GRAPHICAL INTERACTION MODELS FOR MULTIVARIATE TIME SERIES[1]

BY R. DAHLHAUS

Universität Heidelberg

Abstract

In this paper we extend the concept of graphical models for multivariate data to multivariate time series. We define a partial correlation graph for time series and use the partial spectral coherence between two components given the remaining components to identify the edges of the graph. As an example we consider multivariate autoregressive processes. The method is applied to air pollution data.

1 Introduction

Graphical interaction models have become an important tool for analyzing multivariate data - for an introduction to the topic, the basic notations and an overview of the different methods see the recent monographs by Cox and Wermuth (1996), Edwards (1995), Lauritzen (1996) and Whittaker (1990). In this paper we extend the concept of undirected conditional independence graphs to multivariate time series. The edges of a conditional independence graph reflect the conditional dependence structure between several variables and give the data analyst an idea of the interaction structure of the observed variables. In particular, it helps to discriminate between direct and indirect correlations between the variables.

In this paper we discuss the concept of graphical models for multivariate time series. The vertex set will consist of the components of the series while the edges will reflect the partial correlation structure of the components given the others. Thus, for Gaussian time series our graph is a conditional independence graph (a generalisation of a concentration graph/covariance selection model) while for non-Gaussian time series it is termed partial correlation graph.

For the characterisation of the edges the partial spectral coherence is used - a standard tool in the frequency domain analysis of time series (cf. Brillinger, 1981, Chapter 8.3 and the references therein). It is a measure for the dependence between two time series after removing the linear time invariant effects of a third (or more) series. The partial spectral coherence has been used for time series by Gersch (1972) in electrophysiological signal analysis and for point processes by Brillinger, Bryant and Segundo (1976) for the identification of synaptic interactions of neurons. Graphical models for time series based on the partial spectral coherence have also been defined by Brillinger (1996).

The paper is organized as follows. We discuss graphical models for time series. In particular, we prove a property of the inverse of the spectral matrix which helps to identify the graph. In Section 3 we prove a separation theorem for time series graphs which is equivalent to the global Markov property of the graph. In Section 4 we consider as a specific example multivariate autoregressive processes. Section 5 contains as an example the analysis of air pollution data.

2 Graphical models for time series

A graph $G = (V, E)$ consists of a set of vertices V, say $V = \{1, \ldots, k\}$ and a set of edges $E \subset \{(a, b) \in V \times V\}$. We only consider undirected graphs, i.e.

2

we assume $(a,b) \in E$ whenever $(b,a) \in E$.

Suppose $\boldsymbol{X}(t) = (X_1(t), \ldots, X_k(t))'$, $t \in \mathbb{Z}$ is a multivariate stationary time series. As the vertices of our graph we want to have the components of the series, i.e. we set $V = \{1, \ldots, k\}$. The basic idea is that an edge (a,b) is missing if the components $X_a(\cdot)$ and $X_b(\cdot)$ are uncorrelated given the other components of the series. To come to an exact definition we now make this idea rigorous. Note, that in the following definitions we are dealing with the stochastic properties of the process and not with empirical values based on observations.

Let $Y_{ab}(t) = (X_j(t), j \neq a, b)$. We remove the linear effects of Y_{ab} from $X_a(t)$ by determining the optimal μ_a and the optimal $1 \times (r-2)$ filter $\{d_a(u)\}$ such that

$$E \left(X_a(t) - \mu_a - \sum_u d_a(t-u) Y_{ab}(u) \right)^2$$

is minimal. The remainder is denoted by $\varepsilon_a(t)$, i.e.

$$\varepsilon_a(t) := \varepsilon_{a|\{a,b\}^c}(t) := X_a(t) - \mu_a^{\text{opt}} - \sum_u d_a^{\text{opt}}(t-u) Y_{ab}(u).$$

In the same way we define

$$\varepsilon_b(t) := \varepsilon_{b|\{a,b\}^c}(t) := X_b(t) - \mu_b^{\text{opt}} - \sum_u d_b^{\text{opt}}(t-u) Y_{ab}(u).$$

We now set $\mathcal{X}_a = (X_a(t); t \in \mathbb{Z})$ and $\mathcal{Y}_{ab} = (Y_{ab}(t); t \in \mathbb{Z})$ and define the relation

$$\mathcal{X}_a \perp\!\!\!\perp \mathcal{X}_b | \mathcal{Y}_{ab} :\Leftrightarrow \text{cov}\big(\varepsilon_{a|\{a,b\}^c}(t), \varepsilon_{b|\{a,b\}^c}(t+u)\big) = 0 \quad \text{for all } u \in \mathbb{Z} \qquad (2.1)$$

leading to the definition of a partial correlation graph.

(2.1) Definition Let $\mathbf{X}(t) = (X_1(t), \ldots, X_k(t))'$ be a multivariate stationary time series and $V = \{1, \ldots, k\}$ the corresponding set of vertices. Let $(a,b) \notin E$ if and only if $\mathcal{X}_a \perp\!\!\!\perp \mathcal{X}_b | \mathcal{Y}_{ab}$. Then $G = (V, E)$ is called a *partial correlation graph* for time series.

3

We might also use the notation concentration graph instead of partial correlation graph - in particular since the graph is a generalisation of an ordinary concentration graph. However, we chose the above name since it better describes the nature of the graph (Furthermore, the inverse spectral matrix which characterizes the edges of the graph - see Theorem 2.4 below - has also never been termed spectral concentration matrix).

For Gaussian time series, ε_a and ε_b defined above are again Gaussian and the above best linear predictor of $X_a(t)$ given $Y_{ab}(t)$ is the best predictor. Furthermore, $\mathcal{X}_a \perp\!\!\!\perp \mathcal{X}_b | \mathcal{Y}_{ab}$ if and only if ε_a and ε_b are independent. In this situation the above graph is a conditional independence graph for time series. Note, that the orthogonality relation $\perp\!\!\!\perp$ defined in (2.1) can be retained if $X_a(t)$ and $X_b(t)$ are vector time series.

An important characterisation of the edges of the graph can be obtained from the partial spectral coherence. An estimate of the partial spectral coherence will also be of importance in identifying the graph from an observed time series. Let

$$c_{ab}(u) = c_{X_a X_b}(u) = \text{cov}\big(X_a(t+u), X_b(t)\big)$$

be the covariance function of the process. If

$$\sum_{u=-\infty}^{\infty} |c_{ab}(u)| < \infty$$

then the (cross-)spectrum between $X_a(t)$ and $X_b(t)$ is defined by

$$f_{ab}(\lambda) = f_{X_a X_b}(\lambda) = \frac{1}{2\pi} \sum_{u=-\infty}^{\infty} c_{ab}(u) \exp(-i\lambda u).$$

Let $f_{XX}(\lambda) = (f_{ab}(\lambda))_{a,b=1,\ldots,k}$ and $c_{XX}(u) = (c_{ab}(u))_{a,b=1,\ldots,k}$. The Fourier inversion formula gives

$$c_{XX}(u) = \int_{-\pi}^{\pi} f_{XX}(\lambda) \exp(i\lambda u) d\lambda.$$

4

A similar Fourier-representation for the process itself holds (cf. Brillinger, 1981, Theorem 4.6.2).

If the components $X_a(t + u)$ and $X_b(t)$ are uncorrelated at all lags u then $f_{ab}(\lambda) = 0$ (and vice versa). Otherwise $f_{ab}(\lambda)$ contains information on the dependence structure, decomposed into different frequencies. $\arg f_{ab}(\lambda)$ is a measure for the time delay of the two signals, also decomposed into different frequency components (cf. Brillinger, 1981 Chapter 6 and 7).

A measure of the dependence between $X_a(t)$ and $X_b(t)$ given $Y_{ab}(t)$ is the partial cross-spectrum of $X_a(t)$ and $X_b(t)$ given $Y_{ab}(t)$

$$f_{X_a X_b | Y_{ab}}(\lambda) := f_{\varepsilon_a \varepsilon_b}(\lambda)$$

where $\varepsilon_a(t)$ and $\varepsilon_b(t)$ are as above. Rescaling leads to the partial spectral coherence

$$R_{X_a X_b | Y_{ab}}(\lambda) := \frac{f_{X_a X_b | Y_{ab}}(\lambda)}{\left[f_{X_a X_a | Y_{ab}}(\lambda) f_{X_b X_b | Y_{ab}}(\lambda) \right]^{1/2}}. \tag{2.2}$$

The solution of the above optimization problem can be found together with the form of the partial cross-spectrum in Brillinger (1981, Theorem 8.3.1). Brillinger proves that

$$f_{X_a X_b | Y}(\lambda) = f_{X_a X_b}(\lambda) - f_{X_a Y}(\lambda) f_{YY}(\lambda)^{-1} f_{Y X_b}(\lambda). \tag{2.3}$$

Since $f_{\varepsilon_a \varepsilon_b}(\cdot) \equiv 0$ if and only if $\mathrm{cov}(\varepsilon_a(t), \varepsilon_b(t + u)) = 0$ for all $u \in \mathbb{Z}$ we obtain the following result.

(2.2) Proposition *Suppose $G = (V, E)$ is a partial correlation graph for a multivariate time series. Then*

$$(a, b) \notin E \quad \text{if and only if} \quad R_{X_a X_b | Y_{ab}}(\cdot) \equiv 0.$$

(2.3) Remark The orthogonality relation $\perp\!\!\!\perp$ can be defined in the same way for arbitrary vector time series $\mathcal{X}_A := (X_a(t); a \in A, t \in \mathbb{Z})$. In particular we

5

have

$$\mathcal{X}_A \perp\!\!\!\perp \mathcal{X}_B | \mathcal{X}_C \quad \Leftrightarrow \quad \mathrm{cov}(\varepsilon_{A|C}(t), \varepsilon_{B|C}(t+u)) = 0 \text{ for all } u \in \mathbb{Z}$$
$$\Leftrightarrow \quad f_{X_A X_B | X_C}(\cdot) \equiv 0$$
$$\Leftrightarrow \quad R_{X_A X_B | X_C}(\cdot) \equiv 0 \qquad (2.4)$$

for disjoint sets $A, B, C \subset V$. Furthermore, (2.2) and (2.3) stay the same in the vector case

We now prove that the partial spectral coherences can be obtained as the negative values of the rescaled inverse of the spectral matrix. Let

$$g(\lambda) := f_{XX}(\lambda)^{-1}$$

and

$$d(\lambda) := \begin{pmatrix} g_{11}(\lambda)^{-1/2} & & 0 \\ & \ddots & \\ 0 & & g_{kk}(\lambda)^{-1/2} \end{pmatrix} g(\lambda) \begin{pmatrix} g_{11}(\lambda)^{-1/2} & & 0 \\ & \ddots & \\ 0 & & g_{kk}(\lambda)^{-1/2} \end{pmatrix}.$$

(2.4) Theorem *Suppose $\boldsymbol{X}(t) = (X_a(t); a \in V)$ is a multivariate time series with spectral density matrix $f_{XX}(\lambda)$. If $f_{XX}(\lambda)$ has full rank, we have with the above notations*

$$d_{ab}(\lambda) = -R_{X_a X_b | Y_{ab}}(\lambda).$$

Furthermore,

$$g_{aa}(\lambda) = 1/f_{X_a X_a | Y_a}(\lambda)$$

where $Y_a(t) = (X_j(t) | j \neq a)$. If $X(t) = (X_1(t), ..., X_k(t))'$ and the $X_A(t)$ are vector processes then $g_{AB}(\lambda) = 0$ (as a matrix) if and only if $f_{X_A X_B | Y_{AB}}(\lambda) = 0$.

6

PROOF. The proof is given in the appendix $\qquad\qquad\square$

Theorem 2.4 has important consequences. First, with view to Proposition 2.2 it means that the missing edges in the partial correlation graph can uniquely be identified from zeroes in the rescaled inverse of the spectral matrix (similar to concentration graphs/covariance selection models where missing edges are characterized by zeroes of the concentration matrix (inverse covariance matrix) - cf. Lauritzen, 1996, Chapter 5). This characterizes for example the restrictions on the parameter space for parametric time series models that have a certain graph (cp. (4.1) below for multivariate autoregressive models). This is for example important for likelihood ratio tests of graphical models.

Second, Theorem 2.4 is the basis for a nonparametric identification of the time series graph where an estimate of the spectral matrix is inverted and rescaled (see Section 5). This is much less computerintensive than e.g. the estimation of the residuals $\varepsilon_{a|\{a,b\}c}(t)$ used in the definition of the graph which would require the calculation of $2\binom{k}{2}$ linear filters.

For certain applications it may be interesting to study the situation where $d_{ab}(\lambda) = 0$ for certain frequency bands leading to the idea of a frequency dependent graph. However, we do not want to pursue this any further.

(2.5) Remark The inverse of the spectral matrix may also be used to explain the effect of a confounder, i.e. an unobserved component X_c of the time series. Suppose $X(t) = (X_1(t), \ldots, X_k(t))'$ is the full time series including $X_c(t), g(\lambda) := f_{XX}(\lambda)^{-1}$, $\tilde{X}(t) = (X_j(t); \ j \neq c)'$ is the observed reduced series and $\tilde{g}(\lambda) := f_{\tilde{X}\tilde{X}}(\lambda)^{-1}$. Then we obtain with some matrix calculations (see also the expressions for B^{-1} in the appendix)

$$\tilde{g}_{ab}(\lambda) = g_{ab}(\lambda) - g_{ac}(\lambda)g_{cc}(\lambda)^{-1}g_{cb}(\lambda)$$

explaining the relations between edges in the full graph $(g_{ab}(\lambda) \not\equiv 0)$ and edges in the reduced graph $(\tilde{g}_{ab}(\lambda) \not\equiv 0)$. In particular, if there is no edge

7

between a and b ($g_{ab}(\lambda) \equiv 0$) but edges between a and c and c and b with $g_{ac}(\lambda) \neq 0$ and $g_{cb}(\lambda) \neq 0$ for the same λ this causes an edge between a and b in the reduced graph (note that it can be shown that $0 < f_{X_c X_c}(\lambda) < \infty$ implies $0 < g_{cc}(\lambda) < \infty$).

We conclude this section with a remark on the possibilities of including ordinary variables into the graph.

(2.6) Remark (Joint graphs for time series and ordinary variables)
In the 'mixed case' where one has time series and ordinary variables we have to consider two different cases:

1. One observes the ordinary variables independently at each time point where one observes the time series. In that case the ordinary variables may be treated as a time series consisting of *iid* observations and having a constant spectrum. This situation is only a special case of the situation considered above. If e.g. all variables are ordinary, then $f_{XX}(\lambda)$ is constant with $2\pi f_{XX}(\lambda)$ being the variance covariance matrix of the observations. In some sense covariance selection models may therefore be regarded as a special case of the time series graph above.

2. With each time series one only has one observation of the ordinary variable (e.g one investigates the blood and an EEG-recording of a patient). In this case the time series may be regarded as one (multivariate) variable leading to a classical graphical model. Usually one would even summarize the information contained in the time series into one (or several) variables (e.g. for EEG data one might consider a variable which reflects the occurrence of an epileptic seizure).

8

3 Markov properties of time series graphs

In this chapter we prove the separation theorem for time series graphs. This is equivalent to the global Markov property.

(3.1) Lemma *Suppose $\boldsymbol{X}(t) = (X_a(t); a \in V)$ is a multivariate time series whose spectral matrix $f_{XX}(\lambda)$ has full rank (for all λ) and $\mathcal{X}_A = (X_a(t); a \in A, t \in \mathbb{Z})$, $\mathcal{X}_B$, $\mathcal{X}_C$ and $\mathcal{X}_D$ are vector time series with disjoint sets $A, B, C, D \subset V$. Then we have*
(i) $\mathcal{X}_A \perp\!\!\!\perp (\mathcal{X}_B, \mathcal{X}_C) | \mathcal{X}_D$ implies $\mathcal{X}_A \perp\!\!\!\perp \mathcal{X}_B | \mathcal{X}_D$,
(ii) $\mathcal{X}_A \perp\!\!\!\perp (\mathcal{X}_B, \mathcal{X}_C) | \mathcal{X}_D$ if and only if $\mathcal{X}_A \perp\!\!\!\perp \mathcal{X}_B | (\mathcal{X}_C, \mathcal{X}_D)$ and $\mathcal{X}_A \perp\!\!\!\perp \mathcal{X}_C | (\mathcal{X}_B, \mathcal{X}_D)$.

PROOF. (i) From (2.4) we get $f_{X_A(X_B, X_C)|X_D} \equiv 0$ which implies $f_{X_A X_B | X_D} \equiv 0$, i.e. $\mathcal{X}_A \perp\!\!\!\perp \mathcal{X}_B | \mathcal{X}_D$.

(ii) Consider the matrix $g(\lambda)$ from Theorem 2.4. If $\mathcal{X}_A \perp\!\!\!\perp \mathcal{X}_B | (\mathcal{X}_C, \mathcal{X}_D)$ and $\mathcal{X}_A \perp\!\!\!\perp \mathcal{X}_C | (\mathcal{X}_B, \mathcal{X}_D)$ then Theorem 2.4 implies $g_{AB}(\cdot) \equiv 0$ and $g_{AC}(\cdot) \equiv 0$. If we set $Y_A \equiv X_A, Y_B = (X_B, X_C)$ and $Y_C \equiv X_D$ then $g(\cdot) \equiv g^Y(\cdot)$ and Theorem 2.4 implies $\mathcal{X}_A \perp\!\!\!\perp \mathcal{X}_C | (\mathcal{X}_B, \mathcal{X}_D)$. The other direction follows in the same way. $\square$

(3.2) Remark For ordinary graphical models assertion (ii) only holds under additional assumptions, for example under the condition that the joint density of the variables involved is positive everywhere (cf. Lauritzen, 1996, Proposition 3.1). For time series graphs also additional assumptions are needed for assertion (ii) to hold. As an example consider the case where $A = \{a\}, B = \{b\}, C = \{c\}, D = \{d\}$, $X_b(t) = X_c(t), X_a(t) = X_b(t) + \varepsilon(t)$ where $\varepsilon(t)$ is an iid sequence independent of the other components and $X_d(t)$ is also independent of all other components. Then $\mathcal{X}_A \perp\!\!\!\perp \mathcal{X}_B | (\mathcal{X}_C, \mathcal{X}_D)$ and $\mathcal{X}_A \perp\!\!\!\perp \mathcal{X}_C | (\mathcal{X}_B, \mathcal{X}_D)$ but we do not have $\mathcal{X}_A \perp\!\!\!\perp (\mathcal{X}_B, \mathcal{X}_C) | \mathcal{X}_D$. In this case the spectral matrix is singular since the columns b and c are identical.

9

We now establish the separation theorem for a time series graph (V, E). For $A, B, S \subset V$ we say that S separates A and B if every *path* from an element from A to an element from B contains at least one vertex from the separating set S (a path is a sequence of distinct vertices $i_1, \ldots, i_m$ with $(i_\ell, i_{\ell+1}) \in E$ for each $\ell = 1, \ldots, m - 1$).

(3.3) Theorem *Suppose $\boldsymbol{X}(t)$ is a multivariate time series with everywhere regular spectral matrix and corresponding partial correlation graph (V, E). Let $A, B, S \subset V$ where S separates A and B. Then we have*

$$\mathcal{X}_A \perp\!\!\!\perp \mathcal{X}_B | \mathcal{X}_S.$$

PROOF. With the assertions of Lemma 3.1 (i) (ii) the result can be proved in exactly the same way as Theorem 3.7 in Lauritzen (1996). □

(3.4) Remark (Markov properties) The relation $\mathcal{X}_A \perp\!\!\!\perp \mathcal{X}_B | \mathcal{X}_C$ constitutes a range of different Markov properties (cf. Lauritzen, 1996, Section 3.2.1). Without repeating all the definitions we remark that the separation theorem implies that the global Markov property holds for a time series graph.

(3.5) Example Let $X_1(t) = a_1 X_1(t - 1) + \varepsilon_1(t)$,

$$X_j(t) = a_j X_j(t - 1) + b_j X_{j-1}(t - t_j) + \varepsilon_j(t) \qquad (j = 2, 3, 4)$$

with some time lags $t_j \in \mathbb{N}_0$. The $\varepsilon_j(t)$ are assumed to be *iid* $\mathcal{N}(0, \sigma^2)$. This means that each process depends on its own past and its predecessor with some time lag t_j. Then all processes are correlated while the conditional correlation graph is

This follows from the next section where the more general class of multivariate autoregressive processes is treated. The separation theorem gives for example in this situation $\mathcal{X}_1 \perp\!\!\!\perp \mathcal{X}_4 | \mathcal{X}_2$.

10

4 Vector autoregressive processes

The class of vector autoregressive processes provides an intuitive example for time series graphs. Let

$$\mathbf{X}(t) = \sum_{j=1}^{p} \Phi_j \mathbf{X}(t-j) + \mathbf{Z}(t)$$

where the Φ_j are $k \times k$ matrices and $\mathbf{Z}(t)$ are $iid \, \mathcal{N}(0, \Sigma)$. Let

$$\Phi(z) := I - \Phi_1 z - \ldots - \Phi_p z^p$$

be the characteristic polynomial of the process. If $\det \Phi(z) \neq 0$ for all $z \in \mathbb{C}$ with $|z| \leq 1$ then the above recursion has a stationary solution (cf. Brockwell and Davis, 1987, Theorem 11.3.1).

In the above model the components $\Phi_{j,ab}$ may intuitively be regarded as the 'influence' from $X_b(t-j)$ on $X_a(t)$, that is we have no influence from component b on a if $\Phi_{ab}(\cdot) \equiv 0$.

The spectral density matrix of $\mathbf{X}(t)$ is (cf. Brockwell and Davis, 1987, Example 11.8.1)

$$f_{XX}(\lambda) = \frac{1}{2\pi} \Phi^{-1}(e^{-i\lambda}) \Sigma \Phi^{-1}(e^{i\lambda})',$$

i.e. we have

$$g(\lambda) = f_{XX}(\lambda)^{-1} = 2\pi \Phi(e^{i\lambda})' \Sigma^{-1} \Phi(e^{-i\lambda}).$$

Suppose now for simplicity $\Sigma = \sigma^2 I_k$. Then we have

$$g_{ab}(\lambda) = \frac{2\pi}{\sigma^2} \sum_{c=1}^{k} \Phi_{ca}(e^{i\lambda}) \Phi_{cb}(e^{-i\lambda}) \tag{4.1}$$

If a and b do not 'influence' jointly another component c then we have $g_{ab}(\cdot) \equiv 0$ if and only if $\Phi_{ab}(\cdot) \equiv 0$ and $\Phi_{ba}(\cdot) \equiv 0$, i.e. with view to Theorem 2.4

11

we obtain the result we would expect. In particular this proves the assertion from Example 3.5.

The above restriction seems to be strange from a first view. The following example demonstrates that this is a natural restriction: Suppose X_1, X_2 and ε are independent and $X_3 = X_1 + X_2 + \varepsilon$. Then X_1 and X_2 are no longer independent conditional on X_3 and the conditional correlation graph will show a connection between X_1 and X_2 contrary to the intuition. In Dahlhaus, Eichler and Sandkühler (1997) we have studied this effect more detailed in the context of point processes.

The situation is very similar to ordinary graphical models where the problem of equivalence of directed and undirected graphs arises (cf. Whittaker, 1990, Chapter 3.5). Both graphs have the same independence interpretations if the directed graph satisfies the so called *Wermuth condition* which forbids that non-connected variables have 'influence' on the same variable. For an arbitrary directed graph the corresponding (undirected) conditional independence graph is a subgraph of the so called *moral graph* where 'parents are married' (cf. Whittaker, 1990, Chapter 3.9; Wermuth, 1980). Relation (4.1) implies that the same holds in the above situation if an influence from b to a ($\Phi_{ab}(\cdot) \not\equiv 0$) is represented by a directed edge (b, a) in a directed graph.

We hesitate to set down the definitions of directed graphs, moral graphs etc. indicated above in a rigorous way for time series. The reason is that we feel that in a careful (and meaningful) definition of 'direction' of time series graphs time should play a major role (e.g. in the model $X_2(t) = bX_1(t-1) + \varepsilon(t)$ the direction should be from X_1 to X_2 due to the time lag). Such a definition therefore requires more and deeper considerations. In particular it will not be totally analogous to the definition of ordinary directed graphs.

5 Identification of the time series graph for air pollution data

Given a k-dimensional multivariate time series of length n the problem arises how to identify the corresponding graph. By Proposition 2.2 the presence of an edge is equivalent to a non-vanishing partial spectral coherence. Therefore, we build the graph by testing whether the different spectral coherences disappear.

For an estimate of the partial spectral coherence we use the characterisation from Theorem 2.4:

We nonparametrically estimate the spectral matrix and invert and rescale this estimate. As an estimator for $f_{ab}(\lambda)$ we take

$$\hat{f}_{ab}(\lambda) = \frac{2\pi}{bT} \sum_s W\left(\frac{\lambda - (2\pi s)/T}{b}\right) I_{ab}^{(T)}\left(\frac{2\pi s}{T}\right)$$

where W is a kernel with $\int_{-\pi}^{\pi} W(\alpha)d\alpha = 1$ and

$$I_{ab}^{(T)}(\lambda) = \left\{2\pi \sum_{t=1}^{T} h\left(\frac{t}{T}\right)\right\}^{-2} \left\{\sum_{t=1}^{T} h\left(\frac{t}{T}\right)\left(X_a(t) - \bar{X}_a\right)\exp(-i\lambda t)\right\}$$
$$\times \left\{\sum_{t=1}^{T} h\left(\frac{t}{T}\right)\left(X_b(t) - \bar{X}_b\right)\exp(i\lambda t)\right\}$$

is the tapered periodogram.

The method was used to analyze a 5-dimensional time series of length 4386 of air pollutants recorded from January 1991 to December 1992 in Heidelberg (6 equidistant recordings a day). 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 gri which plays a major role in these reactions, in particular in the generations of ozone.

The original data were recorded with a distance of 30 minutes. Figure 5.1 shows the daily course of the five variables averaged over 61 consecutive days in summer. CO and NO increase early in the morning due to traffic and, as
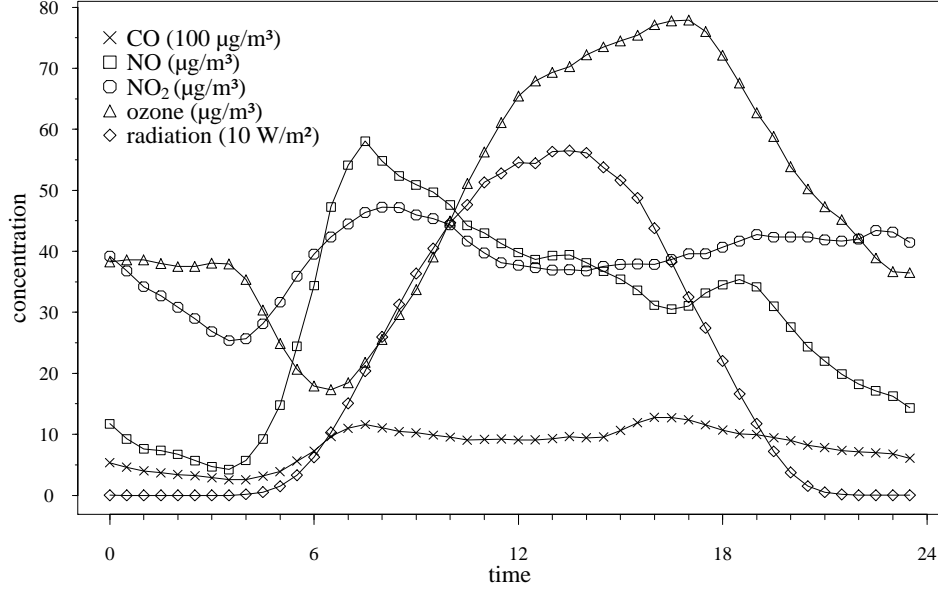
13

Figure 5.1: Average of the daily measurements of CO, NO, NO$_2$, O$_3$ and *gri* over 61 days in summer.

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 5.1 indicates that all variables are correlated at different lags.

Beside the original series we will also analyze the residual series after subtracting the (local) average course as shown in Figure 5.1 (trend corrected data). The original series contained a few missing values (less than 2%) which were completed by interpolation of the residual series with splines. For the final investigation each 8th value of the original series was taken (i.e. 6 values per day).

Figure 5.2 shows above the diagonal the squared empirical coherences $|\hat{R}_{X_a X_b}(\lambda)|^2$ of the original series with

$$\hat{R}_{X_a X_b}(\lambda) := \frac{\hat{f}_{ab}(\lambda)}{[\hat{f}_{aa}(\lambda)\hat{f}_{bb}(\lambda)]^{1/2}}.$$
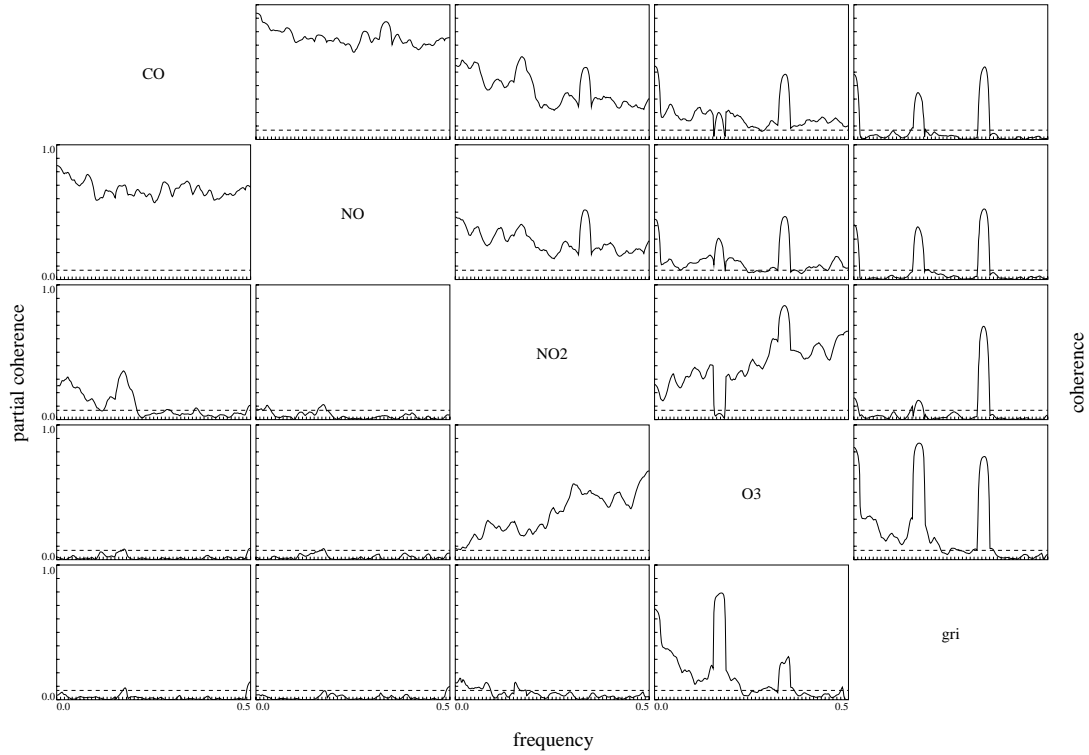
14

Figure 5.2: Spectral coherences (above diagonal) and partial spectral coherences (below diagonal) for air pollution data.

These plots show strong dependencies between all of the variables. The peak at frequency $\pi/3$ and its harmonic at $(2\pi)/3$ corresponds to the periodic behaviour of length 6 (one day). Below the diagonal we have plotted the partial coherences $|\hat{R}_{X_a X_b | Y_{ab}}(\lambda)|^2$ as defined in Section 2. The dashed line in the plots is a 95%-test bound (as described below) for the supremum of the estimates under the hypothesis $R_{X_a X_b | Y_{ab}}(\cdot) \equiv 0$.

The test indicates that some of the processes are uncorrelated given the remaining processes. Although the test bound was slightly exceeded at a few frequencies we found it reasonable to draw the time series graph as in Figure
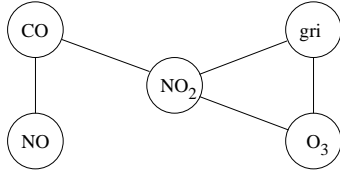
15

Figure 5.3: Estimated partial correlation graph for air pollution data.

5.3 from the partial spectral coherences.

The graph correctly reflects the creation of O_3 from NO_2 and the fact that the global radiation intensity plays a major role in the process of O_3-generation. Furthermore, CO and NO are correlated (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 and still not completely understood (cf. the monograph on this topic by Seinfeld, 1986). 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.

Part of these reactions can be explained by a photochemical theory (cf. Seinfeld, 1986, Section 4.2). This theory is confirmed by the above graph: the edge between *gri* and NO_2 represents the photolysis of NO_2 and the edge between CO and NO_2 supports that most of NO_2 is generated via a radical reaction where CO is involved. A bit surprising is the missing edge between NO and NO_2. This missing edge and the edge between CO and NO indicate that mainly the concentration of CO (and not of NO) is responsible for the generation of NO_2. This means in particular that NO_2 is generated via a radical reaction (where CO plays a major role) and not in a direct reaction (where CO is not involved). It is remarkable that a direct correlation analysis indicates the opposite since the spectral coherence between NO and NO_2 in Figure 5.2 is highly significant.
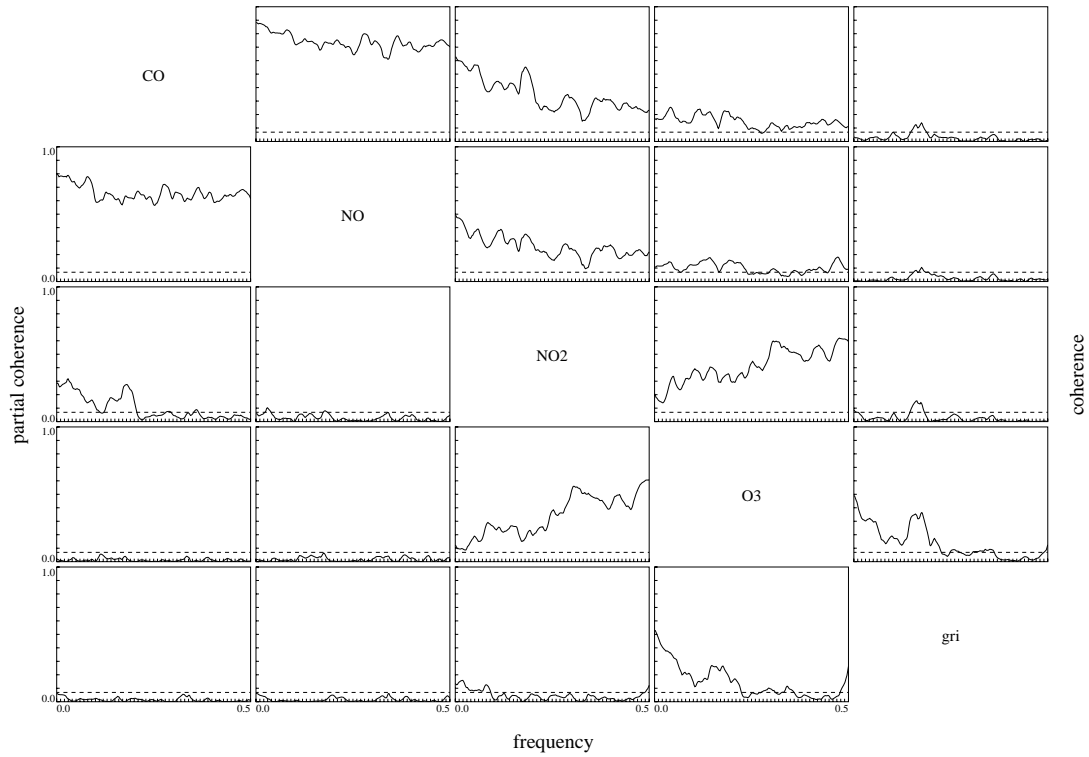
16

Figure 5.4: Spectral coherences (above diagonal) and partial spectral coherences (below diagonal) for the trend corrected air pollution data.

One may raise the question to what extent the above findings are due to the daily up and down of the five variables. For this reason we have repeated the analysis with the residual series (obtained as described above). The plots in Figure 5.4 show that the analysis of the residual series leads exactly to the same graph.

Figure 5.5 shows above the diagonal the cross correlations and below the diagonal the partial cross correlations of the trend corrected series (i.e. estimates of the correlations used in (2.1)). These partial cross correlations have been calculated by the inverse Fourier transform of estimates of the
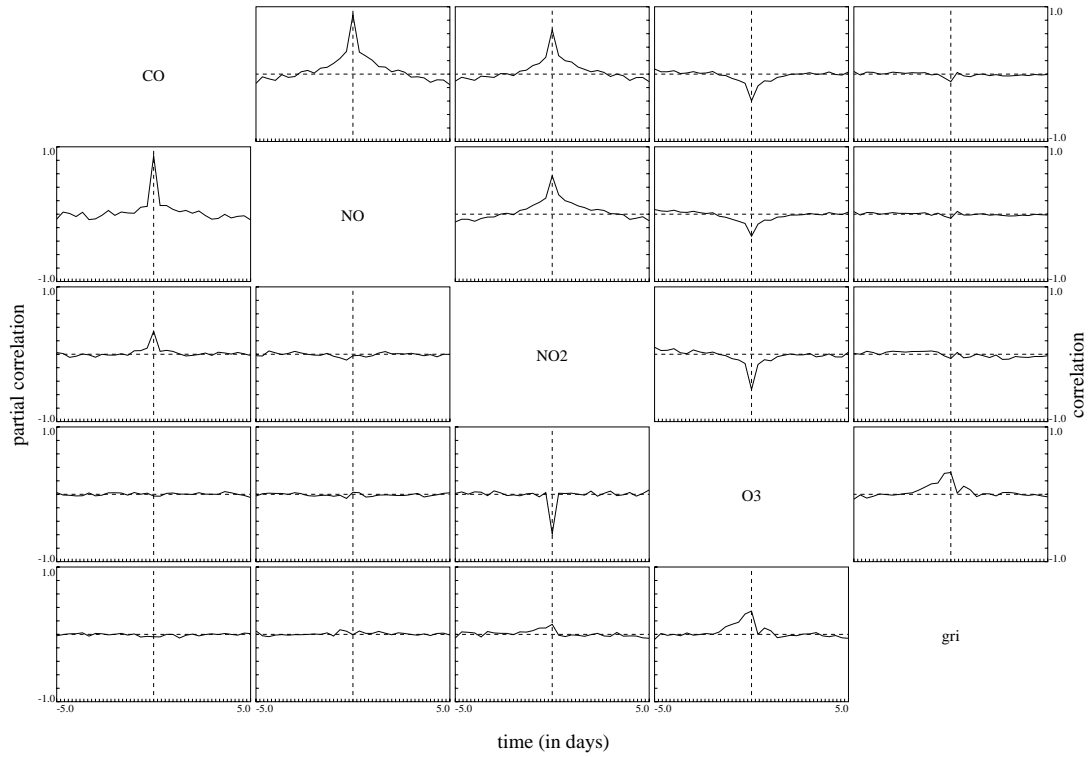
17

Figure 5.5: Correlations (above diagonal) and partial correlations (below diagonal) for trend corrected air pollution data.

partial cross spectra, which in turn have been calculated by using Theorem 2.4. The partial correlations lead to the same graph as before. Remarkable is the negative correlation between NO_2 and O_3 and the positive correlation between O_3 and gri with negative lag (a high global radiation leads with some time lag to an increase of ozone which at the same time reduces the amount of NO_2).

The dashed line in Figures 5.3 and 5.4 is a 95%-test bound of an approximate distribution of

$$\sup_{\lambda} |\hat{R}_{X_a X_b | Y_{ab}}(\lambda)|^2 \tag{5.1}$$

18

derived under the hypothesis that $R_{X_a X_b | Y_{ab}}(\cdot) \equiv 0$. Let $H_k = \int_0^1 h(x)^k dx$ and $c_T = \frac{\pi}{bT} \frac{H_4}{H_2^2} \int W(\alpha)^2 d\alpha$. It can be concluded that the real and the imaginary part of

$$c_T^{-1/2} \left(\hat{R}_{X_a X_b | Y_{ab}}(\lambda) - R_{X_a X_b | Y_{ab}}(\lambda) \right)$$

are asymptotically independent and standard normally distributed leading under the hypothesis $R_{X_a X_b | Y_{ab}}(\cdot) \equiv 0$ to a χ_2^2 distribution for $c_T^{-1} |\hat{R}_{X_a X_b | Y_{ab}}(\lambda)|^2$ (cf. Dahlhaus et al., 1997, Section 2 and Brillinger, 1981, Section 8.8). We then have taken the supremum of $m(n)$ independent χ_2^2 distributions where $m(n)$ is the maximum number of frequencies λ such that the smoothing intervals of the spectral estimates do not overlap, i.e. the dashed line is $c_T \chi^2_{2,(1-\alpha)^{1/m(n)}}$ (for the air pollution data $m(n)$ was 31).

It is very difficult to determine the exact asymptotic distribution of (5.1) - in particular since the values of $\hat{R}_{X_a X_b | Y_{ab}}(\lambda)$ are dependent for neighbouring λ.

Furthermore, the problem of multiple testing has to be addressed. An alternative would be to determine the "best approximating graph" by using a model selection criterion which penalizes the complexity of the graph in an adequate way.

We also mention that the partial correlation graph only reflects linear dependencies. Nevertheless, the above method may be useful for exploration of the dependence structure even if nonlinear dependencies are present (an example for neuron nets is given in Dahlhaus et al., 1997).

Appendix

<u>Proof of Theorem 3.2.</u> Suppose B is a regular matrix of the form

$$B = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}$$

19

Direct verification gives

$$B^{-1} = \begin{pmatrix} E^{-1} & -E^{-1}F \\ -GE^{-1} & B_{22}^{-1} + GE^{-1}F \end{pmatrix}$$

where $E = B_{11} - B_{12}B_{22}^{-1}B_{21}, F = B_{12}B_{22}^{-1}$ and $G = B_{22}^{-1}B_{21}$. To prove the first part we can assume without loss of generality $a = 1$ and $b = 2$. We set

$$B = f_{XX}(\lambda) \quad \text{with} \quad B_{11} = \begin{pmatrix} f_{11}(\lambda) & f_{12}(\lambda) \\ f_{21}(\lambda) & f_{22}(\lambda) \end{pmatrix}.$$

E then is a 2×2 matrix whose <u>rescaled</u> inverse is

$$\begin{pmatrix} 1 & -e_{12}/(e_{11}e_{22})^{1/2} \\ -e_{21}/(e_{11}e_{22})^{1/2} & 1 \end{pmatrix}$$

which proves the first part. The second part follows by setting $B = f_{XX}(\lambda)$ and $B_{11} = f_{11}(\lambda)$. For processes with vector components the assertion follows similarly.

Acknowledgements

References

Brillinger, D.R. (1996). Remarks concerning graphical models for time series and point processes. Revista de Econometria 16, 1-23.

Brillinger, D.R. (1981). Time Series: Data Analysis and Theory. Holt, Rinehart and Winston, New York.

Brillinger, D.R., Bryant, H.L. and Segundo, J.P. (1976). Identification of synaptic inter-
actions. Biol. Cybernetics 22, 213-228.

Brockwell, P.J. and Davis, R.A. (1987). Time Series: Theory and Methods. Springer-
Verlag, New York.

Cox, D.R. and Wermuth, N. (1996). Multivariate Dependencies. Chapman and Hall,
London.

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

Edwards, D. (1995). Introduction to Graphical Modelling. Springer Verlag, New York.

Gersch, W. (1972). Causality or driving in electrophysiological signal analysis. Mathe-
matical Biosciences 14, 177-196.

Lauritzen, S.L. (1996). Graphical Models. Oxford University Press, Oxford.

Seinfeld, J.H. (1986). Atmospheric Chemistry and Physics of Air Pollution. John Wiley
& Sons, New York.

Wermuth, N. (1980). Linear recursive equations, covariance selection and path analysis.
J. Amer. Stat. Ass. 75, 963-972.

Whittaker, J. (1990). Graphical Models in Applied Multivariate Statistics. John Wiley,
Chichester.