988) for directed acyclic graphs.

3.1 Global Markov properties for PC-graphs

Partial correlation graphs generalize concentration graphs for multivariate variables. Thus the corresponding global Markov property is naturally based on the separation in undirected graphs. For an undirected graph $G = (V, E)$ and disjoint subsets $A, B, S \subseteq V$ we say that S separates A and B, denoted by $A \bowtie B \,|\, S \;\; [G]$, if every path between A and B in G necessarily intersects S. Then the following result was shown by Dahlhaus (2000).

Theorem 3.1 *Let $G_{\mathrm{PC}} = (V, E_{\mathrm{PC}})$ be the partial correlation graph of X. Then for all disjoint subsets A, B, S of V*

$$A \bowtie B \,|\, S \;\; [G_{\mathrm{PC}}] \;\Rightarrow\; X_A \perp X_B \,|\, X_S.$$

We say that X satisfies the global Markov property for G_{PC}.

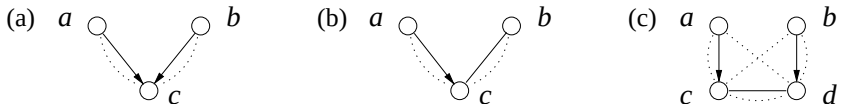

FIG. 4. Configurations in mixed graphs which lead to an additional edge between vertices a and b in the moral graph: (a) flag, (b) immorality, and (c) 2-biflag.

Example 2.2 (contd) In the graph in Fig. 3 the subset $\{3,4\}$ separates $\{1\}$ and $\{5\}$. Thus it follows by Theorem 3.1 that $X_1 \perp X_5 \,|\, X_{\{3,4\}}$.

3.2 Global Markov properties for TSC-graphs

Separation in undirected graphs can also be used to define global Markov properties for directed graphs, chain graphs, or mixed graphs. This approach is based on "moralization" of subgraphs induced by certain subsets of vertices.

Let $G = (V, E)$ be a mixed graph. The undirected subgraph $G^{\mathrm{u}} = (V, E^{\mathrm{u}})$ is obtained from G by removing all directed edges from G. Further a subset A of V induces the subgraph $G_A = (A, E_A)$ where E_A is obtained from E by keeping edges with both endpoints in A. As motivated in the previous section a mixed graph can be moralized by completing certain subgraphs. Following Andersson *et al.* (2001) we call a triple (a, b, c) an immorality (resp, a flag) if $a \longrightarrow c \longleftarrow b$ ($a \longrightarrow c - b$) is a subgraph of G but a and b are not adjacent in G. Further the subgraph induced by (a, b, c, d) is called a 2-biflag if it contains the subgraph $a \longrightarrow c - d \longleftarrow b$ and a and b are not adjacent. The possible forms of immoralities, flags, and 2-biflags in mixed graphs are depicted in Fig. 4 where dotted lines between vertices indicate possible further edges. Most of these edges, however, are excluded in the case of chain graphs. We note that in mixed graphs with multiple edges a subgraph can be both an immorality and a flag.

Definition 3.2 *Let $G = (V, E)$ be a mixed graph. The moral graph $G^{\mathrm{m}} = (V, E^{\mathrm{m}})$ derived from G is defined to be the undirected graph obtained by completing all immoralities, flags, and 2-biflags in G and then converting all directed edges in G to undirected edges.*

In TSC-graphs the orthogonality relations of the process can be described by the global AMP Markov property for chain graphs (Andersson *et al.* 2001). Here we give a slightly different definition of the global AMP Markov property which is based on marginal ancestral subgraphs instead of extended subgraphs. We directly give the general definitions for mixed graphs which include chain graphs as a special case.

As in Frydenberg (1990) we call a vertex a an ancestor of u if either $a = u$ or there exists a directed path $a \longrightarrow \ldots \longrightarrow u$ in G. For $U \subseteq V$ let $\mathrm{an}(U)$ be the set of ancestors of elements in U. The marginal ancestral graph $G\langle A\rangle = (\mathrm{an}(A), E\langle A\rangle)$ induced by A is given by the subgraph $G_{\mathrm{an}(A)}$ with additional undirected edges $a - b$ whenever a and b are not separated by $\mathrm{an}(A)\backslash\{a, b\}$ in G^{u}. Due to these additional edges the subprocess $X_{\mathrm{an}(A)}$ satisfies the pairwise Markov properties for the marginal ancestral graph $G\langle A\rangle$.

Theorem 3.3 *Let $G_{\mathrm{TS}} = (V_{\mathrm{TS}}, E_{\mathrm{TS}})$ be the TSC-graph of X and A, B, S disjoint subsets of V_{TS}. Then if $A \bowtie B \,|\, S$ $[G_{\mathrm{TS}}\langle A \cup B \cup S\rangle^{\mathrm{m}}]$ we have*

$$\{X_a(t), (a,t) \in A\} \perp \{X_b(t), (b,t) \in B\} \,|\, \{X_s(t), (s,t) \in S\}.$$

We say that X satisfies the global AMP Markov property for G_{TS}.

Proof Since for random variables X, Y_1, Y_2, Z the orthogonality relations $X \perp Y_i \,|\, Z$, $i = 1, 2$ always imply $X \perp (Y_1, Y_2) \,|\, Z$, the pairwise and the global AMP Markov property for G_{TS} are equivalent (Andersson *et al.* 2001). □

Example 2.2 (contd) Since X is a VAR(2) process the directed edges in the graph in Figure 1 have at most lag 2. The same holds for the edges in $G_{\mathrm{TS}}\langle A \cup B \cup S\rangle^{\mathrm{m}}$ for all sets A, B, and S. Thus it follows that $X(t-3)$ and $X(t)$ are conditionally orthogonal given $X(t-1)$ and $X(t-2)$.

The global AMP Markov property can be used to derive general causality relations from the graph involving subprocesses of X. More precisely, for $A, B \subseteq S$ X_A is noncausal for X_B relative to X_S whenever $\bar{A}_t \bowtie B_t \,|\, \bar{S}_t \backslash \bar{A}_t$ $[G_{\mathrm{TS}}\langle \bar{S}_t \cup B_t\rangle^{\mathrm{m}}]$ where $\bar{A}_t = \{(a,s), a \in A, s \leq t\}$ and $A_t = \{(a,t), a \in A\}$. In the special case where $A = \{a\}$, $B = \{b\}$ and $S = V \backslash \{a, b\}$ this gives the pairwise causality relations relative to the full process X.

3.3 Global Markov properties for causality graphs

In causality graphs each component $X_a = \{X_a(t)\}$ of the process is represented only by a single vertex a in the graph. Thus each vertex a has to be interpreted either as $\overline{X}_a(t)$ or as $X_a(t)$ depending on the type and the direction of the adjacent edges. All information on the pairwise causality relations is encoded only in the type and the direction of the edges. Since the directions of the edges are removed by moralization we cannot hope to retrieve unidirectional causalities by this method. Nevertheless, the moralization criterion can be used to derive global conditional orthogonality relations between the components of the series, i.e. in this case the vertex a stands for X_a as in PC-graphs.

Theorem 3.4 *X satisfies the global AMP Markov property for G_{C}, i.e.*

$$A \bowtie B \,|\, S \;\; [G_{\mathrm{C}}\langle A \cup B \cup S\rangle^{\mathrm{m}}] \;\Rightarrow\; X_A \perp X_B \,|\, X_S.$$

For a proof see Eichler (2001).

Remark 3.5 If we set $A = \{a\}$, $B = \{b\}$, and $S = V \backslash \{a, b\}$ the corresponding marginal ancestral graph is the moral graph $G_{\mathrm{C}}^{\mathrm{m}}$ of G_C itself irrespectively of a and b. Thus we have for all $a, b \in V$

$$\{a\} \bowtie \{b\} \,|\, V \backslash \{a, b\} \;\; [G_{\mathrm{C}}^{\mathrm{m}}] \;\Leftrightarrow\; a \text{ — } b \notin E_{\mathrm{C}}^{\mathrm{m}} \;\Rightarrow\; X_a \perp X_b \,|\, X_{V \backslash \{a,b\}},$$

i.e. the process X satisfies the pairwise Markov property for the moral graph $G_{\mathrm{C}}^{\mathrm{m}}$. In particular G_{PC} is a subgraph of $G_{\mathrm{C}}^{\mathrm{m}}$ which together with Theorem 3.1 implies that X satisfies also the global Markov property for the moral graph $G_{\mathrm{C}}^{\mathrm{m}}$.

The global AMP Markov property for G_{C} is stronger than the global Markov property for $G_{\mathrm{C}}^{\mathrm{m}}$ or G_{PC}. To illustrate this we consider the simple causality graph $1 \longrightarrow 3 \longleftarrow 2$. With $A = \{1\}$, $B = \{2\}$ and $S = \varnothing$ the global AMP Markov property yields $X_1 \perp X_2$. On the other hand the moral graph G_C^{m} is complete and consequently no conditional orthogonality relation can be read off the graph.

In TSC-graphs the derivation of causal relations by the moralization criterion is based on the fact that the past and the present of a component at time t are represented by different subsets of vertices in the TSC-graph. To preserve the causal ordering when moralizing causality graphs Eichler (2000) introduced a special separation concept for mixed graphs which is based on the idea of splitting the past and the present of certain variables and considering them together in a chain graph. For any subset B of $U \subseteq V$ this splitting of the past and the present for the variables in B can be accomplished by augmenting the moral graph $G_{\mathrm{C}}\langle U\rangle^{\mathrm{m}}$ with new vertices b^* for all $b \in B$ which then represent the present of these variables, that is, b^* corresponds to $X_b(t)$, whereas all other vertices stand for the past at time t, for example, a for $\overline{X}_a(t)$. Let $B^* = \{b^* | b \in B\}$ be the set of augmented vertices. The new vertices in B^* are joined by edges such that we obtain a chain graph with two chain components $\mathrm{an}(U)$ and B^* which reflects the AMP Markov properties of the process. More precisely we define the augmentation chain graph $G\langle U\rangle_{B^*}^{\mathrm{aug}}$ as the mixed graph $(\mathrm{an}(U) \cup B^*, E^{\mathrm{aug}})$ with edge set E^{aug} such that for all $a_1, a_2 \in U$ and $b_1, b_2 \in B$

$$\begin{aligned} a_1 \text{ — } a_2 \notin E^{\mathrm{aug}} &\Leftrightarrow a_1 \text{ — } a_2 \notin E\langle U\rangle^{\mathrm{m}}, \\ a_1 \longrightarrow b^* \notin E^{\mathrm{aug}} &\Leftrightarrow a_1 \longrightarrow b_1 \notin E, \\ b_1^* \text{ — } b_2^* \notin E^{\mathrm{aug}} &\Leftrightarrow b_1 \bowtie b_2 \,|\, B\backslash\{b_1, b_2\} \;\; [G^{\mathrm{u}}]. \end{aligned}$$

Let $Y = (Y_v)_{v \in \mathrm{an}(U)}$ and $Y^* = (Y_b^*)_{b \in B}$ random vectors with components $Y_v = \overline{X}_v(t)$ and $Y_b^* = X_b(t)$. Then (Y, Y^*) satisfies the pairwise AMP-Markov property with respect to $G\langle U\rangle_{B^*}^{\mathrm{aug}}$. This suggests defining the global Markov property for causality graphs by the global AMP Markov property in augmentation chain graphs.

Definition 3.6 *X satisfies the global causal Markov property for the mixed graph $G = (V, E)$ if for all $U \subseteq V$ and all disjoint partitions A, B, C of U*

$$\begin{aligned} A \bowtie B^* \,|\, U\backslash A \;\; [(G\langle U\rangle_{B^*}^{\mathrm{aug}})^{\mathrm{m}}] &\Rightarrow X_A \nrightarrow X_B \;\; [X_U], \\ A^* \bowtie B^* \,|\, S \cup C^* \;\; [(G\langle U\rangle_{U^*}^{\mathrm{aug}})^{\mathrm{m}}] &\Rightarrow X_A \nsim X_B \;\; [X_U]. \end{aligned}$$

The assumptions on the spectral matrix in Section 2 now guarantee the equivalence of pairwise and global causal Markov properties.

Theorem 3.7 *Let $G_{\mathrm{C}} = (V, E_{\mathrm{C}})$ be the causality graph of X. Then X satisfies the global causal Markov property with respect to G_{C}.*

For a proof see Eichler (2000), Theorem 3.8.

Example 2.2 (contd) As already mentioned, an intuitive interpretation of the causality graph in Fig. 2 suggests that X_1 has only an indirect effect on X_4

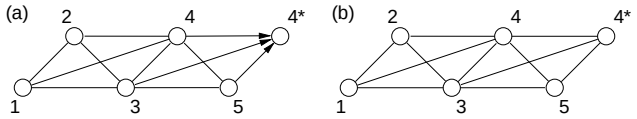

FIG. 5. Illustration of the global causal Markov property for the vector autoregressive process in Example 2.2: (a) Augmentation chain graph $G\langle\{1,3,4\}\rangle^{\mathrm{aug}}_{\{4^*\}}$ and (b) its moral graph $(G\langle\{1,3,4\}\rangle^{\mathrm{aug}}_{\{4^*\}})^{\mathrm{m}}$.

which is mediated by X_3. That this interpretation is indeed correct can now be shown by deriving the relation $X_1 \nrightarrow X_4\ [X_{\{1,3,4\}}]$ from the corresponding augmentation chain graph $G\langle\{1,3,4\}\rangle^{\mathrm{aug}}_{\{4^*\}}$.

Since the ancestral set generated by $\{1,3,4\}$ is equal to the full set V we start from the moral graph G^{m} in Fig. 3. Augmenting the graph with a new vertex 4^* and joining this with vertex 4 and its parents 3 and 5 by arrows pointing towards 4^* we obtain the augmentation chain graph $G\langle\{1,3,4\}\rangle^{\mathrm{aug}}_{\{4^*\}}$ in Fig. 5 (a). As the graph does not contain any flag or 2-biflag the removal of directions yields the moral graph $(G\langle\{1,3,4\}\rangle^{\mathrm{aug}}_{\{4^*\}})^{\mathrm{m}}$ in Fig. 5 (b). In this graph the vertices 1 and 4^* are separated by the set $\{3,4\}$ and hence the desired noncausality relation follows from Theorem 3.7. Similarly it can be shown that X_1 and X_4 are contemporaneously partially uncorrelated relative to the same subprocess $X_{\{1,3,4\}}$ (Eichler 2000, Example 3.9)

3.4 The p-separation criterion

The moralization criterion in the previous sections is not a separation criterion in G_{TS} of G_{C} itself because the ancestral graphs and augmented moral graphs appearing in the definitions of the global AMP and global causal Markov property vary with the subsets A, B, and S. For AMP chain graphs models Levitz *et al.* (2001) presented a pathwise separation criterion, called p-separation, which is similar to the d-separation criterion for directed acyclic graphs. In Eichler (2001) it is shown that the global AMP Markov property and the global causal Markov property for causality graphs equivalently can be formulated by an adapted version of the p-separation criterion.

4 Inference for graphical time series models

4.1 Conditional likelihood for graphical autoregressive models

We first discuss the case of fitting VAR(p) models restricted with respect to a time series chain graph. In practice the true TSC-graph is unknown and one has to apply model selection strategies to find the best approximation for the true graph. Depending on the model discrepancy this leads to different model selection criteria such as AIC or BIC. More precisely, let $G = (V_{\mathrm{TS}}, E)$ be a graph from the class $\mathscr{G}_{\mathrm{TS}}(p)$ of all time series chain graphs whose edges have at most lag p and which are invariant under translation in the sense that for all $a, b \in V$, and all $t, s, u \in \mathbb{Z}$ we have $(a, t-u) \longrightarrow (b, t) \notin E$ if and only if

$(a, s-u) \longrightarrow (b, s) \notin E$ and (a, t) — $(b, t) \notin E$ if and only if (a, s) — $(b, s) \notin E$. We consider Gaussian VAR(p) models of the form

$$X(t) = A(1)X(t-1) + \ldots + A(p)X(t-p) + \varepsilon(t), \qquad \varepsilon(t) \overset{\text{iid}}{\sim} \mathcal{N}(0, \Sigma_G)$$

where the parameters $A_G = (A(1), \ldots, A(p))$ and $K_G = \Sigma_G^{-1}$ satisfy the following constraints

$$A_{ba}(u) = 0 \text{ if } (a, t) \longrightarrow (b, t+u) \notin E \text{ and } K_{ab} = 0 \text{ if } (a, t) — (b, t) \notin E.$$

We call a vector autoregressive model with these constraints on the parameters a VAR(p, G) model.

Given observations $X(1), \ldots, X(T)$ let $T_0 = T - p$. The conditional log likelihood function of a Gaussian VAR(p, G) model for the data can be written as (cf. Lütkepohl 1993)

$$\mathscr{L}_T\big(A_G, K_G\big) = \frac{d}{2}\log(2\pi) - \frac{1}{2}\log|K_G| + \frac{1}{2T_0}\sum_{t=p+1}^{T} e(t)' K_G e(t),$$

where $e(t) = X(t) - \sum_{u=1}^{p} A(u)X(t-u)$. Differentiating the log-likelihood we obtain equations for the maximum likelihood estimate which can be solved numerically by an iterative algorithm (Dahlhaus and Eichler 2001). Furthermore we can use model selection criteria for selection of the optimal VAR(p, G) model among all models with $1 \le p \le P$ and $G \in \mathscr{G}_{\text{TS}}(p)$. Since in practice the number of possible models is too large one has to use special model selection strategies.

Alternatively we can fit VAR(p) under the restrictions of a mixed graph G from the class of all causality graphs $\mathscr{G}_{\text{C}}$. Such models are a special case of the VAR(p, G) models since by Proposition 2.5 for each $G \in \mathscr{G}_{\text{C}}$ there exists a chain graph $G' = (V_{\text{TS}}, E') \in \mathscr{G}_{\text{TS}}(p)$ which implies exactly the same constraints on the parameter as G, i.e. we have for all $t \in \mathbb{Z}$ (a, t) — $(b, t) \notin E'$ if and only if a — $b \notin E$ and $(a, t) \longrightarrow (b, t+u) \notin E'$ if and only if $u > p$ or $a \longrightarrow b \notin E$.

4.2 Partial directed and partial contemporaneous correlation

We now present a nonparametric approach for estimating TSC-graphs which is based on the definition of the TSC-graphs itself. For random variables X, Y, and Z let $\text{corr}_{\text{L}}(X, Y|Z)$ denote the correlation between X and Y after the linear effects of Z have been removed. Thus we have $\text{corr}_{\text{L}}(X, Y|Z) = 0$ if and only if $X \perp Y \,|\, Z$. We define the partial directed correlation at lag $u > 0$ by

$$\pi_{ba}(u) = \text{corr}_{\text{L}}\left(X_b(t), X_a(t-u)\big|\overline{X}_V(t)\backslash\{X_a(t-u)\}\right)$$

and the partial contemporaneous correlation by

$$\pi^{\circ}_{ba} = \text{corr}_{\text{L}}\left(X_b(t), X_a(t)\big|\overline{X}_V(t) \cup \{X_{V\backslash\{a,b\}}(t)\}\right).$$

Then the time series chain graph G_{TS} of a process $X = \{X(t)\}$ can be characterized in terms of the partial directed and partial contemporaneous correlation,

$$(a, t-u) \longrightarrow (b,t) \notin E_{\mathrm{TS}} \Leftrightarrow \pi_{ba}(u) \neq 0,$$
$$(a,t) \text{ — } (b,t) \notin E_{\mathrm{TS}} \Leftrightarrow \pi^{\circ}_{ba} \neq 0.$$

When estimating the partial directed and partial contemporaneous correlations from observations $X(1), \ldots, X(T)$ it is clear that only the observed part of the past can be taken into account. Therefore we define

$$\hat{\pi}_{ba}(u) = \widehat{\mathrm{corr}}_{\mathrm{L}} \left(X_b(p), X_a(p-u) \big| \{X(1), \ldots, X(p-1)\} \backslash \{X_a(p-u)\} \right)$$

for some fixed p, with a similar definition for $\hat{\pi}^{\circ}_{ba}$. In Dahlhaus and Eichler (2001) an iterative scheme for the computation of these estimators has been suggested.

The estimators $\hat{\pi}_{ba}(u)$ and $\hat{\pi}^{\circ}_{ba}$ can be used for testing for the existence of an edge in the TSC-graph G_{TS}. More precisely, if the edge $(a, t-u) \longrightarrow (b,t)$ is absent in the true TSC-graph then $T\hat{\pi}^2_{ba}(u)$ is asymptotically χ^2_1-distributed. Similarly we can test for the absence of undirected edges in the TSC-graph using the partial contemporaneous correlation. For model selection these tests have to be applied repeatedly for various edges which raises the usual problems associated with multiple testing.

In Proposition 2.5 we have seen that aggregation of the edges in the TSC-graph leads to the causality graph of the process. The idea of aggregation suggests to use the sum of $\hat{\pi}^2_{ba}(u)$ for testing for the absence of a directed edge in the causality graph. Let $\hat{R}_p$ being the empirical covariance matrix of X and $\hat{H}$ a $p \times p$ matrix with entries $\hat{H}_{uv} = (\hat{R}_p^{-1})_{bb}(u-v)/(\hat{R}_p^{-1})_{bb}(0)$. Correcting for the dependence between the $\hat{\pi}_{ba}(u)$ at different lags we then obtain as a test statistic

$$S_{ab} = T \; \hat{\pi}'_{ba} \hat{H}^{-1} \hat{\pi}_{ba},$$

where $\hat{\pi}_{ba}$ is the vector of partial directed correlations. Under the null hypothesis that X_a is noncausal for X_b the test statistic is asymptotically χ^2_p-distributed. An equivalent test statistic has been presented by Tjøstheim (1982) for testing for causality in multivariate time series (without considering causality relations between the variables at the different time points themselves).

4.3 Partial correlation graphs

Inference for PC-graphs is in a certain sense more complicated than for TSC-graphs or causality graphs. Suppose for example that one wants to test by a likelihood ratio test the hypothesis that the edge a — b is missing in the PC-graph of the VAR model in Example 2.2. This requires the calculation of the maximum likelihood estimator under the restrictions (2.4). There does not exist an analytic solution for this problem whereas numerical solutions seem to be difficult due to the particular nonlinear structure of the restrictions.

Instead we suggest a nonparametric method based on the characterization of a missing edge by a vanishing partial spectral coherence in Proposition 2.7. By

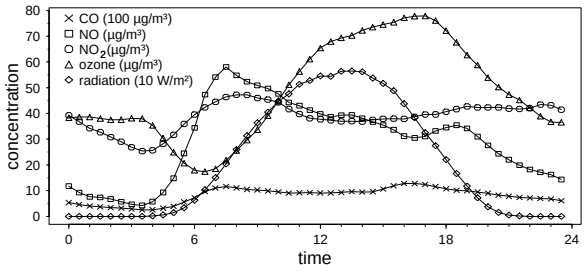

FIG. 6. Average of daily measurements over 61 days in summer.

this proposition the partial spectral coherence can be obtained by an inversion and rescaling of the spectral density matrix. We therefore estimate the spectral matrix and invert and rescale this estimate. As an estimator for $f_{ab}(\lambda)$ we take a kernel estimator based on the periodogram (cf. Dahlhaus, 2000).

We then test the hypothesis $a \text{ — } b \notin E_{\mathrm{PC}}$ by using an approximation of the distribution of $\sup_{\lambda\in[-\pi,\pi]} |\hat{R}_{ab|V\setminus\{a,b\}}(\lambda)|^2$ under the hypothesis, i.e. under $R_{ab|V\setminus\{a,b\}}(\lambda) = 0$ (for more details see Dahlhaus 2000, Section 5).

Alternatively the test can be based in the time domain on the partial correlations of X_a and X_b after removing the linear effects of $X_{V\setminus\{a,b\}}$ (see Eichler *et al.* 2000).

5 Applications

In order to illustrate the methods presented in this chapter we discuss two real data example from chemistry and neurology. For further examples we refer to Gather *et al.* (2002), who analyzed time series from intensive care monitoring by the help of partial correlation graphs, and Eichler *et al.* (2000), who discussed the identification of neural connectivities from spike train data.

5.1 Application to air pollution data

The method was used to analyze a 5-dimensional time series of length 35088 of air pollutants recorded half-hourly from January 1991 to December 1992 in Heidelberg. The recorded variables were CO and NO (mainly emitted from cars, house-heating and industry), NO_2 and O_3 (created in different reactions in the atmosphere) and the global radiation intensity I_{rad} which plays a major role in the generation of ozone. Details on these reaction can be found e.g. in Seinfeld (1986, Chapter 4). Figure 6 shows the daily course of the five variables averaged over 61 successive days in summer. CO and NO increase early in the morning due to traffic and, as a consequence, also NO_2 increases. O_3 increases later due to the higher level of NO_2 and the increase of the global radiation. Figure 6 indicates that all variables are correlated at different lags.

The following analysis is based on the residual series after subtracting the (local) average course as shown in Fig. 6. The original series contained a few missing values (less than 2%) which were completed by interpolation of the

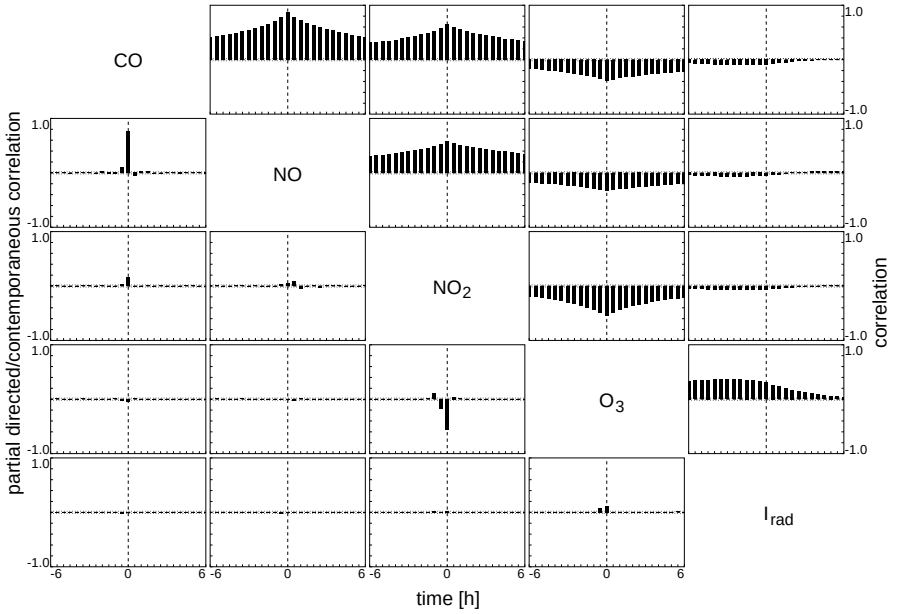

FIG. 7. Correlations (above diagonal) and partial directed and partial contemporaneous correlations (below diagonal) for air pollution data.

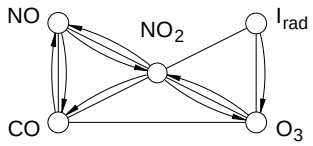

FIG. 8. Causality graph for the air pollution data.

residual series with splines.

Figure 7 shows above the diagonal estimates for the correlations between the variables at different lags and below the diagonal estimates for the partial directed and partial contemporaneous correlations as discussed in Section 4.2. Since $\pi_{ab}(u)$ is defined only for $u > 0$ we have plotted $\pi_{ba}(-u)$ for $u < 0$ and π°_{ab} for $u = 0$. Furthermore horizontal dashed lines indicate simultaneous 95% test bounds for the test $\pi_{ab}(u) = 0$ (π°_{ab}=0). These dashed lines are close to zero and hardly visible. The test bounds lead to the TSC-graph (not plotted) and the causality graph in Fig. 8.

The graph reflects the creation of O_3 from NO_2 and the decrease of O_3 leading to an increase of NO_2 (indicated by the very strong negative partial contemporaneous correlation between O_3 and NO_2 in Fig. 7). The graph also confirms that the global radiation plays a major role in the process of O_3-generation. Furthermore, there is a very strong partial contemporaneous correlation between

CO and NO (both are emitted from cars etc.). The meaning of the other edges (and of some of the missing edges) is less obvious. Chemical reactions between air pollutants are very complex. In particular, one has to be aware of the fact that NO_2 and O_3 are not only increased but also decreased by several chemical reactions and that several other chemicals play an important role.

Some of these reactions can be explained by a photochemical theory (cf. Seinfeld 1986, Section 4.2). This theory is confirmed by the above graph: First the edge between I_{rad} and NO_2 represents the photolysis of NO_2. Second the edges between NO and NO_2 are present but Fig. 7 shows that the partial directed and partial contemporaneous correlations are less strong than the correlation between CO and NO_2. This indicates that mainly the concentration of CO (and not of NO) is responsible for the generation of NO_2 which means that NO_2 is mainly generated via a radical reaction (where CO is involved) and not in a direct reaction (where CO is not involved). This also is in accordance with the photochemical theory for ozone generation.

The example shows the effect of the discretization interval (30 minutes) on the graph. Obviously this interval is too large to resolve the direction in the chemical reactions from the time flow. Thus most of the causality is detected as undirected partial contemporaneous correlation instead of directed Granger causality. There is one exception: The strong partial contemporaneous correlation between CO and NO is due to the fact that both are emitted at the same time from cars, i.e. we have a confounder which is causal for both CO and NO.

The example also reveals the need for a more quantitative analysis in that one wants to discriminate between major effects (such as NO_2 — O_3 — I_{rad}) and minor effects (such as CO — O_3) leading to the idea of edges with different grey levels or colors reflecting the "strength" of a connection. This has to be pursued, in particular for applications with large data sets.

An analysis of the same data set discretized at a 4 hour interval with PC-graphs can be found in Dahlhaus (2000).

5.2 Application to human tremor data

The second example in this section is concerned with the identification of neurological signal transmission pathways in the investigation of human tremor. Tremor is defined as the involuntary, oscillatory movement of parts of the body, mainly the upper limbs. It has been shown that patients suffering from Parkinson's disease show a tremor-related cortical activity which can be detected in the EEG time series by cross-spectral analysis (e.g. Timmer *et al.* 2000).

In one experiment the EEG and the surface electromyogram (EMG) of the left hand wrist extensor muscle in a healthy subject have been measured during an externally enforced oscillation of the hand with 1.9 Hz. The EMG signal was band-pass filtered to avoid aliasing effects and undesired slow drifts. Additionally the signal was digitally full wave rectified. The resulting time series reflects the muscle activity encoded in the envelope of the originally measured signal. The EEG recordings were performed using a 64-channel EEG system. The potential

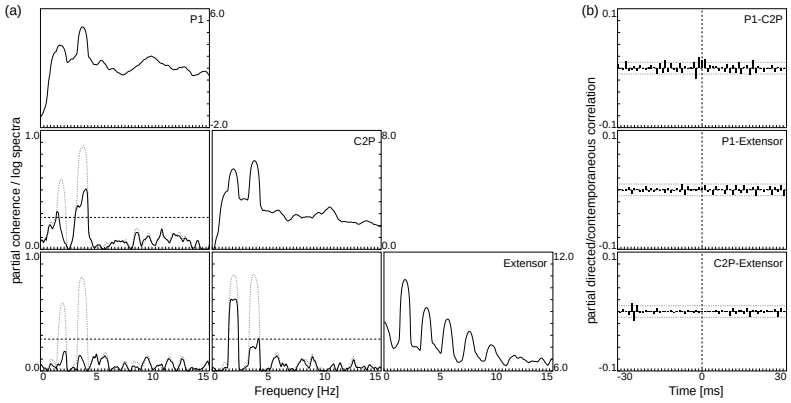

FIG. 9. Human tremor data: (a) On diagonal: logarithm of auto-spectra; below diagonal: spectral coherences (dotted) and partial spectral coherences (solid). (b) Partial directed and partial contemporaneous correlations.

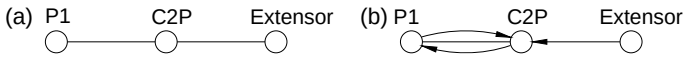

FIG. 10. (a) Causality graph and (b) partial correlation graph for the human tremor data.

field measured over the scalp was transformed into the reference free current distribution which reflects the underlying cortical activity by calculating second spatial derivatives. All data were simultaneously sampled at 1000 Hz. For the analysis a subseries of length $T = 40000$ has been used.

In a preliminary analysis significant coherence has been found between the EMG and the EEG of channels P1 and C2P. Figure 9 (a) displays estimates for the spectra on the diagonal and estimates for the ordinary and the partial spectral coherence between the respective time series below the diagonal. The higher harmonics in the spectra indicate that the process has a nonlinear dynamic.

A significant coherence at the 5% level is found at the tremor frequency and its higher harmonics between all three channels. The partial coherence between the extensor EMG and the EEG of channel P1 falls below the 5% threshold. Consequently there is no edge between P1 and the Extensor in the corresponding partial correlation graph which is given in Fig. 10 (a). The graph suggests that the coherence between P1 and the EMG signal does not represent a direct connection, but that the signal transfer is mediated via C2P.

For a further investigation of the interrelations and in particular the directions of signal transmission we also estimated the partial directed and partial contemporaneous correlations which are given in Fig. 9 (b). The processes seem to be only weakly correlated and the interrelations which were clearly discernable in the frequency domain are nearly undetectable in the time domain. Using the threshold for the identification of the edges we obtain the causality graph

in Fig. 10 (b). The graph reveals that the cortical activity at location C2P is directly caused by the oscillation of the hand whereas the cortical activity at P1 is only indirectly influenced via C2P. The edges between P1 and C2P indicate that signals between the two regions in the cortex are transmitted in both directions with very short delay. On the other hand the delay of about 25 ms in the transmission of signals from the extensor to C2P corresponds quite well to the length of the pathway.

The example shows that due to the periodic nature of tremor oscillation frequency domain methods are much better suited for the identification of the corresponding neurological connectivities. Therefore it seems to be advantageous to have a frequency domain method for the estimation of causality graphs. A first step in that direction is the partial directed coherence (Sameshima and Baccalá 1999), but so far there does not exist any test based on this statistic. In the context of neurophysiological point process data Dahlhaus *et al.* (1997) suggested an indirect method: They first identified the PC-graph and then used in addition the partial phase information for inference on the directions in the interrelation between the point processes.

6 Discussion

The recent development of causal Markov models has shown that graphical models provide a general framework for describing causal relations. With this chapter we want to contribute to the discussion of causality concepts and their relation to graphical models. Focusing on the temporal aspect of causality—an effect cannot precede its cause in time—we have based our approach to causal modelling on the concept of Granger causality. Causal Markov models are then defined by merging Granger causality with mixed graphs.

In applications the causality graph describing the causal relations between the variables is typically unknown. Due to the time series setting the directions of the edges in the graph and thus of the causal effects can be identified from the data. However there are also various problems which may lead to wrong identification of causal effects.

First Granger causality measures only association between variables at different times and therefore cannot distinguish between direct causal effects and an association which is due to an unmeasured confounder. This problem was already discussed briefly by Granger (1969) when introducing his notion of causality. Granger assumed in his definition that one conditions on "all the available information in the universe" and he was of course aware of the fact that in practice inference is only based on the available information contained in the process under study. If the effects of a confounder on two variables are lagged in time the resulting association between lagged variables falsely leads to a directed edge between the corresponding vertices in the causality graph. However in certain situations the global causal Markov properties can be used to distinguish between a direct causal and a confounding effect. As an example we consider the human tremor data in Section 5.2. The causality graph in Fig. 10 shows a

direct feedback between the two EEG channels P1 and C2P. The two channels might also be influenced by a third region in the brain. On the other hand a bivariate analysis of P1 and the extensor shows that the latter is still causal for the former. This contradicts the assumption of a confounder which would imply that the hand movements are noncausal for the cortical activity in channel P1. Therefore C2P must have a direct causal effect on P1 but we cannot exclude that part of the association is due to a confounder. For more details we refer to Section 4 of Eichler (2000).

Second, the identification of nonlinear causal effects is hampered by the restriction to linear methods. Here, graphs defined in terms of conditional independence or strong Granger causality provide an alternative capturing the complete dependence structure of the process. The results in Sections 2 and 3 also hold for these graphs with one notable exception: The pairwise and the global Markov properties for the TSC- and the causality graphs in general are no longer equivalent but only under additional assumption on the process. Alternatively the graphs can be defined in terms of the so-called blockrecursive Markov properties which are still equivalent to the global ones. For details we refer to Section 5 of Eichler (2000) and Eichler (2001). However, we mention that it is hardly possible to make inference for arbitrary nonlinear causal relationships since strictly speaking we have only one observation of the multivariate process. This problem could only be overcome by assuming specific nonlinear time series models.

A third limitation arises from the fact that the measured variables typically evolve continuously over time but are observed only at discrete times. If the time between a cause and its effect is shorter than the sampling interval then the corresponding discretized component series are only contemporaneously correlated and a directed edge is detected falsely as an undirected one. (We note that contemporaneous causality may also be due to confounding by an unmeasured process.) For the same reasons an indirect causality mediated by measured variables may lead to a directed edge in the causality graph.

The third problem can be avoided by replacing the discrete version of Granger causality by local Granger causality and local instantaneous causality which have been introduced by Comte and Renault (1996) for the investigation of causal relations between time-continuous processes. We note that local Granger causality is related to the notion of local independence (Schweder 1970; Aalen 1987) which has been used by Didelez (2001) for the definition of directed graphs for marked point processes.

Finally we note that the concept of Granger causality is not restricted to stationary time series but also applies to more general situations such as nonstationary time series, panels of time series, or point processes. We conjecture that the graphical modelling approach can be applied in these cases in a similar way.

Bibliography

Aalen, O.O. (1987). Dynamic modelling and causality. *Scandinavian Actuarial Journal*, 177-190.

Aigner, D.J. and Zellner, A. (eds.) (1988). Causality. *Journal of Econometrics*, 39, No.1/2, North Holland, Amsterdam, p. 1-234.

Andersson, S.A., Madigan, D., and Perlman, M.D. (2001). Alternative Markov properties for chain graphs. *Scandinavian Journal of Statistics*, 28, 33-85.

Brillinger, D.R. (1981). *Time Series: Data Analysis and Theory*, McGraw Hill, New York.

Comte, F. and Renault, E. (1996). Noncausality in continuous time models. *Econometric Theory*, 12, 215-256.

Dahlhaus, R. (2000). Graphical interaction models for multivariate time series. *Metrika*, 51, 157-172.

Dahlhaus, R., Eichler, M., and Sandkühler, J. (1997). Identification of synaptic connections in neural ensembles by graphical models. *Journal of Neuroscience Methods*, 77, 93-107.

Dahlhaus, R. and Eichler, M. (2001). Statistical inference for time series chain graphs. In preparation.

Dawid, A.P. (2000). Causal inference without counterfactuals. *Journal of the American Statistical Association.*, 86, 9-26.

Dempster, A.P. (1972). Covariance selection. *Biometrics*, 28, 157-175.

Eichler, M. (1999). *Graphical Models in Time Series Analysis*. Doctoral thesis, University of Heidelberg, Germany.

Eichler, M. (2000). Granger-causality graphs for multivariate time series. Technical report, University of Heidelberg, Germany.

Eichler, M. (2001). Markov properties for graphical time series models. Technical report, University of Heidelberg, Germany.

Eichler, M., Dahlhaus, R., and Sandkühler, J. (2000). Partial correlation analysis for the identification of synaptic connections. Technical report, University of Heidelberg, Germany.

Frydenberg, M. (1990). The chain graph Markov property. *Scandinavian Journal of Statistics*, 17, 333-353.

Gather, U., Imhoff, M., and Fried, R. (2002). Graphical models for multivariate time series from intensive care monitoring. *Statistics in Medicine*, to appear.

Geweke, J. (1984). Inference and causality in economic time series. In Z. Griliches and M.D. Intriligator (eds.), *Handbook of Econometrics*, Vol. 2, North-Holland, Amsterdam, 1101-1144.

Granger, C.W.J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37, 424-438.

Haavelmo, T. (1943). The statistical implications of a system of simultaneous equations. *Econometrica*, 11, 1-12.

Hsiao, C. (1982). Autoregressive modeling and causal ordering of econometric variables. *Journal of Economic Dynamics and Control*, 4, 243-259.

Lauritzen, S.L. and Wermuth, N. (1989). Graphical models for association between variables, some of which are qualitative and some are quantitative. *Annals of Statistics*, 17, 31-57.

Lauritzen, S.L. (2000). Causal inference from graphical models. In E. Barndorff-Nielsen, D.R. Cox, and C. Klüppelberg (eds.) *Complex Stochastic Systems*, CRC Press, London

Levitz, M., Perlman, M.D., and Madigan, D. (2001). Separation and completeness properties for AMP chain graph markov models. *Annals of Statistics*, 29, 1751 - 1784.

Lütkepohl, H. (1993). *Introduction to multiple time series analysis*. Springer, Berlin.

Lynggaard, H., and Walther, K.H. (1993). *Dynamic Modelling with Mixed Graphical Association Models*. Master's Thesis, Aalborg University.

Neyman, J. (1923). Statistical problems in agricultural experimentation (with discussion). *Journal of the Royal Statistical Society Suppl.*, 2, 107-180.

Pearl, J. (1988). *Probabilistic Inference in Intelligent Systems.* Morgan Kaufmann, San Mateo, California.

Pearl, J. (1995). Causal diagrams for empirical research (with discussion). *Biometrika*, 82, 669-710.

Pearl, J. (2000). *Causality*, Cambridge University Press, Cambridge, UK.

Robins, J. (1986). A new approach to causal inference in mortality studies with sustained exposure periods - application to control of the healthy worker survivor effect. *Mathematical Modelling*, 7, 1393-1512.

Rubin, D.B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66, 688-701.

Sameshima, K. and Baccalá, L.A. (1999). Using partial directed coherence to describe neuronal ensemble interactions. *Journal of Neuroscience Methods*, 94, 93-103.

Schweder, T. (1970). Composable Markov processes. *Journal of Applied Probability*, 7, 400-410.

Seinfeld, J.H. (1986). *Atmospheric Chemistry and Physics of Air Pollution.* John Wiley, Chichester.

Timmer, J., Lauk, M., Köster, B., Hellwig, B., Häußler, S., Guschlbauer, B., Radt, V., Eichler, M., Deuschl, G., and Lücking, C.H. (2000). Cross-spectral analysis of tremor time series. *Int. J. of Bifurcation and Chaos*, 10, 2595-2610.

Tjøstheim, D. (1981). Granger-causality in multiple time series. *Journal of Econometrics*, 17, 157-176.

Wright, S. (1921). Correlation and causation. *Journal of Agricultural Research*, 20, 557-585.