Parameter dependences arising from calibration of a riverine diatom model - representation in terms of posterior conditional distributions

Ulrich Callies · Christopher G. Albert · Udo von Toussaint

Received: date / Accepted: date

Abstract We address the analysis and proper representation of posterior dependence among parameters obtained from model calibration. A simple water quality model for the Elbe River (Germany) is referred to as an example. The joint posterior distribution of six model parameters is estimated by Markov Chain Monte Carlo sampling based on a quadratic likelihood function. The estimated distribution shows to which extent model parameters are controlled by observations, highlighting issues that cannot be settled unless more information becomes available. In our example, some vagueness occurs due to problems in distinguishing between the effects of either growth limitation by lack of silica or a temperature dependent algal loss rate. Knowing such indefiniteness of the model structure is crucial when the model is to be used in support of management options. Bayesian network technology can be employed to convey this information in a transparent way.

Keywords Model calibration · Overparametrization · Posterior parameter dependence · Markov Chain Monte Carlo · Bayesian network

1 Introduction

Mathematical ecosystem models differ with regard to mathematical complexity and the number of free parameters being involved. Even the most complex models vastly simplify reality, including arbitrary choices with respect to their structure. It remains as an ever-present challenge to suitably balance model complexity with the amount of data available for model calibration. Having agreed on a certain structure, over-parametrization, i.e. parameters not being controlled by observational evidence, is a ubiquitous problem. Different parameter combinations may possibly produce very similar results (e.g. Fedra et al., 1981; Beck, 1987; Oreskes et al., 1994; Spear, 1997; Brun et al., 2001). Simultaneously strengthening one process and weakening another may have no substantial overall effects. As a result, some parameters can be set to unrealistic values without contradiction with the data.
under study. In the light of inevitable model uncertainties, Fedra (1980) proposed a replace-
ment of predictions that pretend an unrealistic precision by multiple predictions covering
predictive uncertainty. He suggested models to be used for more qualitative discrimination
between different options rather than for detailed predictions.

Identifiability of model parameters needs the observed part of model output to be sen-
sitive to these parameters (Wu et al., 2019). However, lacking identifiability must be distin-
guished from model output insensitivity (Gupta and Razavi, 2018). Local sensitivity analy-
ses based on local derivatives of some model output are suitable when a model is essentially
linear. By contrast, global sensitivity analysis (GSA) takes a sampling approach to appor-
tion model output uncertainties to uncertainties of single input parameters or combinations
thereof (Saltelli et al., 2000). Model output variance may be decomposed in terms of or-
thogonal partial variances with an increasing number of input parameters contributing to
them (Sobol’ indices; Sobol’, 1993). In practice, a large number of Monte Carlo simulations
needed often makes calculation of higher order partial variances infeasible. Sudret (2008)
proposes the use of surrogate models based on polynomial chaos expansion (PCE), orig-
inally developed by Wiener (1938), as a possible way out of this difficulty. Using PCE,
Sobol’ indices can be calculated analytically.

Referring to the problem of ‘equifinality’ of different acceptable models, Beven and
Freer (2001) developed the generalized likelihood uncertainty estimation (GLUE) method-
ology in which model simulations are ranked according to their performance. Interrelation-
ships between parameters that lead to satisfactory results are implicitly represented by the
respective subset of Monte Carlo simulations. In contrast to Bayesian Monte Carlo (BMC)
(Dilks et al., 1992), for instance, GLUE does not employ likelihood in a statistically rig-
orous sense. For a comparison of GLUE with formal Bayesian approaches, including also
Markov Chain Monte Carlo (MCMC), see Vrugt et al. (2008) or Camacho et al. (2015),
for instance. Ratto et al. (2001) combined GLUE and GSA, replacing model output vari-
ability in GSA by the variability of an informal GLUE likelihood measure. Callies et al.
(2008) applied the GSA-GLUE approach to a model simulating chlorophyll \(a\) concentra-
tions at station Geesthacht Weir on the River Elbe in Germany. The present study revives
this simple model in an even further simplified version (neglect of all algae species other
than diatoms, assumption of constant maximum growth rates of diatoms, no shading by
mineral compounds), published by Scharfe et al. (2009). The Lagrangian concept of the
model, originally suggested by Schroeder (1997), nicely fits the experimental approach of
Lagrangian sampling campaigns, trying to follow water parcels during their transport (e.g.
de Ruyter van Steveninck et al., 1992; Hardenbicker et al., 2016).

According to Karrasch et al. (2001), diatoms dominate algae biomass in the Elbe River.
Our diatom-based model fits chlorophyll \(a\) observations quite well, despite its simplicity.
For the Rhine river, de Ruyter van Steveninck et al. (1992) identified a potential silica lim-
itation effect during an experiment in 1990. Generally it is very difficult to identify those
among different biological processes that really control observed phytoplankton growth.
Numerical models provide a means to at least formalize different hypotheses, estimate their
consequences and compare them with observational evidence. Complexity of our example
model is much lower than that of most other mechanistic models trying to resolve processes
in more detail (e.g. Schöl et al., 2002). A key feature of the model is its option (depending
on how parameter values are set) to explain sporadic sharp decreases of chlorophyll \(a\) con-
centrations by diatoms suffering from lack of silica. At station Geesthacht, very low silica
concentrations (below 0.1 mg Si/l) are observed during summer. However, the model offers
also temperature dependent grazing rates as a potential alternative mechanism. Hardenbicker
et al. (2016) report an experimental study on major differences between the plankton dynam-
ics in the two rivers Rhine and Elbe. They try to substantiate the hypothesis that much lower
phytoplankton densities in the Rhine than in the Elbe river might be due to grazing by inva-
sive bivalves being more abundant in the Rhine than in the Elbe river. Similarly, Waylett
et al. (2013) argue that between-year differences in grazing are likely to explain interannual
variability of phytoplankton loss observed in the upper Thames. Although the temperature
dependent loss rate assumed in our model is very simplistic compared to real world con-
ditions, model calibration nevertheless can be hoped to indicate how distinguishable such
temperature dependent mechanisms are from silica related effects.

From a mathematical point of view, the problem of over-parametrization should be al-
leviated by striving for a reduced number of model parameters. Some details addressed in
the model might be discarded. Alternatively, variables might be aggregated into a set of few
compound variables. However, lumped parameters like principal components, for instance,
often lack a clear physical interpretation. A simplified pure input-output model might be
successful in predicting variables of interest but can usually not explain why certain things
are going to happen. In process oriented simulations, each model parameter has a specific
meaning. Knowing about its interpretation will be important for any informed management
action based on integrated assessments (Hesse and Krysanova, 2016). The method proposed
in this study therefore keeps the full set of model parameters. However, instead of providing
for each of these parameters its most probable value (possibly together with an error bar),
we provide a description of the joint probability density of the full set of parameters.

We propose using well established Bayesian network (BN) technology (Pearl, 1988;
Kjaerulff and Madsen, 2008) to describe inter-relationships between parameters that jointly
establish the basis of successful simulations in agreement with existing data. The underly-
ing ensemble of successful simulations is produced performing Markov Chain Monte Carlo
(MCMC) simulations. A BN represents a joint distribution of multivariate data by its fac-
torization in terms of conditional probabilities. Various software packages are available for
that purpose. Evidence provided for any subset of parameters can be spread across the whole
network, potentially changing the marginal distributions of all other parameters.

The problem with constructing a saturated BN (retaining all possible interactions) is the
dimension and overall size of conditional probability tables needed. However, as long as all
parameters are allowed to interact, conditional marginal distributions can also be obtained by
sub-sampling from the set of successful parameter combinations generated by MCMC. In a
first step we will follow this direct approach. A benefit from using a BN based on conditional
probability tables fitted to the data arises when the goal is to focus on just the most important
interaction patterns. A BN displays such interaction structure in terms of a directed acyclic
graph (DAG). Often a DAG is seen as a means to represent cause-effect relationships (Pearl,
2000; Peters et al., 2017). Although this concept is not applicable for the example under
study, we explore graph simplification to visualize key parameter dependences.

This paper is organized in the following way. In Section 2 we first describe observa-
tional data, the model used to simulate them and the way Markov Chain Monte Carlo is
implemented in the context of model calibration. A brief introduction is given to graphi-
cal modelling, encompassing both Gaussian graphical models and Bayesian networks. A
special tool for analysing conditional posterior marginal distributions of MCMC parameter
samples is presented. In Section 3 we first illustrate model output uncertainties that arise
from posterior model parameter uncertainties. Then, dependences between calibrated pa-
rameters are explored looking at conditionalized marginal distributions. Finally, a Bayesian
network with simplified parameter dependences is devised based on the results from fitting
a Gaussian graphical model to the MCMC parameter samples. Section 4 provides a compre-
hensive discussion followed by some conclusions.
Fig. 1 The Elbe River with station Geesthacht where the chlorophyll $a$ and silica observations under study were taken. Some aspects of model forcing were obtained from stations Neu Darchau (river discharge), Schnackenburg (temperatures in 1997) and Schmilka (silica). The map also indicates the four most important tributaries.

2 Materials and methods

2.1 Observations

Observations were taken by an automatic monitoring station operated by the GKSS Research Centre Geesthacht (now Helmholtz-Zentrum Hereon) at Geesthacht Weir on the Elbe River (Elbe km 586), located some 40 km upstream of the city of Hamburg (Fig. 1). Geesthacht Weir separates the riverine part of the Elbe River from its estuary issuing into the North Sea. Quasi-continuous observations of several parameters (see Scharfe et al., 2009) obtained from an automated flow-through unit are available for the years 1997-2001. Here, we focus on concentrations of chlorophyll $a$ (obtained from observed fluorescence, using calibration based on high-performance liquid chromatography (HPLC)) and silica, observed during March-October. Chlorophyll $a$ data were collected quasi-continuously, silica on an hourly basis.

As input for our model (Section 2.2) we used 24 h means of water temperature at Geesthacht Weir, centered at noon of each day. Only for year 1997 data from station Schnackenburg (Elbe km 475) had to be used to fill existing data gaps. Global radiation on an hourly basis was available from GKSS Research Centre located in close vicinity to the weir. For model initialization, silica concentrations observed at station Schmilka near the Czech-German border (Elbe km 4, see Fig. 1) were used. This station marks the end of the upper reach of the river (about 370 km long) with a mean river discharge of about 310 m$^3$/s (compared to about 730 m$^3$/s at Geesthacht Weir 580 km further downstream). Observations at station Neu Darchau (Elbe km 536) were used to characterize river discharge.

1 From the former ARGE ELBE, since 2010 part of the "Flussgebietsgemeinschaft Elbe (FGG Elbe)"; https://www.fgg-elbe.de
2.2 Modelling chlorophyll $a$ concentrations at Geesthacht Weir

The simple model first published by Scharfe et al. (2009) aims at the simulation of chlorophyll $a$ concentrations observed at station Geesthacht Weir. Values at different times result as the end points of Lagrangian trajectories initialized at Schmilka (Fig. 1). Simulations are scheduled in such a way that each day one trajectory arrives exactly at noon. This scheduling is based on an estimated overall travel time $t = \tau$ obtained as function of discharge $Q$,

$$\frac{\tau}{\tau_{ref}} = \left(\frac{Q_{ref}}{Q}\right)^{1/3}$$

with reference values $Q_{ref} = 270$ m$^3$/s and $\tau_{ref} = 10$ days. This simple formula was found to reasonably agree with existing flow time data (Böhme et al., 2002). For a detailed illustration of the general Lagrangian approach, showing also example trajectories, the reader is referred to the original paper of Scharfe et al. (2009).

For each parcel travelling down the Elbe river the following equation for chlorophyll $a$ concentration $C_{chl}$ is solved until the parcel reaches station Geesthacht Weir,

$$\frac{dC_{chl}}{dt} = \left[\mu(t) - \sigma(t)\right]C_{chl}$$

with growth rate $\mu$ and loss rate $\sigma$. Simulations for different times are completely independent from each other, any temporal coherence at Geesthacht Weir is brought about just by the fact that external forcing (light and temperature) will be the same for trajectories overlapping in time. Each trajectory is assumed to start with the same low chlorophyll $a$ concentration ($10 \mu$g chl $a$/l), a value expected to be largely overwritten during the water parcel’s 580 km journey. What might contribute to the success of the very simple approach is that in nature there seem to be only small contributions of chlorophyll $a$ from the major tributaries (Hardenbicker et al., 2016).

In Eq. (2), both $\mu$ and $\sigma$ depend on environmental conditions and therefore vary with time $t$. The value of $\mu(t)$ results as the triple product of a constant maximum growth rate, $\mu_0$, a light dependent limitation factor $F_{light}(t)$ and a second limitation factor $F_{Si}(t)$ that reflects possible effects of limited availability of silica:

$$\mu(t) = \mu_0 F_{light}(t) F_{Si}(t)$$

Both $F_{light}(t)$ and $F_{Si}(t)$ can assume values between 0 and 1.

The model does not explicitly consider the water depth coordinate $z$. Light limitation is specified just as a vertical average over water depth $D$. Given radiation intensity $I(t)$ at the water surface, the following parametrization of $F_{light}$ is assumed,

$$F_{light} = \frac{1}{D} \int_0^D \frac{I(t) \exp^{-\lambda(t)z}}{\sqrt{K_{light}^2 + F(t) \exp^{-2\lambda(t)z}}} dz$$

where the ‘Smith formula’ (Smith, 1936) with some half-saturation constant $K_{light}$ has been used. Light attenuation $\lambda$ is assumed to be proportional to chlorophyll $a$ concentration with a constant factor $\lambda_S$:

$$\lambda(t) = \lambda_S C_{chl}(t)$$

In Eq. (4), water depth $D$ is treated as a constant, indicating that its value remains unchanged during each trajectory calculation. However, for each simulated trajectory the value of $D$ is
adjusted to water discharge observed at station Neu Darchau (about 50 km upstream of Geesthacht) at the time when this trajectory reaches Geesthacht Weir. A polynomial formula well reproduces the empirical relationship between discharge and water depth, slightly enhancing, however, small values of $D$ (see Scharfe et al., 2009, their Fig. 3).

The potential for chlorophyll $a$ development may depend on the amount of silica apportioned to each simulated trajectory in agreement with observations at station Schmilka. The simple model concept assumes that an initial reservoir of silica, $C_{Si}(t_0)$, is continuously depleted due to assimilation of silica by algae. Parameter $f_{Si}$ specifies the silica content in relation to chlorophyll $a$ in the algae biomass:

$$\frac{dC_{si}}{dt} = -\mu(t)f_{Si}C_{chl}$$  \hspace{1cm} (6)

Eq. (6) does not take into account any sources of silica like releases from the sediment or additional inputs from tributaries. Following Scharfe et al. (2009), we initialized Eq. (6) with concentrations observed at station Schmilka near the Czech-German border. Initial concentration were constrained, however, by a minimum value of 2 mg Si/l. A half-saturation constant $K_{Si}$ is introduced to specify limitation factor $F_{Si}$ in Eq. (3):

$$F_{Si}(t) = \frac{C_{Si}(t)}{K_{Si} + C_{Si}(t)}$$  \hspace{1cm} (7)

In our experiment parameter $K_{Si}$ is set to the fixed value of 0.1 mg Si/l.

Scharfe et al. (2009) identified in each of the five years 1997-2001 a short period when the model consistently failed to reproduce a fast increase of chlorophyll $a$ concentrations after a late spring chlorophyll $a$ minimum. This model deficiency could not be fixed by any parameter adjustment, indicating relevance of some processes not considered within the model framework (e.g. dominance of algae other than diatoms). The authors therefore decided to modify the model in such a way that in each year assimilation of silica is abandoned during a 1-2 week period (see Scharfe et al., 2009, Fig. 11 therein). In this study we adopt this approach to prevent that the large short-term discrepancies dominate the overall model evaluation. In all time series shown in this paper, the special periods will be highlighted.

Loss rate $\sigma$ (including also respiration) in Eq. (2) is assumed temperature dependent only if temperature $T$ exceeds 20°C, otherwise it is set constant:

$$\sigma(t) = \begin{cases} 
\sigma_0 & \text{for } T < 20^\circ C \\
\sigma_0a^{T(t)-20^\circ C} & \text{for } T \geq 20^\circ C
\end{cases}$$  \hspace{1cm} (8)

The interpretation of coefficient $a$ being greater than one remains unspecific but could cover an increased zooplankton grazing rate, for instance.

From the above equations we selected six parameters for this calibration study. These parameters are the maximum growth rate $\mu_0$, half-saturation constant $K_{light}$, light attenuation constant $\lambda_S$, algal silica content $f_{Si}$, loss rate $\sigma_0$ and coefficient $a$ for loss rate temperature dependence above 20°C.

2.3 Markov Chain Monte Carlo (MCMC)

Markov Chain Monte Carlo (MCMC) (Neal, 1993; MacKay, 2003; Yustres et al., 2012; von der Linden et al., 2014) is a method to get access to unbiased samples from potentially high-dimensional probability distributions. These samples can then be used to compute quantities
of interest like parameter means and variances. Here we are interested in samples distributed
as the posterior distribution \( p(\hat{\theta} | \hat{d}) \) for the model parameters \( \hat{\theta} \) given the data \( \hat{d} \). The
key idea underlying the MCMC approach is an iterative exploration of a target probability
distribution such that the distribution of the samples asymptotically converge to it. Proposal
steps (i.e. changing the parameter values from \( \hat{\theta} \) to \( \hat{\theta}' \)) which increase the probability are
always accepted. If, however, the probability is reduced then such a parameter step is only
accepted with an acceptance probability \( \alpha(\hat{\theta}, \hat{\theta}') \) of
\[
\alpha(\hat{\theta}, \hat{\theta}') = \min\left(1, \frac{p(\hat{\theta}' | \hat{d})}{p(\hat{\theta} | \hat{d})}\right)
\]
(9)
for symmetric proposal functions (Press et al., 2007). If a proposed step is not accepted
then the old parameter vector \( \hat{\theta} \) is added again to the chain of sampled parameter values,
otherwise the new vector \( \hat{\theta}' \) and in addition \( \hat{\theta} \) is updated to \( \hat{\theta}' \). Under some weak technical
conditions (like ergodicity, detailed balance) the distribution of this chain converges to the
desired distribution \( p(\theta | \hat{d}) \) (Hastings, 1970).

Let now \( \hat{\theta} \) denote the vector of six model parameters \( \theta_k \) (see Tab. 1) we wish to calibrate
based on existing observations. Let \( \hat{d}_t \) be \( M \) observations of chlorophyll \( a \) depending on
time \( t \) and \( m(\hat{\theta}) \) corresponding chlorophyll \( a \) simulations. Then model calibration will be
based on the minimization of the following cost function \( J(\hat{\theta}) \),
\[
J(\hat{\theta}) = \frac{1}{2M} \sum_{t=1}^{M} \left[ m(\hat{\theta}) - d_t \right]^2 - \sum_{k=1}^{6} \ln p(\theta_k)
\]
(10)
with an assumed observational error \( \sigma_{chl} \). The MCMC search algorithm should not explore
unrealistically large values for parameters or parameter combinations insufficiently con-
trolled by observations. For that reason, for each parameter \( \theta_k \) a prior distribution of the
following form is introduced:
\[
p(\theta_k) = \frac{2}{\pi} \frac{b_k}{b_k^2 + \theta_k^2}
\]
(11)
Scaling ensures that \( \int_0^\infty p(\theta_k) d\theta_k = 1 \). To adjust coefficients \( b_k \) in Eq. (11), we first specify
for each parameter \( \theta_k \) values \( \theta_k^* \) that to be exceeded should be quite unlikely (see Table 1).
Probability \( P^* \) of finding \( \theta_k \leq \theta_k^* \) reads
\[
\int_0^{\theta_k^*} p(\theta_k) d\theta_k = P^*
\]
(12)
With Eq. (11) and \( \int (b_k^2 + \theta_k^2)^{-1} d\theta_k = \arctan(\theta_k/b_k)/b_k \), coefficients \( b_k \) satisfying Eq. (12)
can then be calculated as:
\[
b_k = \theta_k^* \tan^{-1}\left(\frac{\pi}{2} P^*\right)
\]
(13)
In the following we assume a value of \( P^* = 0.9 \).

All parameters \( \theta_k \) considered in this study are also constrained to positive values. In the
MCMC algorithm that is warranted by flipping negative proposed parameter test values into
the positive range by taking absolute values. This approach maintains the detailed balance
necessary for the convergence of MCMC (von der Linden et al., 2014). Freni and Mannina
(2010) studied the implications of choosing prior distributions on uncertainty analysis. Here
we sought to keep prior constraints as weak as possible.
| $\mu_0$ | $\lambda_g$ | $K_{\text{light}}$ | $f_{\text{Si}}$ | $\sigma_0$ | $\alpha$ |
|---------|-------------|-----------------|----------------|-----------|---------|
| Eq. (3) | Eq. (5)     | Eq. (4)         | Eq. (6)        | Eq. (8)   |         |
| 3.5     | 0.05        | 500.            | 0.4            | 2.0       | 2.0     |
| $d^{-1}$ | (m $\mu g$ Chl)$^{-1}$ | W/m$^2$ | mg Si/mg C | $d^{-1}$ | -       |

2.4 Graphical modelling

Graphical models can be used to highlight key interrelationships between parameters, discarding dependences of minor importance. Data are represented in terms of nodes (or vertices) for each variable and a number of edges connecting them. Edges in undirected Gaussian graphical models (e.g. Edwards, 1995; Whittaker, 1990) represent partial correlations (i.e. correlations between pairs of variables when all other variables are held constant). By contrast, directed edges in Bayesian networks (e.g. Kjaerulf and Madsen, 2008; Pearl, 1988) represent conditional probability tables for response variables (child nodes) given the values of all explanatory variables (parent nodes) from which edges are pointing into them.

2.4.1 Gaussian graphical models (GGMs)

Undirected edges in a GGM represent non-zero pairwise partial correlations conditioned by all the rest of variables. Covariance selection developed by Dempster (1972) provides a general framework to assess whether or not the set of constraints displayed by a GGM contradicts observations. See Whittaker (1990) for a comprehensive presentation of the concept. More recent development of variational methods are described in Jordan (2004). Callies (2005) and Callies and Scharfe (2015) applied graphical modelling for analysing interaction structures from water quality observations, Taeb et al. (2017) used it to characterize dependencies among water reservoirs.

The basic idea of graphical Gaussian modelling is to modify a sample correlation matrix $S$ within the limits of observational uncertainty in such a way that small partial correlations are replaced by zero values. Partial correlation matrix $S_p$ is closely related to the precision matrix $S^{-1}$, zero valued elements of the two matrices coincide (see Whittaker, 1990). Among all correlation matrices that satisfy the constraints of a given GGM some matrix $V$ will fit the data best. The difference between the log-likelihoods of sample correlation matrix $S$ and $V$ provides an entropy type measure of the amount of information in the data against the interaction structure hypothesized by the graph $G$. The deviance $\text{dev}_S(G)$ 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). Specific properties of $V$ imply that the deviance assumes the following simplified form (Edwards, 1995):

$$\text{dev}_S(G) = N \ln \left( \frac{|V|}{|S|} \right)$$

(14)

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

With Eq. (14) it is straightforward to evaluate the effects of either excluding another edge from the existing graph (edge excluding deviance, EED) or re-establishing a link previously removed (edge inclusion deviance, EID). In this study we will not strictly test whether a
graph simplification is statistically significant or not. Due to the very large number of samples \((N = 10^6)\), formal statistical significance will always be satisfied. We rather try to stop graph simplification at a point when the smallest EED of graph \(G\) is clearly larger than the largest EID among all edges already discarded. It should be noted that this graph simplification is a manual procedure, complicated by the fact that removal or establishment of an edge generally will affect the relevance of all other edges.

### 2.4.2 Bayesian networks (BNs)

Contrary to Gaussian graphical models, Bayesian networks are directed acyclic graphs (DAGs) (Pearl, 1988; Kjaerulff and Madsen, 2008). Nodes of the BN represent random variables with usually discrete states, often (as in our case) obtained by binning a continuous variable into a certain number of categories. For each node, a table specifies the conditional probabilities for its states, given all possible combinations of states of the node’s ancestors according to the DAG. The joint distribution for \(N\) variables \(X_N\) is then given as a product of conditional probabilities,

\[
P(X_1, ..., X_N) = \prod_{X_i \in \{X_1, ..., X_N\}} P(X_i \mid Pa(X_i))
\]

where \(Pa(X_i)\) denotes the set of all parent nodes of node \(X_i\). For root nodes without parents (applies to at least one node in a DAG), the conditional probability \(P(X_i \mid Pa(X_i))\) is replaced by the simple prior distribution \(P(X_i)\). Structuring a BN in terms of parent and child nodes can often be related to the concept of causality (Pearl, 2000; Peters et al., 2017). However, for nodes representing parameters of a process model such an interpretation is not applicable.

Interactive BN software provides a convenient tool to explore parameter dependences empirically. However, depending on how many state categories are used for each parameter, the maximum number of \(M - 1\) parent nodes (with \(M\) denoting the number of parameters) can be a serious limitation for the application of BN software. That is why for saturated graphs (with all edges being maintained) specification of conditional marginal distributions by sampling directly from the data (see Section 2.5.2) may be preferable to specification of conditional probability tables. Omission of edges from the graph can much reduce the problem of dimensionality. For nearly multinormal distributions, graphical Gaussian modelling may guide such simplification of a BN.

### 2.5 Pre- and postprocessing

For analyzing measured data and the posterior distribution of model parameters two techniques have been implemented in the interactive proFit (Albert et al., 2019) toolkit for probabilistic reduced order model fitting, using GPflow (de G Matthews et al., 2017; van der Wilk et al., 2020) as a backend and visualization via Plotly/Dash. On the one hand this concerns the estimation of noise in measured time-series data in absence of a parameterized model. On the other hand the analysis of the posterior distribution of the calibrated parameters is facilitated by on-the-fly visualization of conditional marginal distributions.

#### 2.5.1 Noise estimation in measured data

In order to introduce a scale for the tolerated deviation in the MCMC calibration of model parameters, the random noise in the measured time-series data \(d\) has to be known or, as
in the present case, estimated. For this purpose a fairly general Gaussian process regression (Rasmussen and Williams, 2006) with a squared-exponential kernel is applied to the data. Characteristic timescale and random noise are left as free parameters and optimized to their maximum-likelihood values based on the data. This yields a decomposition of a kernel-smoothed representation of the original data plus a Gaussian noise term $\sigma_n$. In order for this estimate to be valid, the random error must be sufficiently close to a normal distribution and the characteristic timescale shouldn’t vary over time.

2.5.2 Marginal and conditional marginal posteriors

For $N$ parameters, the marginal distribution of the posterior for each model parameters $\theta_k$, given measured data $\vec{d}$ is given by

$$p(\theta_k | \vec{d}) = \int_{-\infty}^{\infty} p(\vec{\theta} | \vec{d}) d^{N-1}\theta_{i\neq k}. \quad (16)$$

Here these marginal distributions $p(\theta_k | \vec{d})$ are approximately computed by taking sums of MCMC data inside histogram intervals. Similar to software for BNs, proFit allows to interactively restrict values of certain parameters to intervals $(\theta^A_l, \theta^B_l)$ of a certain histogram bar and observe the influence on conditional marginal distributions

$$p(\theta_k | \vec{d}, \theta^A_l, \theta^B_l) = \frac{p(\theta_k, \theta^A_l, \theta^B_l | \vec{d})}{p(\theta^A_l | \vec{d})} = \frac{\int_{-\infty}^{\theta^B_l} \int_{\theta^A_l}^{\theta^B_l} p(\vec{\theta} | \vec{d}) d\theta_1 d^{N-2}\theta_{i\neq k, l}}{\int_{-\infty}^{\theta^B_l} \int_{\theta^A_l}^{\theta^B_l} p(\vec{\theta} | \vec{d}) d\theta_1 d^{N-1}\theta_{i\neq l}} \quad (17)$$

of each parameter. This enables a fast exploration in parameter spaces that are too high-dimensional to be visualized directly.

3 Results

3.1 MCMC sampling

Based on the cost function defined in Eq. (10), MCMC was used to explore the posterior joint distribution of parameters $\vec{\theta}$. An observational error (standard deviation) of $\sigma_{chl} = 5 \mu g chl/l$ has been determined via the maximum-likelihood estimate of the noise term from a Gaussian process regression as described in section 2.5.1. 1000 parallel chains were calculated with 1000 iterations each, starting at randomized locations in parameter space. For each chain, five warm-up trajectories with 500 iterations were calculated. To reduce the computational burden, chlorophyll $a$ concentrations at Geesthacht Weir were simulated for just every third day. Acceptance rates close to 35 % for all parameters indicated a reasonable choice of the individual proposal step size of the MCMC algorithm (MacKay, 2003; von Toussaint, 2011).

Simulations based on the $10^6$ feasible parameter combinations obtained from MCMC are summarized in Fig. 2. All time series were simulated based on the same parameters, optimized for chlorophyll $a$ observations at Geesthacht Weir during the five year period 1997-2001 (see Table 2). Black lines represent simulations based on those parameters for which the five year cost function takes its minimum. Observations (daily mean values) are shown in blue. Only every third observation has a simulated counterpart.

In Fig. 2, the spread in model outputs arising from parameter uncertainties is represented by the means of Box-Whiskers-Plots. 50 % of simulations are close too each other (magenta
Fig. 2 Chlorophyll $a$ observations (blue) and corresponding simulations optimized to reproduce chlorophyll $a$ observations in the five year period 1997-2001. Black lines represent the simulation for which the cost function (Eq. (10)) with $\sigma_{chl} = 5 \mu g$ chl/l assumes a minimum value. Box-Whiskers-Plots represent the spread among simulations based on the $10^6$ feasible parameter sets obtained from MCMC. Yellow bars indicate periods during which in the model assimilation of silica was abandoned (see Section 2.2).
Table 2 Parameters underlying Fig. 3 (cost: first term in Eq. (10), evaluated for 1997-2001; prior: second term in Eq. (10))

| Minimum cost function | $\lambda_C$ | $K_{\text{light}}$ | $f_{\text{Si}}$ | $\mu$ | $\sigma_0$ | colour in Fig. 3 | cost / prior |
|-----------------------|-------------|-------------------|----------------|------|-------------|-----------------|-------------|
|                       | (m $\mu$g Chl)$^{-1}$ | W/m$^2$ | mg Si/mg C | - | d$^{-1}$ | - |             |
| 0.0118 | 41.9 | 0.168 | 1.25 | 1.19 | 0.150 | black | 14.2 / 4.8 |
| Max. chlorophyll $a$ on 11 May | 0.0054 | 216 | 0.145 | 1.34 | 1.89 | 0.156 | green | 23.6 / 6.6 |
| Max. chlorophyll $a$ on 10 July | 0.0172 | 2.8 | 0.214 | 1.50 | 0.62 | 0.026 | brown | 26.2 / 4.8 |
| Min. chlorophyll $a$ on 31 July | 0.0081 | 29.8 | 0.296 | 2.82 | 0.71 | 0.011 | red | 25.7 / 5.9 |

![a) Chlorophyll $a$, 2001](image1)

![b) Silica, 2001](image2)

Fig. 3 Upper panel: Data (blue) and best fitting simulation (black) of chlorophyll $a$ including uncertainties (Box-Whiskers-Plots), copied from Fig. 2e. Additionally, three simulations are shown that produce the maximum simulation at 11 May (green), 10 July (brown) or the minimum value on 31 July (red). The underlying parameter sets are listed in Table 2. Lower panel: Corresponding data, simulations and simulation uncertainties for SiO$_2$. Yellow bars indicate periods during which in the model assimilation of silica was abandoned (see Section 2.2).

coloured boxes). However, there is a surprisingly large spread between the extremes for each day. One may ask how this goes together with the positive evaluation as a reasonable simulation.

First, larger deviations may occur in simulations for specific years because model performance was evaluated for the five year period in total. Second, the relevance of different parameters for model output depends on environmental conditions (e.g. temperature, availability of silica) at the time of interest, so that particularly large values of some parameter may lead to large anomalies at certain times while being of minor importance during other, probably longer, periods. To illustrate this effect, Fig. 3a combines the optimum simulation...
for 2001 with three other simulations that produce the most extreme chlorophyll $a$ values for 11 May, 10 July and 31 July, respectively. Table 2 compares the parameter values these three simulations are based on with the optimum (i.e. minimum cost function) simulation as a reference.

In Fig. 3a, the green curve deviates from the others in that it shows particularly large peak values in May and June. This might be explained by a large maximum growth rate $\mu_0$ in combination with a large (compared to the reference) half-saturation constant $K_{\text{light}}$ (Table 2). The latter assumption partly compensates for the large $\mu_0$ but at the same time makes growth rate $\mu$ more sensitive to variable light conditions (see Eq. (4)). The brown curve to some extent follows an opposite approach, which results in a rather smooth simulation. At the end of July and August it is then the red simulation that much overestimates two minima of chlorophyll $a$. This simulation is based on a very large parameter $a$ (Table 2) which governs temperature dependence of algae loss rates above the 20°C threshold. The pronounced dips coincide with short periods of high temperatures (Scharfe et al., 2009, Fig. 9g therein).

Although algal silica content $f_{Si}$ underlying the red curve in Fig. 3a is the highest among the three example simulations (Table 2), the very low chlorophyll $a$ concentrations imply low consumption of silica and therefore coincide with peak concentrations of this nutrient (Fig. 3b). The upper bounds of silica ranges indicated in Fig. 3b are more or less identical with concentrations prescribed at Schmilka were trajectories start (in the first half of June the imposed lower bound of 2 mg Si/l can be noticed) and arise from zero consumption of silica. It is interesting to see that generally silica simulations look quite reasonable, although they were not used for model calibration. In fact it turned out that their inclusion did not much affect the overall outcome of model calibration (not shown). Note that the simulated unrealistic increase of silica in June is due to the abandoning of silica consumption during the shot period highlighted in yellow.

3.2 Principal component analysis of feasible parameter combinations

Posterior parameter dependences greatly influence model behaviour, but their effects are not easily recognized in higher dimensions. Conventional principal component analysis of the parameter correlation matrix may be applied. Scaling is needed to get rid of different physical dimensions. If the six selected parameters in our study were strictly independent from each other, each of them (and also each principal component) would contribute 16.7% of overall parameter variability. It turns out, however, that already the first two principal components (PCs) describe 80% to overall parameter uncertainty (see Fig. 4). The spectrum of eigenvalues $\lambda_i$ can be used to estimate the statistical degree of freedom (dof) in parameter space (Fraedrich et al., 1995):

$$\text{dof} = \frac{M^2}{\sum_{i=1}^{M} \lambda_i}$$  \hspace{1cm} (18)

Here we get $\text{dof}=2.7$ for logarithmized (except $f_{Si}$) data in $M = 6$ dimensions. The logarithm was applied as five marginal distributions showed tails towards large values (see Section 3.3, Fig. 5a). Empirical orthogonal functions (EOFs) describe the structure of parameter covariation underlying each mode of variability (von Storch and Zwiers, 1999). According to Fig. 4, the two leading EOFs do not suggest a grouping of parameters or separation between different processes. Only the third EOF (explaining 13.8% of variance) clearly focusses on an interplay between parameters $\lambda_S$ and $K_{\text{light}}$ (see Eqs. (4)) and (5)).
3.3 Exploring conditional marginal distributions

Discrete marginal parameter distributions, with continuous values of each parameter being binned into 10 classes, are shown in Fig. 5a. Different colours were used to better distinguish between parameters related to different processes. Most distributions have tails towards large values. The only parameter showing a symmetric distribution is silica content $f_{Si}$.

We now study implications of assigning specific values to parameter subsets. Confining the value of one parameter may narrow the feasible ranges of other parameters and possibly shift the peaks of their marginal distributions. Fig. 5b compares consequences of assigning an either low or high value to maximum algal growth rate $\mu_0$. The respective choice impacts other parameters to a variable extent. The parameter probably most affected is $K_{light}$, while impact on parameter $a$, for instance, remains small. Fig. 5c illustrates the effects of additionally assuming an either low or high algal silica content $f_{Si}$. It turns out that this very much affects parameters $\sigma_0$ and $a$, while much smaller effects occur for $\lambda_S$, for instance. This is a first indication that silica content is in fact a key variable in the overall model concept.

The widths of marginal parameter distributions depend on the scaling of model-observation discrepancies in the likelihood function Eq. (10), achieved by specifying standard deviation $\sigma_{chl}$. Fig. 6a shows the results of choosing $\sigma_{chl} = 1 \mu g\ chl/l$ instead of $5 \mu g\ chl/l$. The assumed high accuracy of observations prevents divergence of the MCMC process and allows to abandon provision of prior information on parameter distributions. Resulting marginal parameter distributions are very concentrated and located near the maxima of those distributions that were derived assuming larger observational uncertainty in combination with an estimated prior distribution (Fig. 5a).

From Eq. (18) the effective statistical dimension of feasible parameter space was estimated to be less than three (dof=2.7). This means that already fixing the values of 2-3 parameters greatly constrains the joint distribution of the six parameters under study. To substantiate this expectation, we constrain Fig. 5a by entering again evidence for the two parameters $\mu_0$ and $f_{Si}$, now selecting those values that are most likely according to Fig. 6a.
### (a) Unconditional marginal distributions

| μ | λ | Kμ0 | pμa | qμa |
|---|---|------|-----|-----|
| 0.0 - 0.6 | 0.05 | 0.00 | 0.00 | 0.00 |
| 0.6 - 1.2 | 0.05 | 0.10 | 0.12 | 0.10 |
| 1.2 - 1.8 | 0.05 | 0.15 | 0.23 | 0.20 |
| 1.8 - 2.4 | 0.05 | 0.20 | 0.31 | 0.28 |
| 2.4 - 3.0 | 0.05 | 0.25 | 0.44 | 0.40 |
| 3.0 - 4.2 | 0.05 | 0.35 | 0.59 | 0.54 |
| 4.2 - 5.4 | 0.05 | 0.40 | 0.62 | 0.57 |
| 5.4 - 6.0 | 0.05 | 0.45 | 0.69 | 0.64 |

### (b) Conditioning on either low or high maximum growth rate μ0

| μ | λ | Kμ0 | pμa | qμa |
|---|---|------|-----|-----|
| 0.0 - 0.6 | 0.05 | 0.00 | 0.00 | 0.00 |
| 0.6 - 1.2 | 0.05 | 0.10 | 0.12 | 0.10 |
| 1.2 - 1.8 | 0.05 | 0.15 | 0.23 | 0.20 |
| 1.8 - 2.4 | 0.05 | 0.20 | 0.31 | 0.28 |
| 2.4 - 3.0 | 0.05 | 0.25 | 0.44 | 0.40 |
| 3.0 - 4.2 | 0.05 | 0.35 | 0.59 | 0.54 |
| 4.2 - 5.4 | 0.05 | 0.40 | 0.62 | 0.57 |
| 5.4 - 6.0 | 0.05 | 0.45 | 0.69 | 0.64 |

### (c) Conditioning on high μ0 plus either low or high algal silica content $f_{Si}$

| μ | λ | Kμ0 | pμa | qμa |
|---|---|------|-----|-----|
| 0.0 - 0.6 | 0.05 | 0.00 | 0.00 | 0.00 |
| 0.6 - 1.2 | 0.05 | 0.10 | 0.12 | 0.10 |
| 1.2 - 1.8 | 0.05 | 0.15 | 0.23 | 0.20 |
| 1.8 - 2.4 | 0.05 | 0.20 | 0.31 | 0.28 |
| 2.4 - 3.0 | 0.05 | 0.25 | 0.44 | 0.40 |
| 3.0 - 4.2 | 0.05 | 0.35 | 0.59 | 0.54 |
| 4.2 - 5.4 | 0.05 | 0.40 | 0.62 | 0.57 |

**Fig. 5** Each line combines 6 histograms that represent posterior marginal distributions of calibrated parameters (black bars). Background colours are used for grouping parameters into those related to algal growth (green), silica content (blue) and algal loss (brown). Grey coloured histograms indicate that specific evidence regarding the respective parameter has been entered. To ease comparison, white contours in conditional distributions reproduce the unconditional distributions.
As a result (Fig. 6b) the marginal distribution for all remaining four variables shrink in reasonable agreement with what one gets assuming high precision data (Fig. 6a).

It is interesting to see how posterior parameter distributions differ when calibration is conducted using data from individual years (Fig. 7). For 1999 and 2001, posterior marginal distributions of algal silica content $f_{Si}$ look similar to the one obtained for the full five year period 1997-2001 (Fig. 5a). Relatively low values of algal silica content $f_{Si}$ specified in agreement with chlorophyll $a$ observations in 1998 (Fig. 7b) favour the hypothesis that in 1998 algal loss might explain a good deal of chlorophyll $a$ concentration variability observed at Geesthacht Weir. The opposite is true for 1997, a year for which the posterior marginal distribution of silica content is shifted towards clearly higher values. For both years 1997 and 1998, marginal posterior distributions do hardly depend on whether or not silica observations were used in addition to chlorophyll $a$ observations (not shown)\(^2\). This observation holds also for 1999 and 2001.

3.4 Bayesian network assuming simplified parameter inter-relationships

So far we obtained marginal distributions by sampling directly from the MCMC output. However, using advanced BN software with a graphical user interface can very much ease exploration of parameter dependences. Unfortunately, for a larger number of parameters, representation of the joint probability by means of a BN with all edges being kept becomes prohibitive due to the high dimensionality of the conditional probability tables needed. Resolving marginal distributions with a lower number of bins reduces this dimensionality. To avoid coarse resolution, a representation assuming a reduced number of connecting edges often would be the preferred option. However, eliciting the proper simplified structure of a BN (Kjaerulf and Madsen, 2008) is much more difficult than just specifying conditional

\(^2\) In 1997 silica observations are available only in autumn.
Fig. 7 Marginal distributions of BNs calibrated using chlorophyll $a$ data from different years. The overall setup agrees with that underlying Fig. 5a, apart from the different time periods model calibration refers to.

Just as principal component analysis, Gaussian graphical modelling relies on the parameter correlation matrix. An undirected GGM represents correlations (partial correlations) between pairs of variables that are not mediated by any third variables. The idea is to adjust probability tables for an already given dependence structure. Here we use Gaussian graphical modelling (Section 2.4.1) as an auxiliary technique, although this concept must be clearly distinguished from the BN approach.
the posterior parameter correlation matrix in a way that generates zero partial correlations. Statistical relevance of any further truncation of the GGM is assessed statistically in the light of existing data. Here we applied a more qualitative concept, looking for a graph in which all missing edges are clearly less relevant than those maintained (see Section 2.4.1). In the truncated graph we found in agreement with this criterion (Fig. 8), six out of 15 edges of the saturated graph could be removed.

Table 3 shows the original correlation matrix $S$ of feasible parameter combinations, EOFs of which were displayed in Fig. 4. Note the particularly strong correlations (either positive or negative) between algal silica content $f_{Si}$ and the two algal loss related parameters $\log(a)$ and $\log(\sigma_0)$ (see Eq. (8)). Table 3 compares $S$ with correlation matrix $V$, simplified to agree with the graphical structure shown in Fig 8. The iterative proportional fitting algorithm (Whittaker, 1990) was applied for adjusting matrix $V$ to conform to this independence structure. Numbers in bold type correspond to edges that were maintained. These correlations generally remain unchanged (see Whittaker, 1990).

Table 3 also shows partial correlation matrices $S_p$ and $V_p$. Parameters $f_{Si}$ and $\log(a)$ provide an example of how much correlation and partial correlation can differ. Partial correlations that correspond with edges missing in Fig. 8 assumed zero values. Other partial correlations just changed their strengths, thereby adapting to the elimination of some mediating variables.

To assess implications of fitting the graphical model, we quantify posterior parameter dependences in terms of the percentages of uncertainty (variability) of every single parameter that can be modelled as a linear function of all other parameters. Table 4 shows these explained variances as derived from $S$ and $V$, respectively. Generally high values conform...
Fig. 9 A BN with directed edges only where undirected edges exist in the GGM (Fig. 8). The BN is shown in a state after evidence for both \( \mu_0 \) and \( f_{Si} \) was entered. Calculations were performed using Netica. Conditional marginal distributions obtained from the truncated BN well reproduce those shown in Fig. 5c. White contours indicate unconditional distributions.

Table 4 Portions of parameter variability that can be modelled as a linear function of all other five parameters. Values are specified for correlation matrices \( S \) and \( V \) (cf. Section 3.4).

|          | \( \log(\lambda_S) \) | \( \log(K_{light}) \) | \( f_{Si} \) | \( \log(\mu_0) \) | \( \log(\sigma_0) \) |
|----------|------------------------|------------------------|--------------|-------------------|-------------------|
| \( S \)  | 74%                    | 72%                    | 94%          | 84%               | 89%               | 96%               |
| \( V \)  | 73%                    | 69%                    | 93%          | 83%               | 87%               | 96%               |

to the low dimensionality (dof=2.7) of the posterior parameter space. Small differences between the results from either \( S \) or \( V \) are in favour of the simplified GGM. Also the leading EOFs obtained from \( V \) resemble those for \( S \) in Fig. 4 (not shown).

Interpretation of the GGM in Fig. 8 is the following. Assume, for instance, that parameters \( f_{Si} \) and \( \sigma_0 \) were given. According to the GGM this would block all interaction between \( a \) and the remaining 3 parameters. Interaction means that changes of any of the 3 parameters \( \mu_0, \lambda_S \) and \( K_{light} \) could compensate for effects of changing \( a \) and vice versa. Given fixed values for \( f_{Si} \) and \( \sigma_0 \), this mechanism would be suppressed.

Fig. 9 shows a BN with directed edges replacing undirected edges of the GGM in Fig. 8. Generally it is not possible to exactly translate a GGM into a directed BN. Conversely, given any BN, a corresponding conditional independence graph can be derived by first connecting all joint predecessors (parent nodes) of all child nodes. Then all directed edges are converted into undirected ones, giving the so-called moral graph (Cowell et al., 1999). The moral graph derived from the BN in Fig. 9, for instance, would also contain an edge between...
\( \lambda_S \) and \( f_{\text{Si}} \), as these two nodes have joint children \( K_{\text{light}} \) and \( \mu_0 \). Such edge is missing in the GGM in Fig. 8). Hence, the GGM tends to be more restrictive than the BN and it can therefore be expected, that the simplified BN (6 edges out of 15 edges were removed; maximum number of parents is 3 instead of 5 for the saturated graph) behaves similar to the saturated BN.

To substantiate this agreement, Fig. 9 replicates the experiment shown in the upper panels of Fig. 5c, assigning small values to both maximum growth rate \( \mu_0 \) and silica content \( f_{\text{Si}} \). Marginal distributions of the remaining four parameters show reasonable (although not perfect) agreement with Fig. 5c.

4 Discussion

The fact that even complex environmental models drastically simplify the natural system has always attracted much scepticism (e.g. Hedgpeth, 1977). Here, a simple model with a small (compared to other models) number of parameters was found to reproduce chlorophyll \( a \) observations at Geesthacht Weir reasonably well. However, even a good fit does not prove the model’s truth (Fedra et al., 1981; Oreskes et al., 1994; Brun et al., 2001). Hornberger and Spear (1980) considered their simple model for the Peel Inlet a speculative scenario that presumes phosphorus to control algal growth. An alternative scenario putting nitrogen at the heart of the analysis was reported by Humphries et al. (1984). Similarly, in the present study it was our hypothesis that consumption of silica might be relevant to explain sudden drops of chlorophyll \( a \) concentration at station Weir Geesthacht. Model calibration shed some light on how (within the a priori specified model structure!) silica limitation and a temperature dependent loss rate could provide competing concepts to explain variability of chlorophyll \( a \) observations.

Our study benefited from the fact that although the simple model is fast, it fits the data quite well. Posterior parameter distributions could be estimated reasonably well even in six dimensions. When small observational errors were assumed, a disregard of prior knowledge did not let MCMC simulations produce infeasible parameter combinations (Fig. 6a).

According to van Straten (1983) "... one may question, whether a model is actually well-structured if the use of parameter constraints is the only way to avoid nonfeasible solutions".

For our example, it seems hardly possible to formulate a statistical model of structural model inadequacy (Wu et al., 2019). According to Camacho et al. (2015), in a Bayesian framework uncertainty in model structure is often aggregated in a more general error term. However, a vague indication in favour of the underlying model structure is the fact that posterior parameter distributions did not much depend on whether or not silica observations were used in the process of model calibration (not shown). Also without silica observations being used for model calibration, simulated silica concentrations took reasonable values. It must also be recalled, however, that the original model consistently failed to reproduce a sharp late spring chlorophyll \( a \) increase, so that in each year silica consumption had to be abandoned for a 1-2 week period. This model deficiency could not be fixed by any choice of parameters, which we see as a confirmation that the model structure is specific and cannot be adjusted any given time series.

Our analysis of posterior model parameter interactions was based on a large number of feasible realizations obtained by MCMC sampling. Modelling eutrophication in the Peel

---

3 Calculations were performed using Netica; https://www.norsys.com
inlet, Hornberger and Spear (1980) and Hornberger (1981) used a behaviour classification in a binary sense, formulating a set of six behavioural conditions to discriminate between simulations being either successful or unsuccessful in mimicking key aspects of the system’s evolution. Spear and Hornberger (1980) found separation-induced correlations between model parameters to not exceed 0.23, which is why the authors did not embark on a deeper analysis of the correlation matrix. According to Spear (1997) conventional multivariate analyses proved to be unhelpful also in other studies using the same approach. In our study, correlations were found to be much higher (see Table 3). We presume that this relates to a) our model being much more controlled by observations and b) the huge number of successful simulations ($10^6$ in our study) affordable with today’s computer power.

The aforementioned studies motivated further developments leading to the GLUE (Generalized Likelihood Uncertainty Estimation; Beven and Binley, 1992) technique, which sometimes is referred to as a pseudo-Bayesian approach. In contrast to MCMC (a formal Bayesian approach), the GLUE approach separates parameter sampling (either uniform Monte Carlo or Latin Hypercube Sampling, for instance) from likelihood evaluation (Beven and Freer, 2001). For higher dimensions the random sampling makes GLUE computationally more expensive than MCMC using sequential sampling (Camacho et al., 2015). Random sampling is also used within the Bayesian Monte Carlo (BMC) approach, a method related to GLUE but using a statistically rigorous likelihood function (Dilks et al., 1992). According to Beven and Freer (2001) the advantage of MCMC might diminish when model output likelihood has a complex shape. That seemed not to be the case in our application.

Tan et al. (2019) contrasted results of the GLUE and MCMC approaches, assessing uncertainties of nine parameters of a crop model. The authors did not address, however, an explicit specification of parameter interactions according to the posterior joint distribution. The same holds for a comparative assessment of the two approaches conducted by Li et al. (2010), referring to two conceptual hydrological models, or Camacho et al. (2015), reporting a study on the calibration of a hydraulic or hydrodynamic model using synthetic data. Our study focussed on parameter interactions and we believe that for that purpose the many samples obtainable from MCMC are a key advantage when it comes to filling a higher-dimensional parameter space. Using BMC to calibrate nine parameters of a simple water quality model, Dilks et al. (1992) found that approximately 60% of model output uncertainty could be related to covariances between model input parameters. For a lake modelling example, Fedra et al. (1981) found that focusing on meaningful simulations did not much constrain individual parameters. However, in higher dimensions they found clustering of successful parameter combinations.

Referring to a binary classification of model runs in terms of simulations being successful ("passes" or "behaviours") or not, Spear et al. (1994) described the interactions between parameters that gave rise to passes by a tree-structured estimation technique. Studying an example from groundwater pathways modelling, the authors found discontinuous localized regions, interactions of which were not reflected in a linear correlation matrix. A similar feature could not be recognized in our study. The reason might be that our analysis was based on a continuous goodness-of-fit index rather than a sharp binary classification.

Discussing half-saturation constants, Mulder and Hendriks (2014) warn that simultaneous calibration of a whole set of model parameters might not reveal the true values one would get in the laboratory. Similarly, Brun et al. (2001) emphasize that fixing selected parameters will usually bias the estimates of other parameters. However, experimental data are sparse and often not representative for the overall description of a complex natural system. An at least approximate description of the joint posterior distribution of all parameters offers a way out of this dilemma. A BN representing this distribution enables users to explore
the extent to which selecting values for any subset of parameters constrains the marginal
distributions of all other parameters (see Fig. 5), thereby explicitly addressing the concerns
raised by Brun et al. (2001).
Marginal distributions constrained by available evidence on some of the model param-
eters can also be calculated directly from MCMC output, without an involvement of spe-
cialized BN software. The advantage is that high-dimensional conditional probability tables
that hamper the analysis need not be specified. However, using BN software with a graph-
ical user interface provides a more convenient approach. BN software is also needed when
aiming at a simplified representation of posterior parameter interactions. Structural analy-
sis might suggest some kind of stepwise calibration of different process related modules
of a model. We demonstrated how in this context graphical Gaussian modelling could be
helpful, given that parameter uncertainty distributions are reasonably well represented by a
simple linear correlation matrix. Although undirected and directed graphs cannot be directly
translated into each other, the undirected graph seems nevertheless more informative than
conventional principal component analysis, for instance.
We assessed model performance in terms of a squared-error loss function (Eq. (10)). A
quadratic measure, penalizing in particular large discrepancies, is suitable to highlight prob-
lems with the simulation of major short-term changes that characterized the chlorophyll \( a \)
time series under study, for instance. Using the squared-error loss function might have been
less revealing if the general model performance had been worse (that is why we modified
the model during the short periods in May/June, indicated in Fig. 2). In case of generally
strong model data discrepancies, a linear measure of model data misfit could have been be
more adequate.
A major problem arises from the fact that, due to variations of environmental conditions,
some biological processes are pivotal only for short periods of time. Using a single scalar
quality criterion, short term discrepancies with regard to different aspects of model outcome
found in different simulations (cf. Fig. 3) will be indistinguishable.
Parameter correlations documented in the BN represent alternative model structures, an
example of the equifinality which according to Beven and Freer (2001) may occur already
for moderate model complexity. Many authors discussed that a model accommodating such
overparametrisation, not pretending existence of a single true model, may be useful (e.g.
Fedra et al., 1981; Oreskes et al., 1994; Reichert and Omlin, 1997; Beven and Freer, 2001).
Given the model structure we used, the aspect worst controlled by data is decreasing chloro-
phyll \( a \) concentrations explained by either algae growth limited by lack of silica or a
strong algal loss rate. According Fig. 5c, choosing a relatively low algal silica content \( f_{Si} \)
(and therefore a low depletion of the silica reservoir) implies a large loss rate \( \sigma_0 \) and a small
coefficient \( a \) governing effects of temperature on the loss rate (e.g. via grazing). By contrast,
for a large silica content the maximum loss rate should be set to a small value; its variation
for high temperatures becomes less constrained by the data.
One must be aware that simple (or even complex) model structures neglect many exter-
nal factors that potentially impact time series observed in nature. Observations from different
years cannot necessarily be treated like the outcomes of repeated well-defined experiments.
External factors not considered in the model (there are plenty of them) might differ in differ-
ent years. Discussing an exceptionally high chlorophyll \( a \) concentration in the River Rhine
in 2011, Hardenbicker et al. (2016) hypothesize a high growth potential of phytoplankton
which, however, most of the time is suppressed by some other environmental factor. From
a modeller’s perspective, Waylett et al. (2013) found that particularly high spring chloro-
phyll \( a \) concentrations in one year could not be explained by physico-chemical factors in
their model. They suggested variable strength of grazing loss rates, possibly depending on
over-winter survival rates of benthic filter feeders, being the most feasible explanation for such interannual differences. According to Fig. 7, our analyses for the two years 1997 and 1998 suggest an either larger (1997) or smaller (1998) impact of silica limitation, assuming an either high (1997) or low (1998) algal silica content. One must be careful to prematurely attribute such differences for individual years to just an overfitted model. Posterior joint parameter distributions represented in a BN at least document existing differences in a most comprehensive way.

An observation that maybe supports the analysis of silica being less relevant in 1998 is the fact that only in 1998 silica concentrations at Geesthacht Weir clearly exceed those upstream at Schmilka near the Czech-German border (cf. Scharfe et al., 2009, Fig. 9d therein). Fig. 10a shows the improved chlorophyll \( a \) simulations after model parameters were calibrated referring specifically to observations in 1998. In agreement with what the BN in Fig. 7 suggested, the model specialized for 1998 assumes a lower algal silica content (0.085 instead of 0.168 mg Si/mg C) and a higher algal loss rate (0.336 instead of 0.150 \( \text{d}^{-1} \)). The qualitative shape of the time series remains similar to the one based on parameters calibrated for all 5 years, just seasonal changes are a bit more pronounced.

A proper choice between the two explanations of chlorophyll \( a \) variability (either silica limitation or loss) would gain importance as soon as the model would be run in a predictive mode. However, extrapolation of a simplified model into a domain of unobserved environmental conditions would be a questionable enterprise. The goal of the present study in support of past data interpretation was rather to improve the description of imponderabilities that remain after model calibration.

5 Conclusions

For a simple model of riverine diatoms we provided a detailed description of the posterior joint distribution of model parameters, constrained by chlorophyll \( a \) observations used for model calibration. We argue that this is the most profound information on a model’s calibration potential users can get. Results showed how within the pre-specified model structure two different processes affected model output in a very similar way, thereby offering different interpretations of features in the observations. Of course it must be kept in mind that (like with any other ecosystem model) already choosing the specific model structure vastly simplified the representation of nature by the neglect of many (in fact the majority of) detailed processes.

Two challenges must be met for the approach we discussed. First, estimation of the joint distribution in a higher dimensional parameter space needs a large number of Monte Carlo simulations. MCMC seems more effective than random sampling, avoiding exploration of those regions in parameter space that produce unrealistic simulations. As MCMC would still be prohibitive for large models, provision of computationally less demanding surrogate models might offer a way out.

The second challenge is a convenient representation of the joint posterior parameter distribution. BN software estimates a set of conditional probability tables that represent results from the model calibration exercise. While direct sub-sampling the original data avoids technical problems with high-dimensional conditional probability tables, BN technology enables a simplified representation of parameter interactions. For large models a graph of model parameter interactions could possibly be assembled from a number of sub-graphs dealing with parameters of different modules of a process oriented model. Proving feasibility of such an approach is left to further research.
Fig. 10 Chlorophyll a observations (blue) and corresponding simulations optimized to reproduce chlorophyll a observations in the year 1998. The silica observations shown were not used for model calibration. Box-Whiskers-Plots represent the spread among simulations based on the $10^6$ parameter sets obtained from MCMC. Black lines represent the simulation for which the quadratic cost function for chlorophyll a data assumes a minimum value. Yellow bars indicate periods during which in the model assimilation of silica was abandoned (see Section 2.2).

Acknowledgements  Development of the water quality model we analysed had been partially funded by the European Commission, General Directorate Information Society, IST Program, through contract IST-1999-11313 (IMPACT project) and through the grant 17604-2001-02 P1B30 ISP IT. The authors acknowledge support from the Reduced Complexity Models grant number ZT-I-0010 funded by the Helmholtz Association of German Research Centers. Some graphs were produced using the Generic Mapping Tools software (GMT) available from www.soest.hawaii.edu/gmt/, Ulrike Kleeberg assisted in preparing Fig. 1.

6 Declarations

Funding

The authors acknowledge support from the Reduced Complexity Models grant number ZT-I-0010 funded by the Helmholtz Association of German Research Centers.

Conflicts of interest

The authors declare that they have no conflicts of interest.
Availability of data and material

Observations Geesthacht Weir will be made available from a repository.

Code availability

The model code will be made available from a repository together with the data being used.

Authors’ contributions

U. Callies: Original draft preparation - Numerical simulations and Bayesian networks. C. G. Albert: Data analysis, calculation and visualization of conditional marginal distributions. U. von Toussaint: Implementation of Markov Chain Monte Carlo algorithm, data analysis.

References

Albert C, Hofmeister R, RedMod Team (2019) proFit: Probabilistic response model fitting with interactive tools. DOI 10.5281/zenodo.3580488
Beck MB (1987) Water quality modeling: A review of the analysis of uncertainty. Water Resour Res 23(8):1393–1442, DOI 10.1029/WR023i008p01393
Beven K, Binley AM (1992) The future of distributed models: model calibration and uncertainty prediction. Hydrol Process 6:279–298, DOI 10.1002/hyp.3360060305
Beven K, Freer J (2001) Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. J Hydrol 249:11–29, DOI 10.1016/S0022-1694(01)00421-8
Böhm M, Eidner R, Ockenfeld K, Guhr H (2002) Ergebnisse der fließzeitkonformen Elbe-Längsschnittbereisung 26.6.-7.7.2000. In: Primärdaten. BfG-1309, Bundesanstalt für Gewässerkunde, Koblenz, p 51
Brun R, Reichert P, Künsch HR (2001) Practical identifiability analysis of large environmental simulation models. Water Resour Res 37:1015–1030, DOI 10.1029/2000WR900350
Callies U (2005) Interaction structures analysed from water-quality data. Ecol Model 187:475–490, DOI 10.1016/j.ecolmodel.2005.01.045
Callies U, Scharfe M (2015) Mean spring conditions at Helgoland Roads, North Sea: Graphical modeling of the influence of hydro-climatic forcing and Elbe River discharge. J Sea Res 101:1–11, DOI 10.1016/j.seares.2014.06.008
Callies U, Scharfe M, Ratto M (2008) Calibration and uncertainty analysis of a simple model of silica-limited diatom growth in the Elbe River. Ecol Model 213:229–244, DOI 10.1016/j.ecolmodel.2007.12.015
Camacho RA, Martin JL, McAnally W, Díaz-Ramirez J, Rodriguez H, Sucs y P, Zhang S (2015) A comparison of Bayesian methods for uncertainty analysis in hydraulic and hydrodynamic modeling. Journal of the American Water Resources Association (JAWRA) 51(5):1372–1393, DOI 10.1111/jawr.12319
Cowell RG, Dawid AP, Lauritzen SL, Spiegelhalter DJ (1999) Probabilistic Networks and Expert Systems. Springer, New York
de Ruyter van Steveninck ED, Admiraal W, Bre ebaart L, Tubbing GMJ, van Zanten B (1992) Plankton in the River Rhine: structural and functional changes observed during downstream transport. J Plankton Res 14:1351–1368, DOI 10.1093/plankt/14.10.1351
26 U. Callies et al.

Dempster AP (1972) Covariance selection. Biometrics 28:157–175
Dilks DW, Canale RP, Meier PG (1992) Development of Bayesian Monte Carlo tech-
niques for water quality model uncertainty. Ecol Model 62:149–162, DOI 10.1016/0304-
3800(92)90087-U
Edwards D (1995) Introduction to Graphical Modelling. Springer, New York
Fedra K (1980) Mathematical modelling - a management tool for aquatic ecosystems? Hel-
goländer Meeresuntersuchungen 34:221–235
Fedra K, van Straten G, Beck MB (1981) Uncertainty and arbitrariness in ecosystems
modelling: A lake modelling example. Ecol Model 13:87–110, DOI 10.1016/0304-
3800(81)90008-9
Fraedrich K, Ziehmann C, Sielmann F (1995) Estimates of spatial degrees of freedom. J
Climate 8:361–369, DOI 10.1175/1520-0442(1995)008<0361:EODOF>2.0.CO;2
Freni G, Mannina G (2010) Bayesian approach for uncertainty quantification in wa-
ter quality modelling: The influence of prior distribution. J Hydrol 392:31–39, DOI
10.1016/j.jhydrol.2010.07.043
de G Matthews AG, van der Wilk M, Nickson T, Fujii K, Boukouvalas A, León-Villagrá P,
Ghahramani Z, Hensman J (2017) GPlow: A Gaussian process library using TensorFlow.
Journal of Machine Learning Research 18(40):1–6, URL http://jmlr.org/papers/v18/16-
537.html
Gupta HV, Razavi S (2018) Revisiting the basis of sensitivity analysis for dynamical earth
system models. Water Resour Res 54, DOI 10.1029/2018WR022668
Hardenbicker P, Weitere M, Ritz S, Schöll F, Fischer H (2016) Longitudinal plank-
ton dynamics in the rivers Rhine and Elbe. River Res Applic 32:1264–1278, DOI
10.1002/rra.2977
Hastings WK (1970) Monte Carlo sampling methods using Markov Chains and their appli-
cations. Biometrika 57(1):97–109, DOI 10.1093/biomet/57.1.97
Hedgpeth JW (1977) Models and muddles. Helgoländer wiss Meeresunters 30:92–104
Hesse C, Krysanova V (2016) Modeling climate and management change impacts on
water quality and in-stream processes in the Elbe River basin. Water 8, 40, DOI
10.3390/w8020040
Hornberger GM (1981) An approach to the preliminar analysis of environmental systems. J
Environg Mgmt 12:7–18
Hornberger GM, Spear RC (1980) Eutrophication in Peel Inlet - I. The problem-defining
behavior and a mathematical model for the phosphorus scenario. Water Res 14:29–42,
DOI 10.1016/0043-1354(80)90039-1
Humphries RB, Hornberger GM, Spear RC, McComb AJ (1984) Eutrophication in Peel
Inlet - III. a model for the nitrogen scenario and a retrospective look at the preliminary
analysis. Water Res 18:389–395, DOI 10.1016/0043-1354(84)90145-3
Jordan MI (2004) Graphical models. Statistical Science 19:140–155, DOI
10.1214/088342304000000026
Karrasch B, Mehrens M, Rosenlöcher Y, Peters K (2001) The dynamics of phytoplankton
bacteria and heterotrophic flagellates at two banks near Magdeburg in the River Elbe
(Germany). Limnologica 31:93–107, DOI 10.1016/S0075-9511(01)80002-5
Kjærvulff UB, Madsen AL (2008) Bayesian Networks and Influence Diagrams - A Guide to
Construction and Analysis. Springer, New York
Kullback S (1959) Information Theory and Statistics. Wiley, New York
Li L, Xia J, Xu CY, Singh VP (2010) Evaluation o the subjective factors of the GLUE
method and comparison with the normal Bayesian method in uncertainty assessment of
hydrological models. J Hydrol 390:210–221, DOI 10.1016/j.jhydrol.2010.06.044
28 U. Callies et al.

Sudret B (2008) Global sensitivity analysis using polynomial chaos expansions. Reliab Eng Syst Safety 93:964–979, DOI 10.1016/j.ress.2007.04.002

Taeb A, Reager JT, Turmon M, Chandrasekaran V (2017) A statistical graphical model of the California reservoir. Water Resour Res 53:9721–9739, DOI 10.1002/2017WR020412

Tan J, Cao J, Cui Y, Duan Q, Gong W (2019) Comparison of the generalized likelihood uncertainty estimation and Markov Chain Monte Carlo methods for uncertainty analysis of the ORYZA_V3 model. Agron J 111:555–564, DOI 10.2134/agronj2018.05.0336

van der Wilk M, Dutordoir V, John ST, Artemev A, Adam V, Hensman J (2020) A framework for interdomain and multioutput Gaussian processes. arXiv:200301115 URL https://arxiv.org/abs/2003.01115

van Straten G (1983) Maximum likelihood estimation of parameters and uncertainty in phytoplankton models. In: Beck MB, van Straten G (eds) Uncertainty and Forecasting of Water Quality, Springer, Berlin, pp 157–171

von der Linden W, Dose V, von Toussaint U (2014) Bayesian Probability Theory: Application in the Physical Sciences. Cambridge University Press, Cambridge, UK

von Storch H, Zwiers FW (1999) Statistical Analysis in Climate Research. Cambridge University Press, Cambridge, UK

von Toussaint U (2011) Bayesian inference in physics. Reviews of Modern Physics 83:943–999, DOI 10.1103/RevModPhys.83.943

Vrugt JA, ter Braak CJF, Gupta HV, Robinson BA (2008) Equifinality of formal (DREAM) and informal (GLUE) Bayesian approaches in hydrologic modeling? Stoch Environ Res Risk Assess 23:1011–1026, DOI 10.1007/s00477-008-0274-y

Waylett AJ, Hutchins MG, Johnson AC, Bowes MJ, Loewenthal M (2013) Physico-chemical factors alone cannot simulate phytoplankton behaviour in a lowland river. J Hydrol 497:223–233, DOI 10.1016/j.jhydrol.2013.05.027

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

Wiener N (1938) The homogeneous chaos. Am J Math 60:897–936

Wu X, Shirvan K, Kozlowski T (2019) Demonstration of the relationship between sensitivity and identifiability for inverse uncertainty quantification. J Comput Phys 396:12–30, DOI 10.1016/j.jcp.2019.06.032

Yustres Á, Asensio L, Alonso J, Navarro V (2012) A review of Markov Chain Monte Carlo and information theory tools for inverse problems in subsurface flow. Computational Geoscience 16:1–20, DOI 10.1007/s10596-011-9249-z