Callies, U.; Scharfe, M.:

Mean spring conditions at Helgoland Roads, North Sea: Graphical modeling of the influence of hydro-climatic forcing and Elbe River discharge

In: Journal of Sea Research (2014) Elsevier

DOI: 10.1016/j.seares.2014.06.008

Mean spring conditions at Helgoland Roads, North Sea: Graphical modelling of the influence of hydro-climatic forcing and Elbe River discharge

Ulrich Callies[a,*], Mirco Scharfe[b]

[a]Institute of Coastal Research,
Helmholtz-Zentrum Geesthacht, Max-Planck-Str. 1, 21502 Geesthacht, Germany
[b]Biologische Anstalt Helgoland,
Alfred Wegener Institute, Helmholtz-Centre for Polar and Marine Research,
P.O. Box 180, 27483 Helgoland, Germany

Abstract

We analyse inter-annual changes of marine observations at Helgoland Roads (nitrate, phosphate, salinity, Secchi depth) in relation to hydro-climatic conditions and Elbe River discharge as potential drivers. Focussing on mean spring conditions we explore graphical covariance selection modelling as a means to both identify and represent the structure of parameter interactions.

While river discharge is able to modify spatial distributions and related gradients in the station's vicinity, atmospherically forced regional transport patterns govern the time dependent local conditions the station is actually exposed to. A model consistent with the data confirms the interplay of the two forcing factors for observations at station Helgoland Roads. Introducing water temperature as a third predictor of inter-annual variability does not much improve the model.

Comparing a Helgoland Roads dependence graph with corresponding graphs for other stations or related model simulations, for instance, could help identify differences in underlying mechanisms without referring to specific realizations of external forcing. With regard to prediction, supplementary numerical experiments reveal that imposing constraints on parameter interactions can reduce the chance of fitting regression models to noise.

Keywords: Helgoland Roads, inter-annual variability, partial correlation, graphical Gaussian model, Monte-Carlo-simulation, robust regression

1. Introduction

Comprehensive monitoring data convey information far beyond that on trends and variability in a collection of univariate time series. As a result of many processes impacting on the variables, observed time series become related to each other in a complex way. In many cases interpretation of correlations in terms of causality is difficult, be it, for instance, that underlying processes are unknown or that feedback loops make a causal structure difficult to detect. Nevertheless multivariate data analysis can essentially provide a fingerprint characterizing system behaviour. In the present

*Corresponding author
Email addresses: ulrich.callies@hzg.de (Ulrich Callies), mirco.scharfe@awi.de (Mirco Scharfe)

study we will explore the usefulness of graphical Gaussian modelling, also known as covariance selection modelling, for this purpose (Whittaker, 1990; Edwards, 1995). This method focusing on the analysis of conditional independence relations seems particularly promising for bridging the gap between pure statistical analyses and causal concepts. Our case study refers to the example of inter-annual variations of seasonal mean conditions at the North Sea monitoring station Helgoland Roads (Wiltshire et al., 2008).

Two conventional techniques (e.g. Johnson and Wichern, 1992) for exploring inter-relationships between several variables are principal component analysis (PCA) and canonical correlation analysis (CCA). Schlüter et al. (2008), for instance, applied these methods to data of 39 variables from different sources in the German Bight. However, the two exploratory techniques mentioned do not explicitly refer to a concept of causation. Both Pearl (2000) and Shipley (2000) proposed directed acyclic graphs (DAGs) as a formal language for that purpose. In the most simple DAG, one variable X effecting a second variable Y would be represented by a diagram made up by two nodes or vertices representing X and Y and an arrow pointing from X to Y. Such a graph corresponds with a decomposition of the joint probability distribution $P(X, Y)$ into a product of the conditional distribution $P(Y \mid X)$ and the marginal distribution $P(X)$. In the same way, any more complex DAG with several variables portrays asymmetric roles played by different variables. Arrows emanating from parent nodes and pointing to child nodes represent a user's knowledge or hypotheses about causal influences among the variables involved. To properly represent the decomposition of a joint probability distribution, a DAG must not contain any directed cycle.

In the introduction of a chapter on model selection Edwards (1995, p. 137) states: *"Any method (or statistician) that takes a complex multivariate dataset and from this claims to identify one true model is both naive and misleading"*. According to Cowell et al. (1999, p. 256) already the number of DAGs having 10 nodes amounts to about 4.2×10^{18} so that in practice any search algorithm must be selective. Prior knowledge about a causal ordering of variables will reduce the number of degrees of freedom considerably. The selection of explanatory and response variables within a multivariate regression framework is a most simple form of such ordering.

A quite general class of linear models is employed in structural equations modelling (SEM, e.g. Grace, 2006). The basic idea of SEM is that a set of linear equations should reflect a user's qualitative *a priori* knowledge about a system's internal mechanisms. Similar to path analysis proposed by Wright (1921), SEM represents the model in terms of a directed graph. SEM allows for the inclusion of latent variables and can be specialized to get models of multivariate linear regression and factor analysis, for instance. Given an adequate parsimonious model structure, a confirmatory analysis of the resulting covariance structure is to be performed (Jöreskog, 1981). Wootton (1994), for instance, found path analysis useful for a comparison of hypotheses he formulated with regard to causal mechanisms in an intertidal marine community. Representing interactions among species by a hierarchy of regression models allowed him to rank his hypotheses based on a comparison of correlation matrices predicted with the one observed.

Whittaker (1990) elaborates on the relationship between multiple regression models and the general concept of conditional independence. Conditional independence relationships can be represented as missing links in undirected Graphical Gaussian Models (GGMs). For multivariate normally distributed data, conditional independence of two variables given fixed values for all the rest of data is reflected by a corresponding zero entry in the inverse covariance (or correlation) matrix. As a means to reduce the dimensionality of a problem, Dempster (1972) proposed the method of covariance selection which simplifies GGMs by setting small elements of the inverse covariance matrix equal to zero. Jöreskog (1981) emphasized successful applications of this technique when

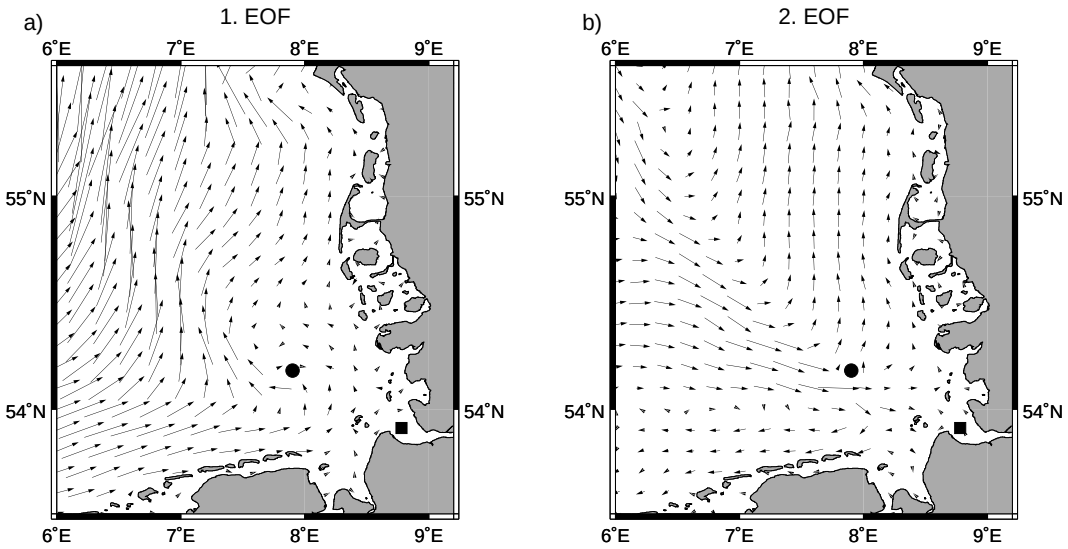

Figure 1: Vector fields of volume transport anomalies (Empirical Orthogonal Functions, EOFs) in the German Bight, obtained by subjecting model based long-term reconstructions to principal component analysis. Explained variances are 83.5% for the first and 11.6% for the second EOF. Black circles indicate the location of the monitoring station Helgoland Roads. Black squares indicate the Elbe estuary.

covariance structures are specified in advance. Magwene (2001), for instance, used GGMs and the notion of mutual information for testing hypotheses about patterns of interactions among phenotypic traits, and Eroukhmanoff and Svensson (2009) chose the same approach for a description of antipredator adaptation.

In contrast with DAGs, GGMs are unable to directly represent causal relationships, they rather estimate main aspects of the interaction structure of a joint distribution of a set of variables. For monitoring data it seems hardly possible either to relate all the variations seen in the data to specific processes or to *a priori* anticipate a causal structure to be satisfied. Different SEMs, for instance, may generate covariance matrices that are indistinguishable (Pearl, 2000) or at least imply the same set of conditional independence relations.

The reason why in this study we focus on GGMs is that recording effects of processes rather than processes themselves is what any monitoring station actually does. We suggest the use of graphical modelling for fingerprinting data sets. Baretta et al. (1998) investigate the sensitivity of marine ecosystem model output with regard to the specification of uncertain forcing parameters and boundary conditions. Even if a model was perfect, insufficient external forcing would still produce discrepancies between model simulations and corresponding observations. The characteristic pattern of covariation in model output, however, can be hoped to remain preserved to some extent. Similar considerations hold for observations at different stations. Our prototypical study tries to establish a specific fingerprint of covariation that characterizes the Helgoland Roads data on the time scale of inter-annual variability.

2. Data

In 1962, the Biologische Anstalt Helgoland started a long-term pelagic monitoring program at Helgoland Roads ($54°11.3'\mathrm{N}, 7°54'E$) in the inner German Bight (North Sea). This program comprises monitoring of water temperature, salinity, Secchi depth, inorganic nutrients and phytoplankton on a work-daily basis (Wiltshire et al., 2008). The Helgoland area (cf. Fig. 1) represents a transitional zone between coastal and open North Sea waters. Coastal waters are strongly influenced by the input of the River Elbe, which represents the most relevant fresh water source in the German Bight. A varying coastal impact on the Helgoland area is assumed to be related to meteorological and hydrodynamic conditions (Stockmann et al., 2010; Scharfe, 2013).

The data underlying our study cover the period 1962-2004. Five variables observed at Helgoland Roads are combined with two physical drivers known to affect variability of local conditions at this station:

Helgoland Roads data: We selected five variables, observed on a work-daily basis: phosphate (PO_4), nitrate (NO_3), salinity (Sal), Secchi depth (Sec), and temperature (WT).

Elbe River discharge: Daily values of Elbe River discharge ($Elbe$) were taken from the river gauge Neu-Darchau (Elbe km 536, Federal Institute of Hydrology, www.bafg.de).

Hydrodynamic transports in the German Bight: In a first step, two-dimensional hydrodynamic simulations stored on an hourly basis in the database coastDat (www.coastdat.de; Weisse and Plüß, 2006) were averaged to obtain vector fields of monthly mean volume transports. In a second step, dominant modes of spatially coherent variability in these fields were identified by PCA. The principal components (PCs) obtained have the meaning of time coefficients (amplitudes) of the two anomaly patterns (i.e. deviations from the long-term mean conditions) shown in Fig. 1. Technically, the vector fields shown are weighting factors (loadings) used for mapping each monthly mean volume transport field to one data point of the corresponding PC time series. In the Earth Sciences this special form of PCA is referred to as Empirical Orthogonal Function (EOF) analysis (e.g. von Storch and Zwiers, 1999). In our study we used only the second PC, hereafter called $Curr$, as this turns out to be much more influential on observations at station Helgoland than the first PC, despite its only minor contribution to the overall variability of volume transports in the German Bight (Stockmann et al., 2010; Scharfe, 2013). In months when $Curr$ is positive, volume transport deviations from the mean conditions are in agreement with Fig. 1b, whereas when $Curr$ is negative the pattern of deviations is similar in structure but with the orientations of all vectors reversed.

As a basis for studying inter-annual variability of mean seasonal conditions, data for spring (Mar/Apr/May), summer (Jun/Jul/Aug), autumn (Sep/Oct/Nov) and winter (Dec/Jan/Feb) were averaged, which resulted in annual data points for each of the four seasons and for each variable. Spring data are shown in Tab. 1. Further data processing included two steps. First, seasonal means of all variables except $Curr$ were square-root transformed to better approximate normal distributions. Second, the seasonal mean time series were de-trended by taking differences between consecutive years.

Our analysis of the resulting 42 (for Secchi depth: 36) annual data points was based solely on correlations. As Secchi depth was not monitored until 1968, all correlations including Secchi depth were estimated based on time series shortened accordingly. We tested the seasonal data sets for multivariate normality using the Energy-Statistic test (Székely and Rizzo, 2005) implemented

Table 1: Annual spring mean (Mar/Apr/May) data underlying this study.

Year	PO_4 [μmol/l]	NO_3 [μmol/l]	Sal [-]	Sec [m]	$Elbe$ [m^3/s]	$Curr$ [-]	WT [$^\circ$C]
1962	0.67	22.1	30.7	-	1102	0.15	5.1
1963	0.40	6.1	32.9	-	659	1.01	2.7
1964	0.66	10.9	32.9	-	543	1.01	4.4
1965	0.68	23.8	31.1	-	1486	0.23	5.5
1966	0.75	20.1	31.2	-	1254	−0.17	5.1
1967	0.94	24.0	30.0	-	1314	−0.85	6.9
1968	0.74	21.7	31.3	2.3	1115	0.03	5.5
1969	0.73	8.8	31.8	3.1	1317	1.08	4.0
1970	0.82	26.6	30.5	2.3	1660	−0.57	3.9
1971	0.63	14.7	31.5	3.4	737	0.35	5.5
1972	0.57	10.2	33.0	3.8	499	0.35	5.4
1973	1.00	18.3	31.2	2.6	693	−0.64	6.5
1974	0.64	11.8	32.9	4.7	557	0.86	6.6
1975	0.89	27.0	32.1	2.4	921	0.08	6.2
1976	0.76	18.0	31.8	2.6	650	0.01	5.0
1977	1.03	16.9	32.2	2.7	829	−0.42	6.2
1978	0.75	20.5	32.7	2.9	929	0.49	5.7
1979	0.81	15.0	32.2	4.2	1584	0.54	3.2
1980	0.61	24.9	31.2	3.4	1277	0.32	5.3
1981	0.84	31.5	31.2	3.7	1366	0.44	5.5
1982	0.72	27.7	30.3	3.6	1066	−0.52	5.5
1983	0.70	26.0	31.5	2.7	1050	−0.04	6.2
1984	0.49	23.5	32.7	4.6	643	0.63	5.2
1985	0.69	26.1	31.9	4.6	805	0.06	4.2
1986	0.70	27.9	32.6	3.4	949	0.57	3.3
1987	0.52	43.5	31.7	3.8	1667	0.24	3.8
1988	0.56	44.5	30.0	3.2	1574	0.14	6.3
1989	0.57	33.1	32.3	3.6	659	−0.03	7.3
1990	0.51	42.0	31.7	3.4	655	−0.87	8.5
1991	0.37	22.8	33.2	4.1	450	−0.10	5.6
1992	0.64	30.7	31.7	3.9	877	−0.35	6.7
1993	0.49	43.1	32.5	4.4	578	0.36	6.6
1994	0.66	73.0	31.7	3.3	1395	−0.30	6.4
1995	0.70	72.2	30.0	2.6	1171	−1.04	7.2
1996	0.24	23.0	31.9	6.1	803	0.45	3.4
1997	0.37	48.8	31.7	4.9	1014	−0.59	5.8
1998	0.84	16.8	32.7	5.2	662	0.25	7.3
1999	0.48	24.7	31.9	4.2	1172	0.51	7.4
2000	0.74	36.7	30.7	3.0	1309	−0.10	7.3
2001	0.29	20.8	31.7	4.4	825	0.50	6.0
2002	0.48	25.7	31.6	3.8	1240	0.46	7.4
2003	0.55	21.5	32.0	4.5	783	0.45	6.3
2004	0.64	21.4	32.4	3.2	685	0.42	6.6

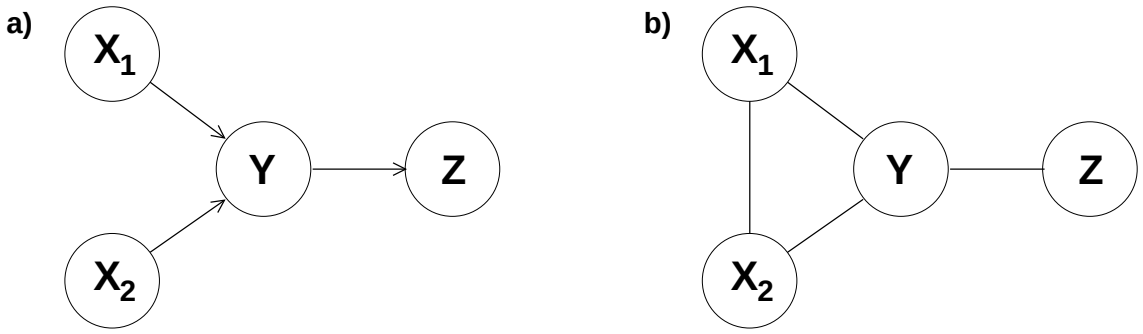

Figure 2: a) Directed graph corresponding with Eqs. (1)-(4). b) Conditional independence graph ('moral graph') derived from the directed graph on the left.

as function 'mvnorm.etest' in the software package 'energy' in R (version 3.0.3). Referring to just those years for which Secchi depth was available, all inter-annual differences of seasonal data passed the test at the 5% significance level.

3. Methods

3.1. Partial correlations

Assume four variables X_1, X_2, Y, and Z to evolve according to the following set of linear equations:

$$X_1 = \varepsilon_1 \tag{1}$$
$$X_2 = \varepsilon_2 \tag{2}$$
$$Y = \alpha X_1 + \beta X_2 + \delta_y \varepsilon_y \tag{3}$$
$$Z = \gamma Y + \delta_z \varepsilon_z \tag{4}$$

These equations are consistent with the causal scheme in Fig. 2a. Coefficients α, β, and γ quantify variable interactions, the ε_i denote independent normally distributed noise with zero mean and variance one, corresponding amplitude factors δ_i allow for an adjustment of the relative importance of these noise terms.

Suppose that Eqs. (1)-(4) are unknown but covariances between the four variables can be obtained from observations. Disregarding for the moment the estimation problem, the covariance matrix $\mathbf{S}$ combining all variances and covariances of the four variables would read:

$$\mathbf{S} = \begin{array}{c|cccc} & X_1 & X_2 & Y & Z \\ \hline X_1 & 1 & 0 & \alpha & \gamma\alpha \\ X_2 & - & 1 & \beta & \gamma\beta \\ Y & - & - & \alpha^2+\beta^2+\delta_y^2 & \gamma\left[\alpha^2+\beta^2+\delta_y^2\right] \\ Z & - & - & - & \gamma^2\left[\alpha^2+\beta^2+\delta_y^2\right]+\delta_z^2 \end{array} \tag{5}$$

To which extent can the structure of Eqs. (1)-(4) be retrieved from matrix $\mathbf{S}$? Mutual independence of input variables X_1 and X_2 is represented by a zero marginal correlation, all the other

elements of $\mathbf{S}$ are non-zero. From Eqs. (1) - (4) we know, however, that correlations between Z and forcing variables X_1 and X_2 do not represent direct causal effects. Zero entries in the inverse covariance matrix $\mathbf{S^{-1}}$ (sometimes called concentration or precision matrix) properly represent this situation:

$$\mathbf{S^{-1}} = \begin{array}{c|cccc} & X_1 & X_2 & Y & Z \\ \hline X_1 & 1+\alpha^2\delta_y^{-2} & \alpha\beta\delta_y^{-2} & -\alpha\delta_y^{-2} & 0 \\ X_2 & - & 1+\beta^2\delta_y^{-2} & -\beta\delta_y^{-2} & 0 \\ Y & - & - & \delta_y^{-2}+\gamma^2\delta_z^{-2} & -\gamma\delta_z^{-2} \\ Z & - & - & - & \delta_z^{-2} \end{array} \tag{6}$$

According to Whittaker (1990, Corollary 5.8.2) *"each off diagonal element of the inverse variance, scaled to have a unit diagonal, is the negative of the partial correlation between the two corresponding variables, partialled on all of the remaining variables, the rest"*. In our example, conditioning on Y completely blocks correlation between Z and X_1 or X_2. Undirected edges in Fig. 2b correspond with the four non-zero off diagonal elements in $\mathbf{S^{-1}}$. Fig. 2b does not distinguish, however, between explanatory and response variables. The marginally uncorrelated parent nodes X_1 and X_2 become 'married' (therefore the term 'moral graph' introduced by Lauritzen and Spiegelhalter, 1988) by conditioning on their common child node Y, which expresses the need for compensating effects of X_1 and X_2 if the value of Y is to be held fixed. This latter aspect is beyond the scope of pure regression modelling.

According to Corollary 5.8.1 from Whittaker's book *"each diagonal element of the inverse variance is the reciprocal of a partial variance"*, i.e. of a variance holding all remaining variables fixed. According to Eq. (6) the partial variance of the final output variable Z, for instance, equals the variance of the noise term in Eq. (4). For X_1 an information flow against the assumed direction of causation manifests itself in the first diagonal element of $\mathbf{S^{-1}}$: The stronger the influence of X_1 on Y is, i.e. the larger coefficient α and the smaller noise term δ_y are (cf. Eq. (3)), the more the assumed unit marginal variance of input variable X_1 narrows after knowing the value of Y. A similar formula holds for X_2. For Y the corresponding diagonal element of $\mathbf{S^{-1}}$ combines information from all three of the remaining variables. First, knowing only X_1 and X_2, the variance of Y becomes δ_y^2. This partial variance, however, further narrows after learning also the value of Z. Z is informative about Y in the same way as Y is about X_1 and X_2, respectively. Generally, the presence of an edge in Fig. 2b indicates that a multiple regression would benefit from including the adjacent variable as another predictor.

3.2. Fitting graphical Gaussian models

Fig. 2b provides an example Graphical Gaussian Model (GGM) that represents a set of jointly Gaussian variables in terms of corresponding vertices and a set of undirected edges connecting them. Edges missing in a GGM constrain interactions between the variables. The absence of an edge connecting X_1 and Z, for instance, reflects conditional independence (zero partial correlation) of these two variables given values for all the rest of variables:

$$\mathrm{corr}(X_1, Z \mid X_2, Y) = \mathrm{corr}(X_1, Z \mid Y) = 0 \tag{7}$$

X_2 could be dropped from the list of variables the partial correlation is adjusted for by making use of the separation theorem (Whittaker, 1990). According to this theorem it is sufficient to condition the correlation on just those variables that separate X_1 from Z in the graph. Dempster

(1972) developed covariance selection modelling as a general framework to assess whether or not the constraints represented by a GGM are consistent with observations. A comprehensive presentation of the concept can be found in Whittaker (1990).

In the following we assume that all variables have been standardized so that covariance matrices become correlation matrices with unit diagonals. Among all correlation matrices that satisfy the constraints of a given graph G (i.e. all matrices with zero entries for elements of their inverse that correspond with edges missing in the graph) some matrix $\mathbf{V}$ will fit the data best. The difference between the log-likelihoods of sample correlation matrix $\mathbf{S}$ and $\mathbf{V}$ provides an entropy type measure of the amount of information in the data (i.e. in matrix $\mathbf{S}$) against the interaction structure hypothesized by the graph G. The deviance $\mathrm{dev}_S(G)$ is defined as twice this difference of log-likelihoods or twice the sample size N times the Kullback-Leibler information divergence between two jointly normal distributions, assuming that their means are equal (Kullback, 1959):

$$\mathrm{dev}_S(G) = N\left[\mathrm{tr}(\mathbf{S}\mathbf{V}^{-1}) - k - \ln\det(\mathbf{S}\mathbf{V}^{-1})\right] \tag{8}$$

This formula involves the trace and determinant of $\mathbf{S}\mathbf{V}^{-1}$, k denotes the number of variables taken into account. Generally the optimum matrix $\mathbf{V}$ among those satisfying the constraints represented in G cannot be specified analytically unless the graph belongs to the special class of decomposable graphs (Whittaker, 1990; Edwards, 1995). In our study we used the iterative proportional fitting algorithm described in Whittaker (1990). Specific properties of $\mathbf{V}$ imply $\mathrm{tr}(\mathbf{S}\mathbf{V}^{-1}) = k$ so that Eq. (8) takes the following simplified form (Edwards, 1995):

$$\mathrm{dev}_S(G) = N\ln\left(\frac{|\mathbf{V}|}{|\mathbf{S}|}\right) \tag{9}$$

If data are normally distributed, the deviance has an asymptotic χ^2 distribution with the degrees of freedom given by the number of edges missing in the graph (Whittaker, 1990).

If G_0 denotes the most simple graph without any edges (representing completely independent variables), its deviance can be used as a measure of overall connectivity in the observed data. The corresponding fitted correlation matrix is the unit matrix so that we obtain

$$\mathrm{dev}_S(G_0) = -N\ln|\mathbf{S}| \tag{10}$$

From Eqs. (9) and (10) we get

$$\mathrm{dev}_S(G) = \mathrm{dev}_S(G_0) - \mathrm{dev}_V(G_0) \tag{11}$$

stating that the deviance of graph G equals a reduction of information against the hypothesis of complete independence of the data, due to the assumed set of pairwise conditional independences.

To decide whether or not a certain edge in a graph should be excluded/included (re-established), one considers the amount by which the graph's deviance would increase/decrease by this measure (deviance differences d_{excl} and d_{incl}, respectively). When a graph is complete (i.e. each pair of vertices is connected), d_{excl} for an edge connecting variables labelled i and j equals the edge exclusion deviance (Whittaker, 1990) specified as a function of the partial correlation $\rho_{ij|rest}$ between the two variables of interest:

$$d_{excl} = -N\ln\left(1-\rho^2_{ij|rest}\right) \tag{12}$$

If the graph is already truncated, however, this simple formula is no longer applicable.

Whittaker (1990) devotes one section of his textbook to a discussion of different graphical model search strategies. Different initial graphs could be chosen in combination with different iterative search procedures, including the complete graph for backward elimination of edges or the graph without any edges for a forward inclusion procedure. Given these plus many other options for the starting point, the good news is that according to Whittaker (1990, p. 252) *"there is little evidence to suggest that its choice affects the final model selected"*. In our analysis we chose backward elimination.

An essential aspect is that fitting a graphical model shifts the focus away from asking for the significance of correlations between pairs of variables. Instead one asks for the significance of partial correlations. The corresponding formula combines several marginal correlations into one variable via the mathematical procedure of matrix inversion (cf. the example in Section 3.1).

3.3. Graphical models and multivariate regression: Monte-Carlo-simulations

For studying effects of graphical modelling on the robustness of linear regression we will refer to the example already presented in Section 3.1. We assume that Y and Z are environmental parameters of interest to be predicted as functions of X_1 and X_2,

$$\widehat{Y} = a_y X_1 + b_y X_2 \tag{13}$$

$$\widehat{Z} = a_z X_1 + b_z X_2 \tag{14}$$

with estimates $\widehat{Y}$, $\widehat{Z}$ and regression coefficients a_i, b_i.

While Eq. (13) is in line with the assumed causal relation Eq. (3), Eq. (14) for Z disregards the role of Y as an intervening variable (cf. Fig. 2a). Whittaker (1990, p. 325) states that while the graphical modeller *"... examines the joint distribution of the response variables, the regression modeller focuses on the marginal distribution of each response"*. Therefore, although regression coefficients can be calibrated for each of the two Eqs. (13) and (14) independently, replacing sample correlation matrix $\mathbf{S}$ by a modified matrix $\mathbf{V}$ that satisfies the constraints of the GGM in Fig. 2b will also affect the estimates of regression coefficients in Eq. (14).

To evaluate the effect of graphical modelling on multiple regression, we performed 1000 Monte-Carlo-simulations in each of which samples of size N (for N=10, 20, 50, and 100) comprising the four variables X_1, X_2, Y, and Z were generated based on Eqs. (1)-(4). To obtain (for large samples) normalized variables with variance one, we chose parameter values $\alpha = \beta = \delta_y = 1/\sqrt{3}$ and $\gamma = \delta_z = 1/\sqrt{2}$. For each of the 1000 simulations then the coefficients in Eqs. (13)-(14) were estimated based on a) the original sample correlation matrix $\mathbf{S}$ and b) the correlation matrix $\mathbf{V}$ modified to satisfy the graphical model displayed in Fig. 2b. The performance of each of the two multiple regression models was then evaluated based on a) the sample data used for model calibration and b) an independent sample (sample size 1000) generated just for validation purposes. The extent to which X_1 and X_2 determine Z, for instance, was specified in terms of the following coefficient of determination or explained variance:

$$R_z^2 = \frac{var(Z) - var(Z - \widehat{Z})}{var(Z)} \tag{15}$$

An analogous coefficient R_y^2 was used for an assessment of $\widehat{Y}$. Note that the decomposition $var(Z) = var(\widehat{Z}) + var(Z - \widehat{Z})$ holds for the calibration data but not for the independent validation data set.

Table 2: Explained variances and loadings for the two leading PCs in each season.

Season	PC	Expl. Variance	PO_4	NO_3	*Sal*	*Sec*	*Elbe*	*Curr*
Winter	1	44%	0.31	0.38	−0.45	−0.47	0.26	−0.52
	2	23%	−0.43	0.53	−0.06	0.24	0.63	0.28
Spring	1	65%	0.34	0.44	−0.43	−0.43	0.36	−0.44
	2	13%	−0.68	0.25	−0.21	0.27	0.58	0.14
Summer	1	33%	0.06	0.50	−0.63	−0.08	0.59	0.00
	2	23%	−0.67	0.36	0.13	0.46	−0.04	0.43
Autumn	1	40%	0.49	0.17	−0.34	−0.55	0.10	−0.54
	2	26%	−0.13	0.57	−0.43	0.15	0.61	0.28

For each simulated sample correlation matrix $\mathbf{S}$ we also recorded the deviance $\mathrm{dev}_S(G)$ (cf. Eq. (9)) of graph G shown in Fig. 2b. This enabled us to calculate the percentage of cases in which the deviance exceeded the 5% point of the χ^2 distribution (here: 6.0 as two edges are missing in Fig. 2b) so that the model (we know to be true) would have been rejected erroneously.

4. Results

4.1. Principal component analysis

As a first step of exploratory data analysis, we subjected standardized seasonal data to PCA, not taking into account water temperature WT. Effects of WT on the structure of the PCs were found to be small (not shown). The special Section 4.4 will be devoted to the role of WT in the context of our analysis.

Tab. 2 lists for each season the percentages of total variance explained by either the first or the second PC. In addition, it shows the corresponding loadings. With the exception of summer the leading patterns of covariation in different seasons are similar (note that all loadings shown are normalized in the sense that in each line the sum of squared values equals one). In summer different roles of *Elbe* and *Curr* manifest themselves much clearer than for other seasons by their contributions to different PCs.

A striking feature to be recognized from Tab. 2 is that in spring the leading mode of covariation is much more dominant than in other seasons, explaining as much as about 65% of total variability. Tab. 3 addresses connectivity from a different perspective. For each individual variable it gives the amount of variance that could be explained by multiple regression on all other five variables listed in the table. Although such prediction makes little sense for variables *Elbe* or *Curr*, the formal procedure quantifies also for them the overall degree to which they are connected with the rest of variables. For *Curr* we see a drastic difference between spring conditions, when this variable seems more central than any of the other variables, and summer conditions, when *Curr* remains more or less unconnected (like also Secchi depth). The variable most correlated with *Curr* in summer is water temperature WT (in Tab. 3 not taken into account) with correlation 0.37 (not shown). For river discharge *Elbe* connectivity assumes its maximum value again in spring; the same holds for PO_4 and *Sal*.

Table 3: Percentages of variance of single variables explainable by multiple regression on all other five variables. Last line: Deviance of a graphical model assuming complete independence (cf. Eq. (10)).

Response Variable	Winter	Spring	Summer	Autumn
PO_4	22%	44%	22%	48%
NO_3	43%	71%	36%	32%
Sal	42%	68%	53%	32%
Sec	58%	66%	5%	70%
$Elbe$	34%	52%	42%	37%
$Curr$	67%	71%	3%	71%
$-N \ln \lvert\mathbf{S}\rvert$	85	160	45	93

In spring also the amount of overall information against the hypothesis of complete independence of all variables (last line in Tab. 3, calculated from Eq. (10)) is by far the largest. The analysis in Section 4.2 trying to disentangle interactions between the set of variables in terms of conditional independence constraints will therefore focus on this particular season.

4.2. Fitting a GGM

A conditional independence model found to be consistent with the six-dimensional data set (excluding water temperature WT) is presented in Fig. 3a. Although a distinction between forcing and response variables is not intrinsic to a GGM, the display already anticipates asymmetric roles of parameters in that explained variances are not indicated for the physical forcing parameters $Elbe$ and $Curr$.

The overall fit of the graph retaining just 7 out of 15 edges seems satisfactory. Its total deviance of 10.4 is below the 5% point (15.5) of the χ^2 distribution for 8 degrees of freedom (the number of links discarded). Re-establishing the strongest link excluded ($Curr - Elbe$) would diminish the graph's deviance by only 3.9, a value close to the 5% point for one degree of freedom (3.8). On the other hand, the deviance difference d_{excl} when discarding the weakest link retained (4.8 for $Curr - PO_4$) exceeds the 5% point of the χ^2 distribution.

The first array in Tab. 4 contrasts the fitted marginal correlation matrix with its unconstrained counterpart. It can be seen that in order to satisfy the constraints of the graph, only those marginal correlations that correspond with edges discarded changed their values (cf. Whittaker, 1990). In contrast, partial correlations (second array in Tab. 4) were all affected by the process of fitting, becoming zero where edges in the graph are missing.

As measures of edge importance, Tab. 4 also lists deviance differences that would result from the exclusion of any edge kept in a graph. The assessment of the strengths of individual links is context dependent. For the unconstrained sample correlation matrix (i.e. a complete graph) the deviance differences d_{excl} can be calculated as functions of partial correlations using Eq. (12). For an already truncated graph this simple relationship is no longer valid. As an example consider the edge connecting $Curr$ and Sal in Fig. 3a. Although the corresponding partial correlation (0.52) remained more or less unaffected by model truncation, the corresponding change in the graph's deviance in the case of the omission of the edge increased from $d_{excl} = 13.3$ to $d_{excl} = 22.7$ (bottom panel of Tab. 4). On the other hand, the deviance differences d_{excl} of the two edges connecting $Elbe$

Table 4: Marginal correlations, partial correlations, and deviance differences as a) sampled (lower triangles) and b) fitted to the GGM in Fig. 3a (upper triangles).

	PO_4	NO_3	Sal	Sec	$Elbe$	$Curr$
Marginal correlations						
PO_4		0.53	-0.43	**-0.60**	0.38	**-0.55**
NO_3	0.39		-0.60	**-0.73**	**0.63**	**-0.71**
Sal	-0.46	-0.67		0.47	**-0.63**	**0.76**
Sec	**-0.60**	**-0.73**	0.60		-0.47	0.57
$Elbe$	0.35	**0.63**	**-0.63**	-0.45		-0.57
$Curr$	**-0.55**	**-0.71**	**0.76**	0.69	-0.44	
Partial correlations						
PO_4		0	0	**-0.36**	0	**-0.22**
NO_3	-0.26		0	**-0.49**	**0.31**	**-0.33**
Sal	0.01	-0.03		0	**-0.34**	**0.54**
Sec	**-0.42**	**-0.47**	0.05		0	0
$Elbe$	0.17	**0.42**	**-0.42**	0.10		0
$Curr$	**-0.26**	**-0.35**	**0.52**	0.12	0.27	
Deviance differences d_{excl} and (d_{incl})						
PO_4		(3.3)	(0.1)	**8.7**	(0.1)	**4.8**
NO_3	2.9		(1.3)	**21.3**	**8.3**	**11.3**
Sal	0.0	0.3		(2.3)	**8.2**	**22.7**
Sec	**8.2**	**10.7**	0.1		(0.1)	(3.1)
$Elbe$	1.2	**8.3**	**8.3**	0.4		(3.9)
$Curr$	**2.9**	**5.5**	**13.3**	0.7	3.1	

Numbers in bold type refer to edges retained in Fig. 3a. Deviance differences put in parentheses indicate the amount by which the graph's deviance would decrease if the edge dismissed was re-established.

with Sal and NO_3, respectively, stayed about the same although both of the two corresponding partial correlations decreased under simplification of the graph.

Fig. 3b presents a graphical model that disregards potential effects of marine currents. The omission of variable $Curr$ generates a new edge connecting Sal and NO_3, removal of which would increase the graph's deviance by 9.9 (not shown). The negligible relevance of the same edge in Fig. 3a ($d_{incl} = 1.3$, cf. Tab. 4) is due to the fact that Fig. 3a already explains the correlation between salinity and nitrate (-0.67) by both variables being influenced by the external drivers $Curr$ and $Elbe$. We will return to this aspect in Section 4.3.

4.3. Assuming asymmetric inter-relationships

The interaction structure between six variables portrayed by the GGM in Fig. 3a does not yet constitute a hierarchy of variables with regard to causal process. Whittaker (1990, p. 72) states: *"At no point do we suggest it is possible for the probability modeller or data analyst to determine the direction of an edge, or the implicit order underlying the graph, other than by presupposition."*

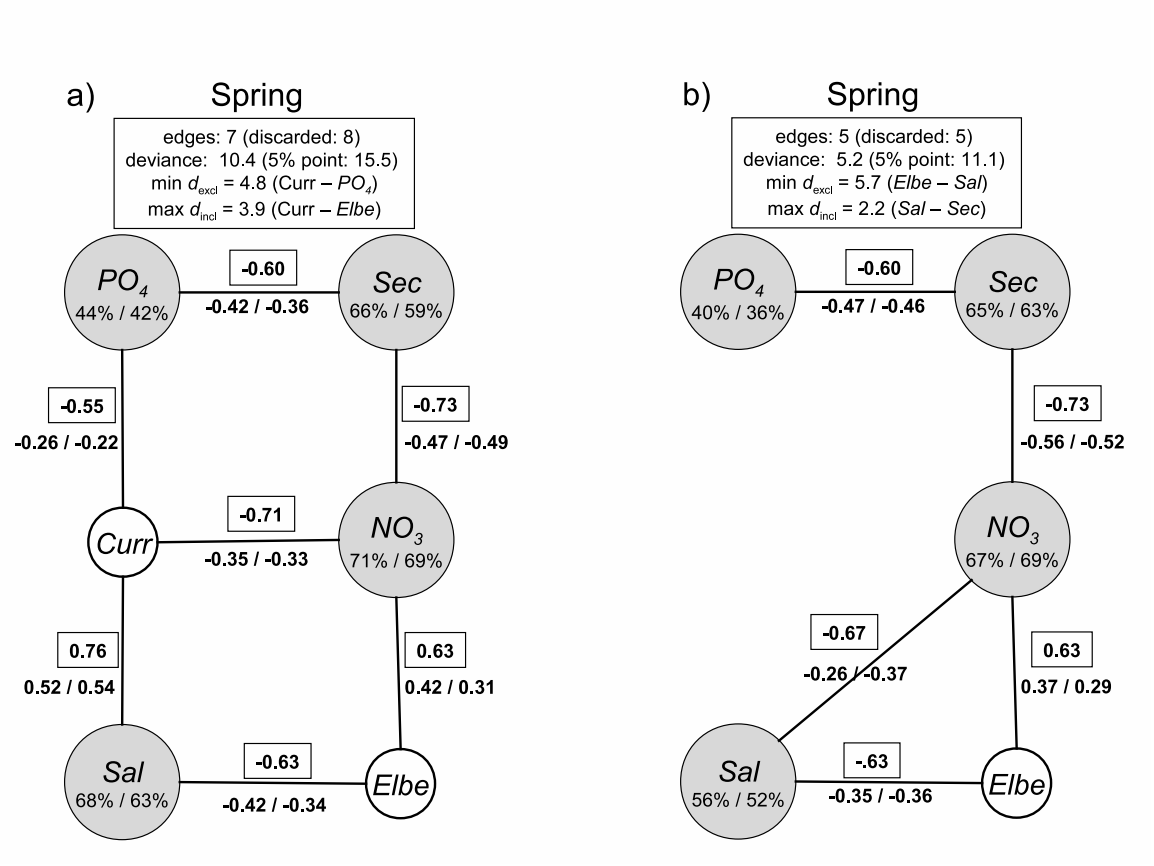

Figure 3: GGMs fitted for spring conditions. Boxes at the top of each panel summarize the quality of fit. Numbers in small boxes attached to edges: marginal correlations. Numbers below these boxes: partial correlations unconstrained/constrained by the graph. Percentages attached to each variable: variances explained by multiple regression on all remaining variables (unconstrained/constrained). More details about graphical model (a) are specified in Tab. 4.

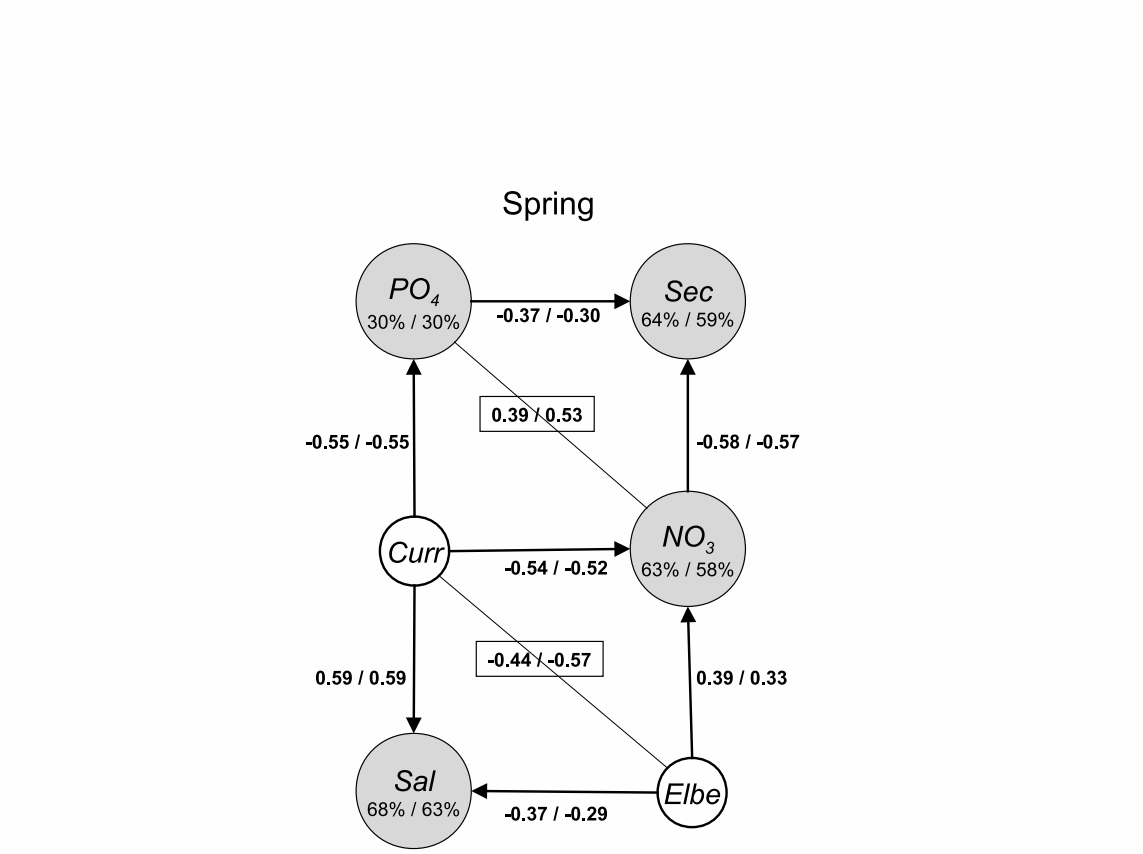

Figure 4: Mixed graph assigning orientations to all undirected edges of the GGM in Fig. 3a. Undirected edges represent marginal correlations given in boxes. Numbers annotated to directed edges: regression coefficients. Attached to each variable: percentages of variance explained by multiple regression on all parent variables. All parameters are specified unconstrained/constrained by the GGM in Fig. 3a.

Definition of directed edges based on external knowledge may, however, supplement the analysis of pairwise conditional independences summarized in a GGM.

For the example in Fig. 3a we suggest grouping the six variables in three causal layers. Parent nodes *Curr* and *Elbe* are assumed to have direct effects on the three child variables PO_4, NO_3, and *Sal*. On grandchild level, *Sec* is assumed to depend on PO_4 and NO_3, a relationship that can be thought of as being mediated by phytoplankton growth, for instance. In the resulting hypothetical quasi-causal model (Fig. 4) all undirected edges of the GGM (Fig. 3a) were converted into directed ones. When regression coefficients in the directed graph are specified based on the correlation matrix fitted to the GGM (i.e. not based on original sample correlations), explained variances of response variables *Sec* and *Sal* are identical with those in Fig. 3a. For PO_4 and NO_3, however, results differ as for these two variables the set of parent variables in Fig. 4 does not include *Sec* which is adjacent in Fig. 3a. Explained variances of nutrients annotated in Fig. 3a and Fig. 4,

respectively, would be the same if we had chosen a one step model using Sec as a third predictor on the same level as $Curr$ and $Elbe$ (reversal of the two arrows pointing into Sec).

The two undirected edges in Fig. 4 represent marginal correlations between the predictors in each of the two consecutive regression steps. The 'marrying' of parent nodes in Fig. 4 is reminiscent of moralising a directed graph (cf. Fig. 2 and the explanation in Section 3.1). It is impossible, however, to construct a directed graph in such a way that the corresponding moral graph reproduces Fig. 3a. It can easily be verified that for a chordless four cycle like the one made up in Fig. 3a by variables $Curr, NO_3, Sal$, and $Elbe$, for instance, replacement of all undirected edges by directed edges would inevitably produce either a cyclic graph (which is forbidden as it would not be related to any joint probability distribution; Whittaker, 1990) or nodes with two parent nodes to be married by a diagonal edge in the moral graph.

Therefore the graphs in Fig. 3a and Fig. 4, respectively, are not strictly equivalent. Information displayed by one type of graph may be obscured in the other. In a directed graph, the conditioning set for independence of two variables includes just the parent variables rather than parent plus child variables (cf. Whittaker, 1990, p. 73). Following this concept the link connecting PO_4 and NO_3 in Fig. 4 would represent the conditional correlation $corr(PO_4,\ NO_3 \mid Curr,\ Elbe)$ which must not be confused with $corr(PO_4,\ NO_3 \mid Curr, Sec)$ that vanishes according to the GGM in Fig. 3a. In our example, a small sampled value (-0.08, not shown) of $corr(PO_4,\ NO_3 \mid Curr,\ Elbe)$ indicates that the correlation between PO_4 and NO_3 needed for a proper prediction of Sec can be thought of as generated by the joint impacts of $Curr$ and $Elbe$. After fitting the GGM, a positive but still small partial correlation (0.22) is consistent with the increase of the corresponding marginal correlation from 0.39 to 0.53 (cf. Tab. 4 and annotation in Fig. 4).

We would like to draw the reader's attention to the fact that both of the two marginal correlations related to the diagonal edges in Fig. 4 changed their values under fitting the correlation matrix to the GGM in Fig. 3a. By contrast, all correlations corresponding with edges kept in Fig. 3a (and converted into directed edges in Fig. 4) remained unaffected by fitting the GGM. Although the adjustment of the marginal correlation between $Curr$ and $Elbe$ from -0.44 to -0.57 might seem substantial, consequences in terms of changes in explained variances of child nodes Sal and NO_3 remain small.

4.4. The role of water temperature

So far we did not take into account water temperature WT as another parameter potentially relevant in a complex web of inter-relationships. Fig. 5 shows GGMs that were adjusted accordingly, in analogy with the two panels in Fig. 3 again with and without consideration of marine transport anomaly $Curr$.

Although the marginal correlation between NO_3 and WT (0.52) is less substantial than correlation between NO_3 and Sal (-0.67, cf. Tab. 4) the explained variance of NO_3 would nonetheless slightly increase (from 69% to 71%, cf. Figs. 3b and 5b) if in a model for NO_3 with three predictors $Elbe$, Sal, and Sec (adjacent nodes in Fig. 3b) predictor Sal was replaced by predictor WT (adjacent node in Fig. 5b). This indicates that information from WT overlaps less with information from $Elbe$ and Sec than information from Sal does. Given the value of predictor Sec, WT is also a useful second predictor for PO_4 (Fig. 5b) but turns out to be dispensable once the hydrodynamic parameter $Curr$ is known (Fig. 5a).

4.5. Graphical modelling and linear regression: Monte-Carlo-simulations

For various practical reasons specific applications will often focus on just a subset of those variables represented in a dependence graph. Variables left out may even be key variables in the

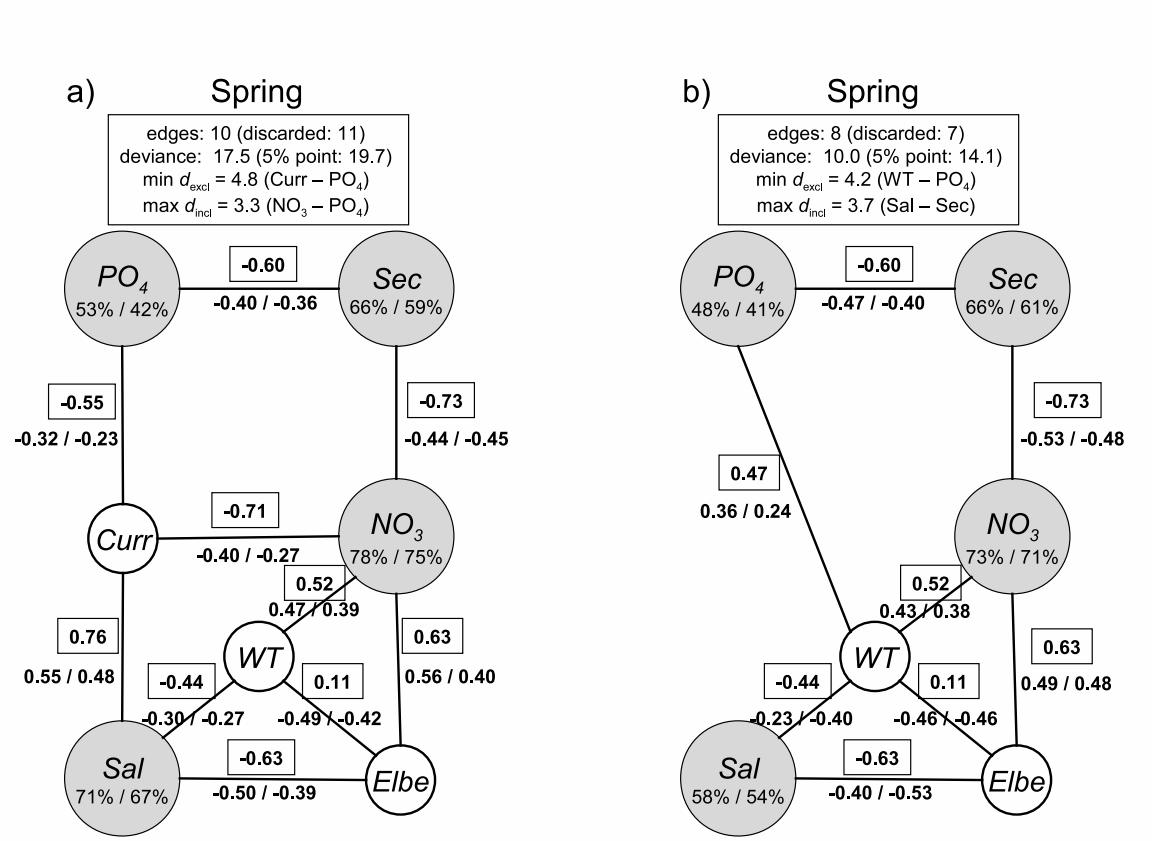

Figure 5: Same as Fig. 3 except that water temperature WT is included.

Table 5: Results of 1000 Monte-Carlo-simulations on the calibration of regression scheme (13)-(14) based on observations simulated with Eqs. (1)-(4).

					Calibration		Validation		
	a_y	b_y	a_z	b_z	R_y^2	R_z^2	R_y^2	R_z^2	$\mathrm{dev}_S(G)$ >6
N=10									
Unconstrained:	0.57±0.23	0.56±0.23	0.41±0.34	0.41±0.32	70±18%	45±22%	56±15%	11±28%	
Constrained:			0.41±0.23	0.40±0.22		37±23%		23±16%	17%
N=20									
Unconstrained:	0.58±0.14	0.58±0.15	0.41±0.21	0.42±0.21	68±12%	40±17%	62±5%	24±11%	
Constrained:			0.41±0.14	0.42±0.15		36±17%		29±6%	12%
N=50									
Unconstrained:	0.58±0.09	0.58±0.08	0.41±0.12	0.42±0.12	67±7%	36±11%	65±2%	30±4%	
Constrained:			0.41±0.09	0.42±0.09		34±11%		32±3%	6%
N=100									
Unconstrained:	0.58±0.06	0.58±0.06	0.41±0.08	0.41±0.08	67±5%	34±8%	66±2%	32±3%	
Constrained:			0.41±0.06	0.41±0.06		34±8%		33±3%	5%

Means and standard deviations of regression coefficients and explained variances obtained for different assumed sample sizes N (cf. Section 3.3). Both R_y^2 and R_z^2 (cf. Eq. (15)) were specified for a) the calibration data (sample size N) and b) independent validation data (sample size 1000). For all Z-related parameters also corresponding values constrained by the GGM in Fig. 2b are shown (for Y-related parameters fitting the correlation matrix to the GGM has no effect). The last column gives the percentage of simulations for which the deviance of the graph (cf. Eq. (9)) exceeded the 5% significance level (6.0) for 2 degrees of freedom (=number of edges discarded from the graph) so that the GGM might have erroneously been rejected.

graph. A hypothetical example is Eq. (14) which disregards the role of Y as a predictor of Z (Fig. 2a). Monte-Carlo-experiments on the calibration of Eqs. (13)-(14) based on observations simulated according to Eqs. (1)-(4), both with and without fitting sample correlations to the GGM in Fig. 2b (cf. Section 3.3), are summarized in Tab. 5. For each of the four regression coefficients in Eqs. (13)-(14) both its mean value and its standard deviation were estimated from 1000 simulations. Coefficients a_y and b_y for response variable Y remain unaffected by fitting the graphical model, as Eq. (13) is consistent with the structure of the data generating model (cf. Eq. (3)). For a_z and b_z, however, the regression coefficients differ. The graphical model is found to reduce the spread of coefficients obtained in the 1000 simulations, while estimated means of the coefficients remain more or less unaffected.

As an indication of over-fitting, during calibration the percentages of explained variance are overestimated while independent validation reveals an underperformance of the calibrated model (true values are R_y^2=66.6% and R_z^2=33.3%). The relevance of over-fitting increases with decreasing sample size N, for small samples the effect is substantial. Fitting sampled correlations to the GGM clearly improves the situation.

The last column in Tab. 5 provides the percentage of 1000 experiments in which the graphical model would have been rejected because its overall deviance is judged significant assuming a χ^2-distribution. Results show that for small samples the probability of the (correct) model being rejected is higher than the theoretical 5% chance. This problem with testing based on small samples is well known (e.g. Edwards, 1995; Shipley, 2000).

5. Discussion

Our analysis of dependences between selected variables either collected at or relevant for station Helgoland Roads was supported by a reasonably broad data basis of 42 inter-annual differences. As a key result of exploratory PCA, in spring one single pattern of covariation was found to account for about 65% of inter-annual variability in the space of six variables (Tab. 2). Tab. 3 suggests hydrodynamic transport anomaly $Curr$ being the most influential forcing parameter in this season. While river discharge is able to modify spatial distributions and related gradients in the station's vicinity, atmospherically forced regional transport patterns govern the time dependent local conditions the station is actually exposed to. Both Wiltshire et al. (2010) and Scharfe (2013) have already discussed variable strength of advection of marine water from the north west (cf. Fig. 1b) as an important factor that controls alternating impacts of either more coastal or more offshore waters in the Helgoland area particularly in spring. The present study confirms these findings. In summer association of $Curr$ with other variables is found to be very weak (Tab. 3). Reduced hydrodynamic variability (Scharfe, 2013) in combination with enhanced biological activity influencing nutrient concentrations could explain this finding.

According to the GGM in Fig. 3a the second forcing variable, river discharge $Elbe$, is directly connected with nitrate but not with phosphate. In part, this could reflect the fact that in the Elbe River concentrations of nitrate (in contrast to those of phosphate) do not follow the theoretical dilution curve but even increase with discharge and preceding rainfall (Hickel et al., 1993). However, the GGM does not necessarily represent a causal chain. The information conveyed by Fig. 3a is that conditional on $Curr$, observations of Sec (or even NO_3) are more informative about PO_4 than those of $Elbe$. Callies (2000) elaborated this kind of sufficiency relation for a four variable problem.

Checking the relevance of water temperature as an additional predictor, we found with regard to inter-annual variability that water temperatures conveyed little information on Sal and NO_3 beyond that already available from $Elbe$ and $Curr$. The relevance of edges connecting WT with one of the other two forcing parameters turned out to be very context dependent. Although sample correlation between WT and $Elbe$ is low (0.11), a corresponding high partial correlation (-0.49 in Fig. 5a) gave rise to a connecting edge. For WT and $Curr$ the situation is the opposite, regardless of a reasonable correlation (-0.44, not shown), the partial correlation is only 0.21 in the context of the graph. The substantial differences between marginal and conditional correlations (even signs are different) seem to indicate compensation for overlapping information from different predictors (collinearity). Inclusion of WT also increases the relevance of other already existing edges. For instance, in Fig. 3b removal of the edge between Sal and $Elbe$ would increase the deviance by 5.7 while in Fig. 5b the corresponding deviance difference of 23.1 is substantially larger (values not shown in the figures). When $Curr$ is included, corresponding numbers 8.2 and 11.4 for edges in Fig. 3a and Fig. 5a, respectively, differ much less.

Given that explained variances of Sal and NO_3 do not benefit all too much from the inclusion of WT, the relevance of WT as an explanatory variable is not as clear as one might have anticipated. It is doubtful that inclusion of WT in the graph contributes to a robust picture of dependences, although it cannot be ruled out that we see the limitations of representing causal effects in terms of linear relationships.

Unlike PCA, fitting a graphical model involves subjective decisions. Different edges may have comparable strengths and the removal of one edge may influence the strengths of edges retained. In our example ambiguity of the interaction structure proved not to be a major issue. Another important decision to be made, however, is when to stop model truncation (cf. Whittaker, 1990, p. 252). Unfortunately, for the GGM in Fig 3a the separation between the minimum deviance difference for

edge exclusion (should be large, here: $d_{excl} = 4.8$ for $Curr - PO_4$) and the maximum deviance difference for edge inclusion (should be small, here: $d_{incl} = 3.9$ for $Curr - Elbe$), respectively, is not fully satisfactory (cf. Tab. 4).

The low weighting of the edge $Curr - PO_4$ in Fig. 3a results from also Sec being adjacent to PO_4. According to Tab. 4 correlations between PO_4 and $Curr$ (-0.55) or Sec (-0.60) are of about the same size. Removal of the edge $Sec - PO_4$ lets the deviance difference d_{excl} for $Curr - PO_4$ substantially increase ($4.8 \rightarrow 14.9$) because then $Curr$ becomes the only predictor of PO_4. Keeping the edge $Curr - PO_4$ in Fig. 3a to some extent already anticipates the causal concept of $Curr$ impacting on PO_4 (cf. Fig. 4). A comparison of Fig. 3a and Fig. 3b also reveals that with regard to a prediction of NO_3 information from $Curr$ much overlaps with information from Sal. Including $Curr$ in the set of variables does not improve predictions of NO_3 but allows for a scheme more consistent with a causal concept.

These ambiguities bring us to the question of whether or not a selected graphical model really represents the data in a proper way. Statistical testing was not a central issue in our study. With regard to the description of an ecosystem one may accept deliberate simplifications, labelling links as irrelevant even when they seem significant in a statistical sense. The crucial criterion is whether or not a given model represents a substantial fraction of variability in the observed data. We used the overall deviance of a graph, closely related to Kullback-Leibler information divergence (Kullback, 1959), for ranking the graph with respect to alternative graphical models. The idea agrees with an approach Johnson (1999) and Anderson et al. (2000), for instance, suggested to replace hypothesis testing.

It should also be mentioned that we found the outcome of PCA to be more or less insensitive to the conditional independence model we fitted to the spring data (not shown). Callies (2005) described a small impact of graphical modelling also for canonical correlation analysis (CCA). In both cases the graphical model gives, however, additional insight into the specific structure of interactions summarized in these more conventional methods of exploratory data analysis.

There is an intrinsic relationship between graphical modelling and linear regression (Whittaker, 1990). To avoid over-fitting, multiple regression schemes should be parsimonious. A well fitted graphical model can be read in the sense that it suggests for each variable in the graph the set of adjacent variables as suitable explanatory variables. Our Monte-Carlo-experiment showed that there is an additional benefit from graphical modelling beyond the proper selection of predictors. Fitting a graphical model including intervening variables may further mitigate over-fitting by taking advantage of knowledge about inter-relationships with variables that are not directly accounted for in a regression model. This interesting effect deserves to be studied in some more detail.

The three causal layers $X_{1,2}$, Y, and Z in the example graph Fig. 2a agree with the hierarchy of three data layers we postulated also for the Helgoland Roads problem, the role of PO_4 and NO_3 as intervening variables in Fig. 4 is similar to the role of Y in Fig. 2a. In analogy with our Monte-Carlo-experiments let us assume that our main focus is on predicting Sec (corresponding with Z in Eq. (14)) as a function of the basic parameters $Curr$ and $Elbe$ (replacing X_1 and X_2). From the sample correlation matrix we would get

$$\widehat{Sec} = 0.61\ Curr - 0.18\ Elbe \qquad (\text{explained} : 50\%) \tag{16}$$

Note that none of the coefficients in Eq. (16) corresponds with an edge existing in Fig. 4. Nevertheless, the procedure of fitting sample correlations to the GGM in Fig. 3a and subsequently re-calculating regression coefficients from the modified correlation matrix (cf. top panel in Tab. 4)

modifies the values of both of the two regression coefficients:

$$\widehat{Sec} = 0.44\ Curr - 0.21\ Elbe \qquad \text{(explained : 35\%)} \tag{17}$$

Like for Z in our Monte-Carlo-experiments, after fitting the graphical model, $\widehat{Sec}$ represents less variance of Sec in the data used for calibration. It is likely that this reduction of variance explained in the calibration data can at least partly be related to a reduction of over-fitting, although for the real world data this hypothesis cannot be validated based on an independent second data set.

6. Conclusions

We propose the interpretation of fitted GGMs as fingerprints to be compared with corresponding graphs fitted to either data from other stations or the results of related model simulations. As long as variables at different stations (or represented in models) interact in similar ways, dependence graphs may resemble each other even when the underlying time series from two stations (or one station and a corresponding model) differ due to different external forcing, for instance. In this sense a GGM may be considered as an inherent attribute of a monitoring station (or numerical model).

For the variables and time scale we selected, marine transports seem to dominate the generation of inter-annual variability of spring conditions at station Helgoland Roads. Manifestation of competing influences of Elbe discharge and hydrodynamic advection of marine water from the north-west is most pronounced for nitrate and salinity. Only marine transports, however, are modelled as adjacent to phosphate which indicates a specific influence that cannot be considered as being mediated via other variables included in the model. Finally, the GGM supports a notion of Secchi depth being a function of both phosphate and nitrate.

The objective of our study was not testing mechanistic hypotheses. Given the complexity of the natural system, it will generally be difficult to disentangle direct and indirect causal effects. In particular, it is quite obvious that the seven variables selected for our study cannot provide us with a full picture of the system's behaviour. For inter-annual variations of seasonal mean conditions even the attempt to represent mechanistic processes might be challenged. Most processes act on shorter (at least seasonal) time scales and a lot of feedback mechanisms may be relevant. A concern that might deserve further studies, for instance, is that effects of marine currents represented in the graph could be spurious due to the fact that underlying weather conditions combine several important environmental factors like wind, light, and temperature in a characteristic way. It must be kept in mind that our analysis characterized correlation patterns. Shipley (2000, p. 1) states: *"As with shadows, these correlation patterns are incomplete - and potentially ambiguous - projections of the original causal processes."*

Our main purpose when fitting graphs to available data was not prediction. Nevertheless, our numerical experiments showed that knowledge about the overall interaction pattern in a larger context can help prevent fitting regression models to noise. The underlying concept is that noise in individual time series is identified by its inconsistency with variations in other variables supposed to be related to the variables of interest. In this way the graphical model exploits structural knowledge for the mitigation of problems that arise from dealing with short time series. The approach also suggests a possibility to benefit from external *a priori* knowledge about causal relationships.

Pearl (2000, p. 22) mentions the *"ruling paradigm of graphical models in statistics, according to which conditional assumptions are the primary vehicle for expressing substantive knowledge"*,

but mentions also the *"primacy of causal over associational knowledge"*. He proposes the use of directed graphs (causal networks) as a language that allows for a proper representation of the effects of manipulations of parent nodes on their child nodes. The design of such network should, however, draw on dedicated experiments which in a large scale environment like the one monitored by station Helgoland Roads are impossible. It is to be emphasized that the directed graph in Fig. 4 suggested based on a preceding analysis of empirical correlations does not comply with the strict requirements for a causal graph. We refer the reader to either Pearl's book or the monograph by Shipley (2000) for an in depth discussion of this issue.

Acknowledgements

We wish to thank all the people who have kept the Helgoland Roads time series alive over the past decades. Our thanks go to the crews of the research vessels AADE and ELLENBOGEN and the people who did the chemical analyses. We would also like to thank Alexandra Kraberg from the Biologische Anstalt Helgoland for organizing the '50 years Helgoland Roads symposium' in 2012 at which the idea for this paper was born. The study was partly conducted within the WIMO project (Scientific monitoring concepts for the German Bight), jointly funded by Niedersächsisches Ministerium für Wissenschaft und Kultur (MWK) and Niedersächsisches Ministerium für Umwelt, Energie und Klimaschutz (MU). We also greatly appreciate Eelke Folmer's constructive comments that were helpful to improve the presentation of our results.

References

D. R. Anderson, K. P. Burnham, and W. L. Thompson. Null hypothesis testing: Problems, prevalence, and an alternative. *Journal of Wildlife Management*, 64:912–923, 2000.

J. W. Baretta, J. G. Baretta-Bekker, and P. Ruardij. Data needs for ecosystem modelling. *ICES Journal of Marine Science*, 55:756–766, 1998.

Ulrich Callies. Comparative forecast evaluation: Graphical Gaussian models and sufficiency relations. *Mon. Wea. Rev.*, 128:1912–1924, 2000.

Ulrich Callies. Interaction structures analysed from water-quality data. *Ecol. Model.*, 187:475–490, 2005.

Robert G. Cowell, A. Philip Dawid, Steffen L. Lauritzen, and David J. Spiegelhalter. *Probabilistic Networks and Expert Systems*. Springer, New York, 1999.

A. P. Dempster. Covariance selection. *Biometrics*, 28:157–175, 1972.

David Edwards. *Introduction to Graphical Modelling*. Springer, New York, 1995.

Fabrice Eroukhmanoff and Erik I. Svensson. Contemporary parallel diversification, antipredator adaptations and phenotypic integration in an aquatic isopod. *PLos ONE*, 4(7):1–11, 2009.

James B. Grace. *Structural Equation Modeling and Natural Systems*. Cambridge University Press, Cambridge, UK, 2006.

W. Hickel, P. Mangelsdorf, and J. Berg. The human impact in the german bight: Eutrophication during three decades (1962-1991). *Helgoländer Meeresunters.*, 47:243–263, 1993.

Douglas H. Johnson. The insignificance of statistical significance testing. *Journal of Wildlife Management*, 63(3):763–772, 1999.

R. A. Johnson and D. W. Wichern. *Applied Multivariate Statistical Analysis*. Prentice Hall, Englewood Cliffs, New Jersey, 1992. (Third Edition).

K. G. Jöreskog. Analysis of covariance structures. *Scand. J. Statist.*, 8:65–92, 1981.

S. Kullback. *Information Theory and Statistics*. Wiley, New York, 1959.

S. L. Lauritzen and D. J. Spiegelhalter. Local computations with probabilities on graphical structures and their application to expert systems (with discussion). *J. Roy. Statist. Soc., Series B*, 50(2):157–224, 1988.

Paul M. Magwene. New tools for studying integration and modularity. *Evolution*, 55(9):1734–1745, 2001.

Judea Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, Cambridge, UK, 2000.

Mirco Scharfe. *Analysis of biological long-term changes based on hydro-climatic parameters in the southern North Sea (Helgoland)*. PhD thesis, University of Hamburg, 2013. In German.

Merja H. Schlüter, Agostino Merico, K. H. Wiltshire, Wulf Greve, and Hans von Storch. A statistical analysis of climate variability and ecosystem response in the German Bight. *Ocean Dynamics*, 58:169–186, 2008.

Bill Shipley. *Cause and Correlation in Biology: A User's Guide to Path Analysis, Structural Equations, and Causal Inference*. Cambridge University Press, Cambridge, UK, 2000.

K. Stockmann, U. Callies, B.F.J. Manly, and K.H. Wiltshire. Long-term model simulation of environmental conditions to identify externally forced signals in biological time series. In F. Müller, C. Baessler, H. Schubert, and S. Klotz, editors, *Long-term Ecological Research*, pages 155–162, Berlin, 2010. Springer.

Gábor J. Székely and Maria L. Rizzo. A new test for multivariate normality. *Journal of Multivariate Analysis*, 93:58–80, 2005.

Hans von Storch and Francis W. Zwiers. *Statistical Analysis in Climate Research*. Cambridge University Press, Cambridge, 1999.

Ralf Weisse and Andreas Plüß. Storm-related sea level variations along the North Sea coast as simulated by a high-resolution model 1958-2002. *Ocean Dynamics*, 56(1):16–25, 2006.

Joe Whittaker. *Graphical Models in Applied Multivariate Statistics*. John Wiley & Sons, Chichester, 1990.

K.H. Wiltshire, A.M. Malzahn, K. Wirtz, W. Greve, S. Janisch, P. Mangelsdorf, B.F.J. Manly, and M. Boersma. Resilience of North Sea phytoplankton spring bloom dynamics: An analysis of long-term data at Helgoland Roads. *Limnol. Oceanogr.*, 53(4):1294–1302, 2008.

K.H. Wiltshire, A. Kraberg, I. Bartsch, M. Boersma, H.-D. Franke, J. Freund, C. Gebühr, G. Gerdts, K. Stockmann, and A. Wichels. Helgoland Roads, North Sea: 45 years of change. *Estuar. Coasts*, 33:295–310, 2010.

J. Timothy Wootton. Predicting direct and indirect effects: An integrated approach using experiments and path analysis. *Ecology*, 75(1):151–165, 1994.

S. Wright. Correlation and causation. *Journal of Agricultural Research*, 20:557–585, 1921.