We introduce a Bayesian method for learning causal graphs for multivariate time series. The method generalizes previous learning procedures for causal graphs by allowing non-decomposability of the graphs and breakpoints in the underlying process.

Learning of causal graphs

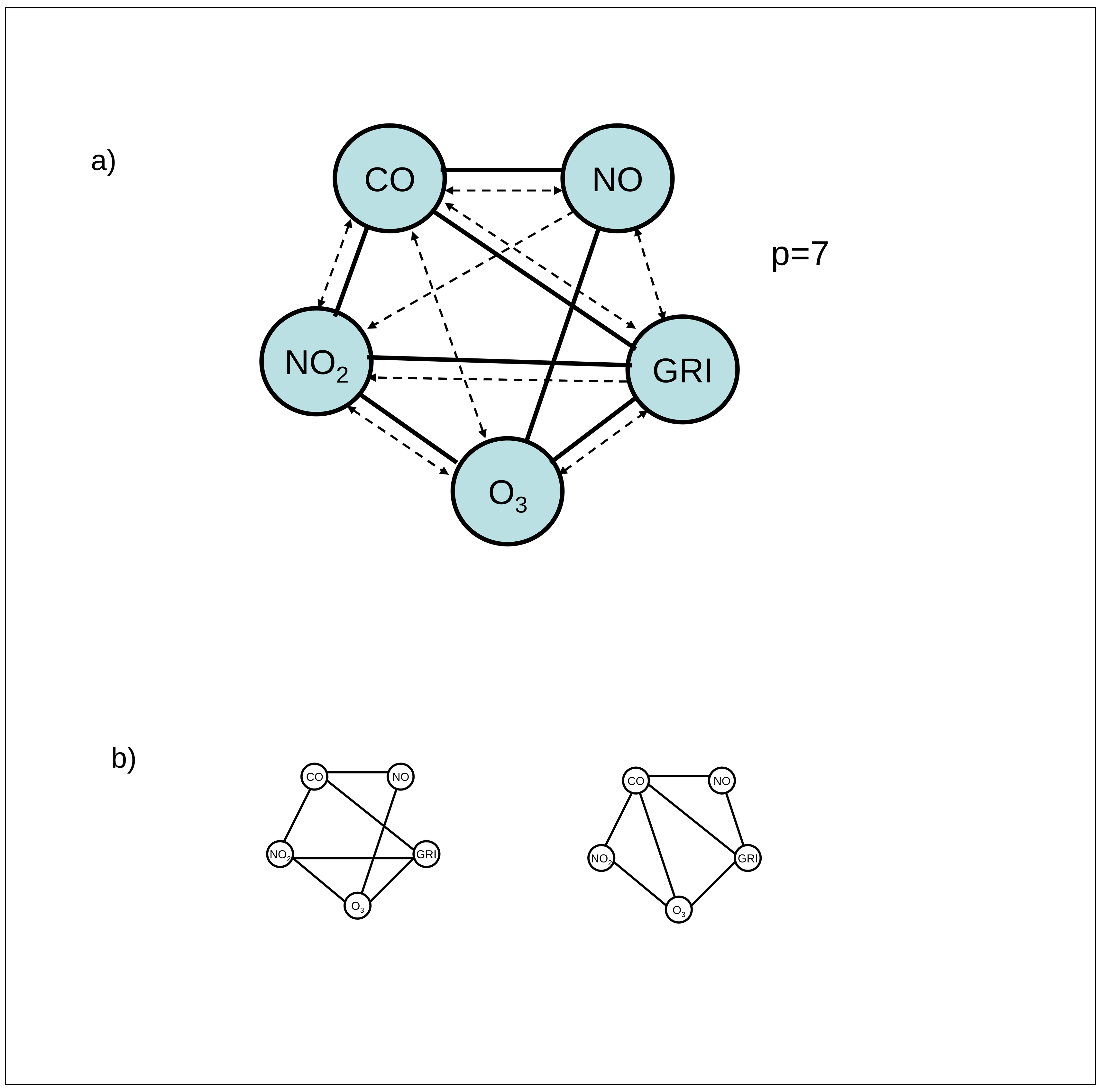

Fig. 3: **a)** Estimated causal graph for air pollution data. **b)** Comparison of graphs estimated by our method (left) and Corader & Villani, (2006) (right). The directed edges were the same in both the graphs. Thus, only the undirected parts of the graphs are shown. Notice that the left graph has a chordless four-cycle (NO, O_3, NO_2, CO) and is therefore non-decomposable.

Introduction

In the recent years, graphical modelling framework has become a standard tool for many statisticians as well as computer scientists. As reviewed by Jordan (2004), applications in fields such as bioinformatics or speech processing, involving a large number of interacting variables, can be seen as special cases of the general graphical model formalism. Graphical models provide both an intuitive interface by which humans can model interacting sets of variables, as well as a data structure that can be utilized in the development of efficient algorithms for statistical inference.

Causal graphs

In graphical models, the joint probability distribution of a set of variables is described by a graph. The nodes in the graph are identified as random variables, while the edges between the nodes describe relationships and independencies between the variables. Grager causality graphs, or simply causal graphs, are graphical models describing the dynamics of a time series. They have both directed and undirected edges, which are defined by the following two conditions:

$$\textit{(i)}\ (a,b) \notin E_1 \Leftrightarrow X_b(t) \perp \overline{X}_A(t) \mid \overline{X}_{V\setminus\{a\}}(t)$$
$$\textit{(ii)}\ (a,b) \notin E_2 \Leftrightarrow X_a(t) \perp X_b(t) \mid \overline{X}_V(t), X_{V\setminus\{a,b\}}(t).$$

The two conditions specify, respectively:

i) *a* is non-causal for *b* if knowing the history of *a* does not affect our predictions about *b*, when we know the history of all other variables.

ii) *a* and *b* are contemporaneously partially uncorrelated.

(see, Fig. 2)

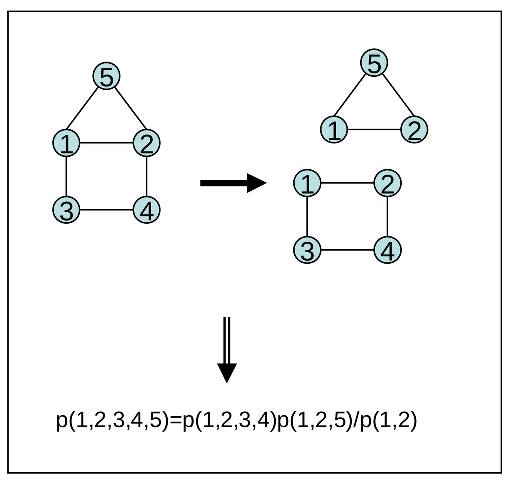

Fig. 1: The conditional joint probability factorizes according to maximal prime components of the undirected part of the graph.

$$\begin{bmatrix} x_t \\ y_t \\ z_t \end{bmatrix} = \begin{bmatrix} a_{11} & \mathbf{a}_{12} & \mathbf{a}_{13} \\ 0 & a_{22} & 0 \\ 0 & 0 & a_{33} \end{bmatrix} * \begin{bmatrix} x_{t-1} \\ y_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} b_{11} & \mathbf{b}_{12} & \mathbf{b}_{13} \\ 0 & b_{22} & 0 \\ 0 & 0 & b_{33} \end{bmatrix} * \begin{bmatrix} x_{t-2} \\ y_{t-2} \\ z_{t-2} \end{bmatrix} + \mathbf{u}_t$$

$\mathbf{u}_t \sim N(0,\Sigma)$, where $\Sigma^{-1}(1,2) = \Sigma^{-1}(2,1) = 0$

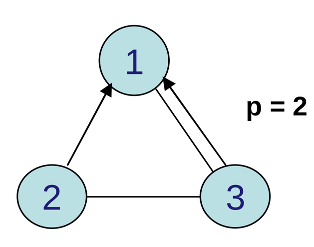

Fig. 2: A simple vector autoregressive process with the corresponding causality graph. Non-zero elements in the coefficient matrices determine the directed edges, while zeros in the inverse covariance matrix determine the missing undirected edges.

Decomposition

Let A, B, and C be subsets of nodes in an undirected graph. A triple (A,B,C) defines a decomposition of G if A,B,C are disjoint, A∪B∪C=V, C separates A from B, and C is a complete subset of V. Decompositions can be used consecutively to divide the graph into maximal prime components. This is very useful, since the joint probability of the variables, conditional on the complete history, can be factorized according to the prime components (Fig. 1).

Learning strategy

We utilize a newly introduced Bayesian information theoretic criterion, BEC (Corander & Marttinen, 2006), which measures the utility of a model in terms of its predictive entropy. An algorithm using deepest descend can be used to find the optimal model. Interpretation in probabilities can be obtained through the relative expected utilities of the models.

Results

The performance of our method has been tested with several simulated data sets. The simulations show that the method does well in inferring the underlying graph structure and lag-length. Also, when breakpoints were added to the time series, the locations of the breakpoints were inferred accurately. Figure 3 shows results from an analysis of real air pollution data that has been previously investigated by Dahlhaus (2000), and Corander & Villani (2006).

References

Corander, J. & Marttinen, P. (2006). Bayesian model learning based on predictive entropy. *Journal of Logic, Language and Information*, DOI: 10.1007/s10849-005-9004-8.

Corander, J. & Villani, M. (2006). A Bayesian approach to modelling graphical vector autoregressions. *Journal of Time Series Analysis*, 27, 141-156.

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

Jordan, M. I. (2004). Graphical models. *Statistical Science (Special Issue on Bayesian Statistics)*, 19, 140-155.

Marttinen, P. & Corander, J. (2006). Bayesian learning of causal graphs for multivariate time-series. Submitted.

Pekka.Marttinen@Helsinki.FI

Pekka Marttinen[a,b], Jukka Corander[a]

[a]Department of Mathematics and Statistics, University of Helsinki

[b]KERMIT, Faculty of Bioscience Engineering, Ghent University