Dealing with uncertainty and model structure deficits in long-term simulations of lake biogeochemistry

Author(s):
Dietzel, Anne

Publication Date:
2012

Permanent Link:
https://doi.org/10.3929/ethz-a-007591174

Rights / License:
In Copyright - Non-Commercial Use Permitted
Dealing with uncertainty and model structure deficits in long-term simulations of lake biogeochemistry

A dissertation submitted to
ETH ZURICH

for the degree of
DOCTOR OF SCIENCES
(Dr. sc. ETH Zürich)

presented by
ANNE DIETZEL
Dipl. Systemwiss., Universität Osnabrück
born September 24, 1983
citizen of Germany

accepted on the recommendation of
Prof. Dr. Peter Reichert, examiner
Prof. Dr. Jukka Jokela, co-examiner
Dr. Thomas Petzoldt, co-examiner

2012
Contents

Summary vii

Zusammenfassung xi

1 Introduction 1
  1.1 Outline of the Thesis ........................................... 6

2 Calibration of computationally demanding and structurally uncertain models with an application to a lake water quality model 9
  2.1 Introduction .................................................... 10
  2.2 Methods .......................................................... 13
     2.2.1 Review of inference in the presence of bias ................. 13
     2.2.2 Transformation ................................................ 18
     2.2.3 Numerical implementation ..................................... 20
     2.2.4 Summary of the approximate calibration and prediction procedure 27
  2.3 Didactical example ............................................... 28
     2.3.1 Data .......................................................... 28
     2.3.2 Model description .............................................. 29
     2.3.3 Prior distribution ............................................. 30
     2.3.4 Results and discussion ....................................... 30
  2.4 Application to biogeochemical and ecological lake model 35
     2.4.1 BELAMO: model description .................................. 36
     2.4.2 Study area ..................................................... 38
# Table of Contents

2.4.3 Data ................................................................. 38
2.4.4 Prior distribution .................................................. 39
2.4.5 Results and discussion ............................................. 41
2.5 Conclusions .......................................................... 49
2.6 Outlook .............................................................. 50
2.7 Acknowledgments ....................................................... 51

A Supplementary data to: Calibration of computationally demanding and structurally uncertain models with an application to a lake water quality model .................................................. 53
A.1 Didactical example: wrong priors ................................................... 53
A.1.1 Results and discussion .................................................. 53

3 Effects of changes in the driving forces on water quality and plankton dynamics in three Swiss lakes – long-term simulations with BELAMO ................................................. 59
3.1 Introduction ............................................................. 60
3.2 Methods ................................................................. 63
  3.2.1 Study area ........................................................... 63
  3.2.2 BELAMO: model description ........................................ 63
  3.2.3 Data ................................................................. 75
  3.2.4 Sensitivity analysis .................................................. 77
  3.2.5 Model calibration .................................................... 78
  3.2.6 Model implementation ................................................ 82
3.3 Results ................................................................. 82
  3.3.1 Sensitivity analysis .................................................. 82
  3.3.2 Model calibration .................................................... 84
  3.3.3 Uncertainty analysis .................................................. 89
  3.3.4 Nutrient mass fluxes .................................................. 91
3.4 Discussion .............................................................. 91
  3.4.1 Bias ................................................................. 92
3.4.2 Biogeochemical and ecological processes ........................................ 97
3.5 Acknowledgments .............................................................................. 98

B Supporting information to: Effects of changes in the driving forces on wa-
ter quality and plankton dynamics in three Swiss lakes - long-term simula-
tions with BELAMO 99
B.1 Data compilation and processing ....................................................... 100
  B.1.1 Chemical and physical variables ............................................... 100
  B.1.2 Biological variables .................................................................. 103
B.2 Lake-specific sensitivity analysis ....................................................... 104
B.3 Results - Model calibration ............................................................... 107

4 Bayesian inference of a lake water quality model with a Gaussian process
emulator 113
4.1 Introduction ..................................................................................... 114
4.2 Methods .......................................................................................... 115
  4.2.1 Derivation of posterior .............................................................. 116
  4.2.2 Gaussian stochastic process emulator ....................................... 118
  4.2.3 Markov chain Monte Carlo sampling based on emulator .......... 120
  4.2.4 Model predictions ..................................................................... 120
4.3 Didactical example ........................................................................... 121
  4.3.1 Model description, observation data and prior distributions ....... 121
  4.3.2 Emulator setup and performance .............................................. 123
  4.3.3 Results and discussion .............................................................. 126
4.4 Application to biogeochemical and ecological lake model .......... 128
  4.4.1 BELAMO: model description .................................................. 128
  4.4.2 Study area ............................................................................... 131
  4.4.3 Observation data ...................................................................... 131
  4.4.4 Prior distributions ................................................................. 132
  4.4.5 Emulator setup ....................................................................... 132
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.4.6 Results and discussion</td>
<td>132</td>
</tr>
<tr>
<td>4.5 Conclusions</td>
<td>137</td>
</tr>
<tr>
<td>4.6 Acknowledgments</td>
<td>140</td>
</tr>
<tr>
<td>5 Conclusions and Outlook</td>
<td>143</td>
</tr>
<tr>
<td>Bibliography</td>
<td>149</td>
</tr>
<tr>
<td>List of Figures</td>
<td>159</td>
</tr>
<tr>
<td>List of Tables</td>
<td>161</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td>163</td>
</tr>
<tr>
<td>Curriculum vitae</td>
<td>165</td>
</tr>
</tbody>
</table>
Environmental models are useful tools to test hypotheses about and enhance the understanding of biogeochemical and ecological processes in the environment. If carefully designed, these tools can quantify the knowledge of the scientific community about the future state of ecosystems under various driving conditions. Such forecasts are important for the design of ecosystem preservation plans and for the assessment of suggested management alternatives.

Lakes are one example of ecosystems that are strongly affected by anthropogenic influences. Important examples of such influencing factors are nutrient loads, toxic chemicals and climate. The explanation and prediction of a lake’s biogeochemistry and directly related plankton dynamics, especially under changes in these influencing factors, is of particular interest if drinking water is extracted from the lake. But also the preservation of biodiversity and the achievement of a good ecological state of the lake are important motivations for the monitoring and modeling of lake ecosystems. On the one hand, a model used for this purpose should be universal and accurate in the sense that it should represent the reality reasonably well, even if driving forces move into ranges that have not been observed in the past. On the other hand, such a model should not be too complex to facilitate its application even if only scarce data are available and to limit its computation time.

Lake models, as all environmental models, suffer from different sources of uncertainty. Especially the simplified representation of the reality leads to model structure deficits. In the case of the studied biogeochemical and ecological lake model BELAMO, systematic deviations of model results from observations are mainly due to the high spatial and structural aggregation within the model. Further causes are the joint calibration of three lakes with the same model with only few lake-specific parameters and the attempt to close the mass balances by mechanistically describing the mineralization processes in the sediment.

In this thesis, three contributions are made to improve the predictions of this model and
to estimate their reliability:

First, an attempt is made to decrease the bias by improving the model. This is done by the introduction of new processes and aspects formerly not considered by the model and improvements in several model formulations. This lead to results that show the ability of the model to represent the essential features of the dynamics of nutrients, dissolved oxygen, phyto- and zooplankton in the three lakes Greifensee, Lake Zurich and Walensee, while only using few lake-specific parameters.

Second, to consider the remaining discrepancies for parameter estimation and prediction, a statistical representation of the bias is added to the output of the deterministic model in addition to observation error. The assumption of normal distributions for the bias and observation errors (on a transformed scale) leads to a likelihood function that can easily be evaluated if the simulations of the deterministic model are available. In the current work, for Bayesian inference of model and error parameters, we suggest an approach to approximate the posterior distribution by a normal distribution and estimate prediction uncertainty by linearized error propagation instead of relying on a computationally expensive Markov chain. This technique is of special interest for environmental models that are typically computationally demanding. It is shown that it results in a reasonable approximation of the posterior compared to the full Markov chain Monte Carlo technique for a didactical example. The application to long-term data of Lake Zurich described by the lake model BELAMO shows the ability of this method to estimate the combined effect of parameter, input and model structure uncertainty on model output also for a computationally demanding environmental model.

Third, as the representation of the posterior by a normal distribution is a very coarse approximation, an attempt is made to get a better approximation without strongly increasing the computational requirements. This is done by using a Gaussian stochastic process emulator to interpolate the posterior probability density between values calculated for a design set of model parameters. This emulator makes it possible to calculate a full Markov chain of the approximate posterior without the need for further model evaluations. A small subsample of the Markov chain is then propagated through the full model to get uncertainty estimates of model predictions. For the didactical example as well as the application of BELAMO to Lake Zurich, the results show that emulating the posterior probability density is computationally much more efficient than the use of the full model and can be more accurate than a normal approximation to the posterior, if the Markov chain subsample used for prediction is not too small.

The current work introduces two main calibration and uncertainty estimation techniques that are suitable for computational demanding and structurally uncertain models. With
that, it makes a significant contribution to the toolbox of systems analysis techniques for environmental modeling in general. When specifically applied to the biogeochemical-ecological lake model BELAMO, it is shown that this model is able to describe the essential features of the observed data over long time periods during which large changes in nutrient loads occurred. However, the careful consideration of systematic and random errors also demonstrated that these predictions are affected by a quite large uncertainty. Nevertheless, the study shows that the suggested methods make it possible to estimate this uncertainty, which enables to analyze its reduction with further model improvements.
Zusammenfassung


Seemodelle, wie alle Umweltmodelle, leiden unter verschiedenen Quellen von Unsicherheiten. Insbesondere die vereinfachte Repräsentation der Wirklichkeit bewirkt zusätzliche Defizite in der Modellstruktur. Im Fall des in dieser Arbeit untersuchten biogeochemischen und ökologischen Seemodells BELAMO, resultieren systematische Abweichungen der Modellresultate von Beobachtungen vor allem aus der großen räumlichen und strukturellen Aggregation durch das Modell. Weitere Ursachen sind die gemeinsame Kalibrierung dreier Seen mit dem gleichen Modell mit nur wenigen see-spezifischen Parametern und der Ansatz die Massenbilanzen durch eine mechanistische Beschreibung der Mineralisierungsprozesse im Sediment zu schließen.
Diese Dissertation leistet drei Beiträge die Vorhersage dieses Modells zu verbessern und seine Zuverlässigkeit abzuschätzen:

Erstens wird ein Versuch beschrieben den systematischen Fehler, auch Bias genannt, zu verringern. Dies wird durch die Einführung zusätzlicher Prozesse und Aspekte, die vorher nicht durch das Modell berücksichtigt wurden, und die Verbesserung verschiedener Modellformulierungen versucht. Das führte zu Modellresultaten, die zeigen, dass das Modell in der Lage ist die wichtigsten Merkmale der Dynamiken der see-internen Nährstoffe, des Sauerstoffs und des Phyto- und Zooplanktons des Greifen-, Zürich- und Walensees mit nur wenigen see-spezifischen Parametern zu beschreiben.


Drittens, da die Annäherung der Posterior-Verteilung durch eine Normalverteilung eine starke Vereinfachung darstellt, wird eine verbesserte Approximation angestrebt ohne dabei den benötigten Rechenaufwand zu stark zu erhöhen. Dies wird durch die Verwendung eines Gauß-Prozess-Emulators erreicht, der die Posterior-Wahrscheinlichkeitsdichte zwischen Werten, die für ein Design-Set von Modellparametern berechnet wurden, interpoliert. Der Emulator ermöglicht es eine Markov-Kette der approximierten Posterior zu berechnen ohne weitere Modellauswertungen zu benötigen. Ein kleines Sample aus der Markov-Kette wird dann durch das volle deterministische Modell propagiert und dient so zur Abschätzung der Vorhersage-Unsicherheit. Für das didaktische Beispiel und die Anwendung von BE-
LAMO auf den Zürichsee zeigen die Resultate, dass die Emulation der Posterior-Wahrscheinlichkeitsdichte recheneffizienter ist als die Verwendung des vollen Modells. Zusätzlich ist sie genauer als die Approximation der Posterior durch eine Normalverteilung, zumindest falls ein ausreichend großes Sample aus der Markov-Kette für die Vorhersage verwendet wird.

Die vorliegende Arbeit stellt zwei Methoden zur Kalibrierung und Unsicherheitsanalyse vor, die auch und besonders für rechenintensive und strukturell unsichere Modelle geeignet sind. Damit leistet sie allgemein einen wichtigen Beitrag zur Sammlung von Techniken der Systemanalyse im Bereich der Umweltmodellierung. Durch die spezielle Anwendung auf das biogeochemisch-ökologische Seemodell BELAMO wird gezeigt, dass dieses Modell in der Lage ist die wichtigsten Merkmale der beobachteten Langzeit-Daten, die große Änderungen der Nährstoffeinträge aufweisen, zu beschreiben. Allerdings hat die sorgfältige Berücksichtigung von systematischen und zufälligen Fehlern auch nachgewiesen, dass diese Vorhersagen durch eine große Unsicherheit beeinträchtigt werden. Dennoch zeigt diese Arbeit, dass die vorgeschlagenen Methoden es ermöglichen diese Unsicherheit abzuschätzen, was auch eine Analyse ihrer Reduktion durch weitere Modellverbesserungen möglich macht.
Chapter 1

Introduction

The conservation of the environment and its biodiversity, the rehabilitation of deteriorated ecosystems and the preservation of ecosystem services are some of the largest challenges of nowadays' society. Among others, surface waters are ecosystems that are highly affected by human activities like discharge of waste water, diffuse losses of herbicides and nutrients from adjacent agriculture, losses of biocides from urban areas, use for hydropower, redirection and canalization of rivers, use for inland navigation and as recreation areas. These activities have severe impacts on water quality and the biological community. This affects human interests and even human health, for example if surface water is used as a source for drinking water. Lake Zurich, for example, is the largest drinking water resource for the city of Zurich. Although in Switzerland and other countries the external driving forces decreased in the last decades due to the installation and improvement of waste water treatment plants, changes in the use of chemicals in households and industry, changes in agricultural practice, lakes are still under risk of increasing population, climate change and potential misuse and overexploitation. Hence, it is important to monitor and control the water quality of lakes, to increase the understanding about the effects of changing driving forces and to predict the outcome of management practices. Mathematical models and associated computer simulation programs are important and widely used tools to fulfill these needs. Besides water quality, concentrations of nutrients and potentially toxic chemicals, the plankton community, as the base of the lake food web and an indicator of the state of a lake, are mostly subject of mathematical models developed for the prediction of lake biogeochemistry and ecological state.

Several lake models have already been developed with different key application areas in mind (Arhonditsis and Brett, 2004; Jørgensen, 2010; Mooij et al., 2010). These include SALMO (Benndorf and Recknagel, 1982), its further developments SALMO-1D and
Chapter 1. Introduction

SALMO-HR (Baumert and Benndorf, 2005; Petzoldt et al., 2005), (DYRESM-)CAEDYM (Bruce et al., 2006; Hamilton and Schladow, 1997; Rinke et al., 2009; Romero et al., 2004; Schladow and Hamilton, 1997; Tanentzap et al., 2007; Trolle et al., 2008) and PROTECH (Elliott et al., 1999b; Elliott et al., 1999a; Elliott et al., 2000; Elliott et al., 2005; Elliott et al., 2006; Elliott et al., 2007; Elliott and Thackeray, 2004; Reynolds et al., 2001). Unlike these lakes models (Mooij et al., 2010), the Biogeochemical and Ecological LAke MOdel (BELAMO) (Omlin et al., 2001b; Omlin et al., 2001a; Mieleitner and Reichert, 2006; Mieleitner and Reichert, 2008) aims for a mechanistic description of the biogeochemical processes as mineralization of organic particles in the sediment layers in addition to those in the water column. This also guarantees closing of the element cycles of phosphorus and nitrogen.

The model BELAMO jointly calculates mass balances of nutrients, oxygen, organic particles, phytoplankton and zooplankton in the lake water column and two sediment layers. Originally, the model was implemented for Lake Zurich (Omlin et al., 2001b; Omlin et al., 2001a). Later, it was successfully applied to the Swiss lakes Greifensee and Walensee as well (Mieleitner and Reichert, 2006). For an analysis of the effect of different plankton aggregation levels it was extended to consider different functional groups of phytoplankton (Mieleitner et al., 2008; Mieleitner and Reichert, 2008). To decrease the computational burden, which was increased by the introduction of more phytoplankton state variables into the model, a 4-box version of the model was implemented that reduced the continuous vertical resolution. The current version of BELAMO describes the epilimnion, hypolimnion, two sediment layers as well-mixed compartments and the interactions between them.

Environmental models such as BELAMO, that are used to describe the past or even predict the future biogeochemical and ecological state of an ecosystem, should represent this state reasonably well as a function of potentially changing external driving forces. This universality requires that the model adequately describes the main processes and underlying mechanisms and is therefore suitable for the prediction of future dynamics even if driving forces change (Mieleitner and Reichert, 2006). The universality of the lake model BELAMO was already investigated in earlier studies. The application to two other lakes with the same climate conditions but different trophic states than Lake Zurich already demonstrated some degree of universality (Mieleitner and Reichert, 2006). Otherwise, the results of the study of introducing functional groups did not increase the universality of the model (Mieleitner and Reichert, 2008) as the concentrations of functional phytoplankton groups turned out to be much more difficult to predict than those of total phytoplankton and zooplankton. Earlier studies were conducted only for a short simulation period of four years. In the current study the uni-
versatility of the model BELAMO is explored for long-term periods of 19 and 30 years (depending on data availability for the lakes), during which the phosphate input loads decreased significantly. Again, this study is done by an application of the model to the same three study lakes, i.e. eutrophic Greifensee, mesotrophic Lake Zurich and oligotrophic Walensee. The study is described in chapter 3 and includes several changes to the model equations and expansions of described processes to increase the universality of the model.

Additionally to the prerequisite of universality, modeling should include a profound estimation of uncertainty of the estimated parameters as well as of model predictions during calibration and prediction periods. Obvious sources of uncertainty are lack of precise knowledge of past and future driving forces and resulting uncertainty in parameter estimates that are also affected by observation errors. Furthermore, as all models are simplified representations of reality, results of environmental models are, in addition to input and parametric uncertainty, affected by errors in model structure. Differently to random measurement errors that often cause independent, random deviations between model results and observations, input and model structural errors lead to systematic deviations of model results from observations. Such systematic deviations, in the literature also referred to as bias, cause problems for the estimation of model parameters from observations if they are not adequately considered in the likelihood function of the model. In particular, the frequently made assumption of randomly and independently distributed observations around the mean model prediction leads to biased estimates of model parameters and their uncertainty if a significant part of the uncertainty results from systematic and not random and independent deviations.

Many attempts have been made to address such systematic deviations. In case the bias can be assumed to be constant in time or described by a functional dependence on influence factors, predictions can be corrected based on the identification of bias in the past (Dee and Da Silva, 1998; Christensen et al., 2008; Buser et al., 2009). If this requirement is not fulfilled, other ways have to be found to deal with the presence of bias in modeling. One way to reduce model bias is to improve the deterministic model. The lake model study in chapter 3 shows such an effort by the identification of relevant processes not yet described by and consequently introduced into the model and improvements of model equations. However, improved model structures will often require more parameters and thus increase the parameter identifiability problem. In addition, due to the complexity of natural systems, even improved model structures will still face the problem of biased output, ideally at a smaller quantitative level. This discussion shows that a technique for dealing with (remaining) bias is still required. A promising way to deal with bias resulting from unidentified input or structural errors
Chapter 1. Introduction

is to account for these by making the model stochastic (Vrugt et al., 2005; Vrugt and Robinson, 2007; Lin and Beck, 2007; Reichert and Mieleitner, 2009; Bulygina and Gupta, 2011). This also increases the model complexity, which often increases the computational burden tremendously. A computationally cheaper alternative to derive reliable uncertainty bounds of model predictions in the presence of bias is a statistical description of the bias in model output based on a Bayesian framework (Craig et al., 1996; Craig et al., 2001; Kennedy and O’Hagan, 2001; Higdon et al., 2004; Bayarri et al., 2007).

The presence of bias leads to an additional problem in model calibration. There is no objective criterion of how to weigh systematic model deviations in one variable against deviations in another variable or not even between different time periods for the same variable. In manual calibration, the modeler accounts for this problem by trying to model the “essential” patterns of observed data based on her or his understanding of the underlying principles. This leads to the necessity that a calibration procedure should be able to account for the fulfillment of calibration objectives of the modeler. To deal with this problem, manual calibration criteria were included explicitly into model calibration by using multi-objective optimization techniques (Yapo et al., 1998; Gupta et al., 1998; Madsen, 2000; Madsen et al., 2002; Gupta et al., 2003a; Boyle et al., 2003; Savić et al., 2011). The resulting “Pareto set” summarizes all combinations of parameter values for which improving one calibration criterion by changing the parameter values is only possible by worsening another criterion. It thus provides an overview of ways of considering the different calibration criteria without making a unique choice of their weights. This can provide essential insight into the calibration problem but is difficult to use as a basis for deriving uncertainty bounds of model predictions.

To consider different calibration objectives but still remain in a probabilistic framework that makes it possible to derive probabilistic parameter estimates and model predictions that account for the effects of parameter, structural and prediction uncertainty, a combination of statistical description of model bias (Kennedy and O’Hagan, 2001; Bayarri et al., 2007) and the ideas underlying multi-objective model calibration (Yapo et al., 1998; Gupta et al., 1998; Gupta et al., 2003a; Boyle et al., 2003) was suggested by Reichert and Schuwirth, 2012. This method consists of the description of system observations as the sum of deterministic model output, bias and observation error. The measurement error is assumed to be uncorrelated and normally distributed, whereas the prior knowledge of the bias is represented by a Gaussian stochastic process. The prior of the bias is chosen in a way that intends to support the description of the main patterns of the observed data by the model by specifying the “acceptable bias” in different model variables that is then used to weigh between different calibration objectives.
For the observation error, an informative prior is chosen that is based on the knowledge of the sampling and measurement processes. The method results in a posterior distribution which considers the modeler’s weights of different criteria as well as prior knowledge of model structure and parameters and observed data. This posterior distribution can then be used to derive probabilistic uncertainty bounds of model predictions.

The technique described by Reichert and Schuwirth, 2012 is an interesting approach for environmental modeling in particular in cases in which the strong coupling of different output variables makes a joint calibration to all outputs difficult. This is particularly the case for the lake model BELAMO because it makes the attempt to close the element mass balances of phosphorus and nitrogen by explicitly modeling mineralization in the sediment instead of using independent source and sink parameters of phosphate, nitrate and ammonia to describe the effect of the sediment on the water column. However, for complex, computationally demanding environmental models, the approach by Reichert and Schuwirth, 2012 is still difficult to apply because of the requirement for tens of thousands of simulation runs to get a Markov chain sample of the posterior. As this study is devoted to long-term simulations of the model BELAMO to test and improve its universality, computational efficiency is crucial. For this reason, besides the test and further development of the model BELAMO, this study also aims to derive approximate implementations of this technique that are computationally more efficient and can be applied for the long-term simulations of the lake model.

As a first approximation, the current study replaces Markov chain sampling of the posterior by the use of a normal distribution that approximates the posterior at its maximum and subsequently uses linearized error propagation to derive approximate uncertainty bounds of model predictions. The procedure is tested for a didactical example also used by Reichert and Schuwirth, 2012 and compared to the results therein. Also, it is applied to long-term simulations of the biogeochemistry of Lake Zurich with the model BELAMO in chapter 2 and used for the estimation of overall prediction uncertainty of the same model jointly applied to all three study lakes in chapter 3. In all application cases, the method is used by separating the available data set of observations into a period where the model is calibrated using these data and a prediction period where observations are only used to evaluate model performance. As the normal distribution appeared to be only a coarse approximation to the posterior distribution of the model, an alternative approach to approximate the posterior was also investigated. This approach is based on the emulation of the model output by a statistical model in the form of a Gaussian stochastic process conditioned to simulation runs for a specified design set of model parameters (Kennedy and O’Hagan, 2001; Higdon et al., 2004; Bayarri et al., 2007). As the application of these techniques to the full, dynamic model output still needs further
development (Bhattacharya, 2007; Conti et al., 2009; Conti and O’Hagan, 2010; Liu and West, 2009; Young and Ratto, 2011; Reichert et al., 2011; Castelletti et al., 2012; Reichert and Albert, ), we use the emulation only of the log posterior density of the model for statistical inference. Predictions are done by some simulations with the full model. The approach is presented in chapter 4 and used for the same two application cases as the posterior approximation technique introduced in chapter 2.

1.1 Outline of the Thesis

The current work expands the toolbox for calibration and prediction uncertainty estimation by a relatively simple, universal and computationally efficient technique that may need case-specific extensions. It exemplifies the suitability of this approach through its application to structurally uncertain and computationally demanding environmental models, especially to a lake water quality model for which universality is tested and enhanced while model deficits are decreased.

It is the goal of chapter 2 to derive a computationally efficient method for the approximation of the posterior of computationally demanding models in combination with an error model that considers model bias. The posterior of the model parameters is approximated by a normal distribution, which is derived by three different algorithms. The presented technique is tested and compared to a full Markov chain Monte Carlo technique for a didactical example of microbial growth. Furthermore, it is applied to long-term simulations of Lake Zurich with the lake water quality model BELAMO.

The focus of chapter 3 is to find model formulations and parameter values for a joint application of BELAMO to three lakes of different trophic state that lead to output of the model consistent with data over a long simulation period, to reduce the systematic model errors and increase model universality. As an important step, the remaining prediction uncertainty of the joint application of the model is also estimated with the technique introduced in 2 (it appeared that this method had to be developed first), while considering the effect of all main sources of uncertainty, including model bias.

Chapter 4 extends the method introduced in chapter 2 by emulating the posterior probability density of the parameters instead of relying on a normal approximation. The emulator is then used to calculate a Markov chain representing the posterior of
which a sub-sample is propagated to the model results to derive prediction uncertainty estimates. This technique intends to provide an improved characterization of parameter and prediction uncertainty at only moderately increased computational requirements. Again, this approach is applied to a didactical example of microbial growth of which the results are combined to the two other available techniques and to long-term simulations of Lake Zurich with BELAMO.

In chapter 5, overall conclusions are drawn and an outlook to future research directions is given.
Chapter 2

Calibration of computationally demanding and structurally uncertain models with an application to a lake water quality model

Anne Dietzel and Peter Reichert
Published in Environmental Modelling & Software 38 (2012), 129–146
Chapter 2. Calibration of computationally demanding models

Abstract Models of environmental systems are simplified representations of the reality. For this reason, their results are affected by systematic errors. This bias makes it difficult to get reliable uncertainty estimates of model parameters and predictions. A relatively simple way of considering this bias when using deterministic models is to add a statistical representation of the bias to the model output in addition to observation error and to jointly estimate model parameters, bias and observation error. When assuming normal distributions for bias and observation error, this leads to a relatively simple likelihood function that can easily be evaluated. Nevertheless, the sampling from the posterior distribution still requires long Markov chains to be calculated which can be prohibitive for computationally demanding models. In order to extend the range of applicability of this technique to computationally demanding models, we suggest to replace Markov chain sampling by a normal approximation to the posterior of the parameters and to estimate prediction uncertainty by linearized error propagation. We tested this procedure for a didactical example and for an application of the biogeochemical-ecological lake model BELAMO to long-term data from Lake Zurich. This is a good test application because the strong coupling of output variables makes it difficult to avoid bias in the results of this model. These tests demonstrate the applicability of the suggested procedure, the approximate reproduction of the results of the full procedure for the didactical example, and meaningful results for the lake model. For the latter, the results demonstrate that the assumption of a realistic likelihood function leads to the conclusion that prediction uncertainty may be high.

Keywords: Bias; Lake water quality model; Structurally uncertain model; Model complexity; Multi-objective calibration.

2.1 Introduction

Models substantially contribute to formalizing and summarizing knowledge, analyzing observations and testing hypotheses about the structure and function of environmental systems. As all models are simplified representations of reality, results of environmental models are, in addition to input and parametric uncertainty, affected by errors in model structure. All of these elements lead to systematic deviations of model results from observations. Such systematic deviations, also called model inadequacy or bias, are defined as deviations of the mean of the real output of the system from the predictions of the expected value of the model (Kennedy and O’Hagan, 2001). The presence of bias is ubiquitous in environmental modeling and, if it can be assumed to be constant
in time or described by a functional dependence on influence factors, predictions can be corrected based on the identification of bias in the past (Dee and Da Silva, 1998; Christensen et al., 2008; Buser et al., 2009). However, in many cases it may not be possible to extrapolate bias easily as it may be time-dependent and does not show a simple dependence on external influence factors or model variables. As long as model bias is not too large, the model might still reproduce key output patterns of the underlying system and the environmental scientist may be content with its explanatory power (being aware that no model will be perfect). However, when trying to identify model parameters from data by using statistical inference and when using the model for probabilistic prediction, model bias leads to a violation of typical statistical assumptions, such as randomly and independently distributed observations around the mean model prediction. This leads to unreliable uncertainty estimates of model parameters and predictions. As the definition of bias is a frequentist concept, it can either be identified by repeated experiments or by residual analysis e.g. of time series calculated at the best estimates of the true parameter values.

Different ways have been suggested to cope with this problem. Reducing model bias by improving the deterministic model, i.e. describing the mean system behavior more realistically, is the most obvious solution. However, it may also be necessary to account for elements of system dynamics not covered by the deterministic model by making the model stochastic (Vrugt et al., 2005; Vrugt and Robinson, 2007; Lin and Beck, 2007; Reichert and Mieleitner, 2009; Bulygina and Gupta, 2011). Both solutions result in a higher model complexity and can increase the computational burden considerably. As a different, computationally cheaper, option to derive reliable uncertainty bounds of model predictions, it was suggested to use a statistical description of bias in model output based on a Bayesian framework (Craig et al., 1996; Craig et al., 2001; Kennedy and O’Hagan, 2001; Higdon et al., 2004; Bayarri et al., 2007).

In addition to unreliable uncertainty bounds, another problem resulting from systematic deviations between model results and data is that frequently used model parameter estimation techniques often cause results that do not fulfill calibration objectives of the modeler. A scientist calibrating a model manually may weigh different calibration objectives, e.g. the peaks and the recessions of a hydrograph, differently. To address this problem, manual calibration criteria were included explicitly by using multi-objective optimization techniques for model calibration (Yapo et al., 1998; Gupta et al., 1998; Madsen, 2000; Madsen et al., 2002; Gupta et al., 2003a; Boyle et al., 2003; Savic et al., 2004). The resulting “Pareto set” summarizes all combinations of parameter values for which improving one calibration criterion by changing the parameter values is only possible by worsening another criterion. It thus provides an overview of ways
Chapter 2. Calibration of computationally demanding models

of considering the different calibration criteria without making a unique choice of their weights. For this reason it is typical that the Pareto set contains solutions that are more satisfying to the modeler than the solution resulting from a single-objective optimization technique that does not explicitly weigh between different calibration objectives. On the other hand, the Pareto set does not provide probabilistic information required for estimating prediction uncertainty.

To assess the parameter, structural and prediction uncertainty under the problems mentioned above, a combination of statistical description of model bias (Kennedy and O'Hagan, 2001; Bayarri et al., 2007) and the ideas underlying multi-objective model calibration (Yapo et al., 1998; Gupta et al., 1998; Gupta et al., 2003a; Boyle et al., 2003) was suggested by Reichert and Schuworth, 2012: System observations are described as the sum of deterministic model output, bias and measurement error. The measurement error is assumed to be uncorrelated and normally distributed, whereas our prior knowledge of the bias is represented by a Gaussian stochastic process. The identifiability problem between model and bias was addressed by specifying a prior for the bias which favors narrow distributions and has a mean of zero. This intends to support the description of the main output patterns by the model rather than by the bias. Identifiability can further be improved by an informative prior of the observation error about which typically more knowledge is available than about the bias. The prior of the statistical description of bias was used to weigh between different calibration objectives. The modeler can choose how much bias he or she is willing to accept in different model output variables. This approach introduces subjective elements into the calibration procedure; but this seems unavoidable in the presence of bias as there is no fundamental criterion that would specify how to “distribute” bias between different model variables. As the modeler can specify different priors for bias in different variables, or even different priors for different parts of a time series of a single variable, this procedure obviously links statistical bias description to multi-objective calibration. The result will not be the full Pareto set corresponding to different calibration objectives, but a posterior distribution which considers the modeler’s weights of different criteria. This posterior distribution can then be used to derive probabilistic uncertainty bounds of model predictions.

The technique as described by Reichert and Schuworth, 2012 is computationally much cheaper than most of the techniques involving stochastic models (Vrugt et al., 2005; Vrugt and Robinson, 2007; Lin and Beck, 2007; Reichert and Mieleitner, 2009; Bulygina and Gupta, 2011) but it still requires tens of thousands of simulation runs to get a Markov chain sample of the posterior. For computationally demanding simulation programs, this may still be prohibitive. As it is important to extend the range of applicability
of this calibration and prediction technique to computationally more demanding models, it is the goal of this paper to derive computationally efficient methods for the approximation of the posterior. The basic idea is to replace Markov chain sampling of the posterior by a normal distribution that approximates the posterior at its maximum and to use linearized error propagation to derive approximate uncertainty bounds of model predictions. Those approximations do not rely on the specific formulation of the likelihood function used herein, but could be used in other contexts where an approximation of the posterior distribution by a normal distribution seems to be meaningful and where a full Markov chain sampling is computationally too expensive. They are specifically useful to apply the technique described by Reichert and Schuwirth, 2012 to computationally demanding environmental models. To test this procedure, we applied it to the same didactical example as used by Reichert and Schuwirth, 2012 and to long-term simulations of the lake model BELAMO.

This paper is structured as follows: In section 2.2 we briefly review the key equations of the technique as described by Reichert and Schuwirth, 2012 and then derive the new approximation for its numerical implementation. The following two sections 2.3 and 2.4 are devoted to a didactical example and the application of the new technique to long-term simulations of the lake model BELAMO. Finally, in section 2.5 we draw our conclusions.

### 2.2 Methods

We briefly summarize the technique suggested by Reichert and Schuwirth, 2012 that combines statistical bias description with ideas of multi-objective model calibration and then discuss options to simplify the numerical implementation of this approach to get an approximate technique that is computationally less demanding.

#### 2.2.1 Review of inference in the presence of bias

**Likelihood function**

As mentioned in the introduction, in the presence of bias, the mean of the true model outcomes is no longer identical to the expected value of the model prediction. This means that the description of the observations as a sum of a deterministic function
Chapter 2. Calibration of computationally demanding models

describing the mean, $y^L_M(x, \theta)$, and an observation error with mean zero, $E^L(\psi)$,

$$Y^L_M(x, \theta, \psi) = y^L_M(x, \theta) + E^L(\psi) \quad (2.1)$$

is no longer valid. In this equation, the vector $Y^L_M$ of random variables represents the observations as described by the model $M$ at the observation layout $L$ that defines the output variables and the time points and locations at which they are observed or evaluated, $y^L_M(x, \theta)$ is a parameterization of the mean model response as a function of inputs, $x$ and model parameters, $\theta$, and $E^L(\psi)$ is a vector of random variables that represents the observation error, which is assumed to have zero mean and may depend on additional parameters, $\psi$. Defining the bias as the difference of the mean of the observations and the expected value of the model

$$b = E[Y^L_M(x, \theta, \psi)] - y^L_M(x, \theta) \quad (2.2)$$

we end with the modified description of the observations by

$$Y^L_M(x, \theta, \psi, \xi) = y^L_M(x, \theta) + b^L_M(x, \xi) + E^L(\psi) \quad (2.3)$$

where $b^L_M(x, \xi)$ is now an unknown deterministic function of the input, $x$, and additional parameters, $\xi$, describing the bias. Note that this bias term combines bias in the model with (potential) bias in the observation process.

The application of equation (2.3) to observed data leads to a significant identifiability problem between model and bias, as we can only get information on the sum of the two. Following the literature on statistical bias description, (Craig et al., 1996; Craig et al., 2001; Kennedy and O’Hagan, 2001; Higdon et al., 2004; Bayarri et al., 2007), we solve this problem by describing our knowledge of the bias by a probability distribution in the Bayesian sense. This makes the bias a random variable, $B^L_M(x, \xi)$, and leads to the following description of our knowledge of new observations:

$$Y^L_M(x, \theta, \psi, \xi) = y^L_M(x, \theta) + B^L_M(x, \xi) + E^L(\psi) \quad (2.4)$$

Equation (2.4) leads to a hierarchical model with the bias, $B^L_M(x, \xi)$, as an intermediate variable. Integrating out this intermediate variable, a likelihood function can be obtained as a function of the parameters, only:

$$f_{Y^L_M|\theta, \psi, \xi}(y^L \mid \theta, \psi, \xi, x) = \int f_{E^L|\psi}(y^L - y^L_M(x, \theta) - b^L \mid \psi) \cdot f_{B^L_M|x}(b^L \mid \xi, x) \, db^L \quad (2.5)$$

In this equation, $f_{E^L|\psi}$ is the probability density of the observation error and $f_{B^L_M|x}$ represents the prior density of the bias given the parameters $\xi$ and external inputs $x$. 

2.2. Methods

We assume normally distributed observation errors with mean zero and covariance matrix \( \Sigma_{E^L} \) (\( n_L \) is the number of observations in layout \( L \))

\[
f_{E^L | \Psi}(\epsilon^L \mid \psi) = \frac{1}{\sqrt{2\pi n_L}} \frac{1}{\sqrt{\det(\Sigma_{E^L}(\psi))}} \exp\left( -\frac{1}{2}(\epsilon^L)^T \Sigma_{E^L}(\psi)^{-1} \epsilon^L \right)
\]  

(2.6)

and a Gaussian stochastic process with mean zero and covariance matrix \( \Sigma_{B^L_M} \) for the bias

\[
f_{B^L_M | \Xi}(b^L \mid \xi, x) = \frac{1}{\sqrt{2\pi n_L}} \frac{1}{\sqrt{\det(\Sigma_{B^L_M}(\xi, x))}} \exp\left( -\frac{1}{2}(b^L)^T \Sigma_{B^L_M}(\xi, x)^{-1} b^L \right).
\]  

(2.7)

The choice for a Gaussian stochastic process is motivated by the fact that this is an easy formulation of our knowledge of the bias including a correlation structure and only a small number of parameters (Bayarri et al., 2007). This formulation makes it possible to separate the errors into a random observation error and a systematic component representing the bias (earlier approaches with auto-regressive error models already considered the dependence structure of the residuals but did not make the attempt to separate observation error and bias (Kuczera, 1983; Bates and Campbell, 2001; Yang et al., 2007)). Under these normality assumptions we can do the integration in (2.5) analytically and obtain the likelihood function

\[
f_{Y^L | \Theta, \Psi, \Xi}(y^L \mid \theta, \psi, \xi, x) = \frac{1}{\sqrt{2\pi n_L}} \frac{1}{\sqrt{\det(\Sigma_{E^L} + \Sigma_{B^L_M})}} \exp\left( -\frac{1}{2} \left[ y^L - y^L_M(x, \theta) \right]^T (\Sigma_{E^L} + \Sigma_{B^L_M})^{-1} \left[ y^L - y^L_M(x, \theta) \right] \right).
\]  

(2.8)

The simplest assumptions of the correlation structures of the variance-covariance matrices in (2.8) would be to assume independence between the different components of the observation error, \( \Sigma_{E^L} \), and correlation coefficients that decrease with increasing distance of the independent input variables, \( x \), for the bias, \( \Sigma_{B^L_M} \). This leads to the expressions:

\[
\Sigma_{E^L, i,j}(\psi) = \begin{cases} 
\sigma_{E^L,i}(\psi)^2 & \text{for } i = j \\
0 & \text{else}
\end{cases}
\]  

(2.9)

and

\[
\Sigma_{B^L_M, i,j}(\xi, x) = \begin{cases} 
0 & \text{if } y^L_i \text{ and } y^L_j \text{ are of different type} \\
\sigma_{B^L_M,i}(\xi)\sigma_{B^L_M,j}(\xi) \exp\left( -\sum_k \beta_k(\xi)(x_{i,k} - x_{j,k})^2 \right) & \text{else}
\end{cases}
\]  

(2.10)
The smoothing parameter $\beta_k$ for input dimension $k$ (e.g. time or space) in this equation relate to the corresponding correlation lengths, $\tau_k = 1/\sqrt{\beta_k}$. Equation (2.10) assumes no correlation between output variables of different type, e.g. concentrations of different substances. This assumption, as well as the specific form of equation (2.10) is not crucial for the procedure but will be used in our applications.

**Inference**

We can infer parameter values from the posterior density that is proportional to the prior $f_{\Theta, \Psi, \Xi}$, of the parameters times the likelihood function (2.5). With observations, $y^{L_1}$, at points in time and/or space described by the observation layout $L_1$, we obtain:

$$f_{\Theta, \Psi, \Xi|Y^{L_1}_M}(\theta, \psi, \xi \mid y^{L_1}, x) \propto f_{\Theta, \Psi, \Xi}(\theta, \psi, \xi) \cdot f_{Y^{L_1}_M|\Theta, \Psi, \Xi}(y^{L_1} \mid \theta, \psi, \xi, x) \quad .$$

(2.11)

Using the normal distributions (2.6) and (2.7) and the resulting likelihood function (2.8), we get:

$$f_{\Theta, \Psi, \Xi|Y^{L_1}_M}(\theta, \psi, \xi \mid y^{L_1}, x) \propto \frac{f_{\Theta, \Psi, \Xi}(\theta, \psi, \xi)}{\sqrt{\det(\Sigma^{E L_1}_M + \Sigma^{B L_1}_M)}} \cdot \exp \left( -\frac{1}{2} \left[ y^{L_1}_M - y^{L_1}_M(x, \theta) \right]^T (\Sigma^{E L_1}_M + \Sigma^{B L_1}_M)^{-1} \left[ y^{L_1}_M - y^{L_1}_M(x, \theta) \right] \right) \quad .$$

(2.12)

(note that $\Sigma^{E L_1}_M$ depends on $\psi$ and $\Sigma^{B L_1}_M$ on $\xi$). To get information about the bias, we can calculate the conditional distribution of the bias given the observations and parameters, which is proportional to the joint distribution of bias and observed outputs:

$$f_{B^{L_1}_M|Y^{L_1}_M, \Theta, \Psi, \Xi}(b^{L_1} \mid y^{L_1}, \theta, \psi, \xi, x) \propto f_{Y^{L_1}_M, B^{L_1}_M|\Theta, \Psi, \Xi}(y^{L_1}, b^{L_1} \mid \theta, \psi, \xi, x)$$

$$= f_{E^{L_1}_M|\Psi}(y^{L_1}_M - y^{L_1}_M(x, \theta) - b^{L_1} \mid \psi) \cdot f_{B^{L_1}_M|\Xi}(b^{L_1} \mid \xi, x) \quad .$$

(2.13)

Again, with the simplifying assumptions (2.6) and (2.7), the conditional distribution of the bias becomes a multivariate normal distribution with mean

$$E[B^{L_1}_M \mid Y^{L_1}_M, \Theta, \Psi, \Xi] = \Sigma^{B L_1}_M (\Sigma^{E L_1}_M + \Sigma^{B L_1}_M)^{-1} \cdot (y^{L_1}_M - y^{L_1}_M(x, \theta))$$

(2.14)

and variance-covariance matrix

$$\text{Var}[B^{L_1}_M \mid Y^{L_1}_M, \Theta, \Psi, \Xi] = \Sigma^{B L_1}_M (\Sigma^{E L_1}_M + \Sigma^{B L_1}_M)^{-1} \Sigma^{E L_1}_M \quad .$$

(2.15)

See Reichert and Schuwirth, 2012 for the derivations.
2.2. Methods

Prediction

In addition to the inference of parameters, we are also interested in predictions at unknown points in time and/or space where observations are not available. By introducing the prediction layout \( L_2 \) and the joint layout \( L_{1,2} = L_1 \cup L_2 \) and using the likelihood function (2.5), we get the following expression for observed outputs at the new layout \( L_2 \) given the observations at layout \( L_1 \) and the model parameters:

\[
 f_{Y_{L_2} | Y_{L_1}^L} (y_{L_2} \mid y_{L_1}, \theta, \psi, \xi, x) = \frac{f_{Y_{L_{1,2}} | \Theta, \Psi, \Xi} (y_{L_{1,2}} \mid \theta, \psi, \xi, x)}{f_{Y_{L_1}^L | \Theta, \Psi, \Xi} (y_{L_1} \mid \theta, \psi, \xi, x)}.
\] (2.16)

The density of the predictions not conditional on model parameters is obtained by multiplying this expression with the posterior density (2.11) of the parameters for layout \( L_1 \) and integrating out the parameters:

\[
 f_{Y_{L_2} | Y_{L_1}^L} (y_{L_2} \mid y_{L_1}, x) = \int f_{Y_{L_{1,2}} | Y_{L_1}^L, \Theta, \Psi, \Xi} (y_{L_{1,2}} \mid y_{L_1}, \theta, \psi, \xi, x) \cdot f_{\Theta, \Psi, \Xi | Y_{L_1}^L} (\theta, \psi, \xi \mid y_{L_1}, x) \, d\theta \, d\psi \, d\xi.
\] (2.17)

Using the likelihood (2.8) and equation (2.16) leads to a normal distribution for predicted observations at layout \( L_2 \), given observations at layout \( L_1 \) and model parameters, with mean

\[
 E[Y_{L_2} \mid Y_{L_1}, \Theta, \Psi, \Xi] = y_{L_1}^L(x, \theta) + (\Sigma_{E_{L_{1,2}}} + \Sigma_{B_{M_{L_{1,2}}}})_{L_2,L_1} (\Sigma_{E_{L_1}} + \Sigma_{B_{M_{L_1}}})^{-1} (y_{L_1} - y_{M_1}^L(x, \theta))
\] (2.18)

and variance-covariance matrix

\[
 \text{Var}[Y_{L_2} \mid Y_{L_1}, \Theta, \Psi, \Xi] = \Sigma_{E_{L_2}} + \Sigma_{B_{M_{L_2}}} - (\Sigma_{E_{L_{1,2}}} + \Sigma_{B_{M_{L_{1,2}}}})_{L_2,L_1} (\Sigma_{E_{L_1}} + \Sigma_{B_{M_{L_1}}})^{-1} (\Sigma_{E_{L_1}} + \Sigma_{B_{M_{L_1}}})^{T} (\Sigma_{E_{L_{1,2}}} + \Sigma_{B_{M_{L_{1,2}}}})_{L_2,L_1}.
\] (2.19)

Again, we are interested in the contribution of the bias. Its conditional distribution becomes a multivariate normal distribution with mean

\[
 E[B_{M_{L_2}} \mid Y_{L_1}, \Theta, \Psi, \Xi] = \left(\Sigma_{B_{M_{L_{1,2}}}}\right)_{L_2,L_1} (\Sigma_{E_{L_1}} + \Sigma_{B_{M_{L_1}}})^{-1} (y_{L_1} - y_{M_1}^L(x, \theta))
\] (2.20)

and variance-covariance matrix

\[
 \text{Var}[B_{M_{L_2}} \mid Y_{L_1}, \Theta, \Psi, \Xi] = \Sigma_{B_{M_{L_2}}} - \left(\Sigma_{B_{M_{L_{1,2}}}}\right)_{L_2,L_1} (\Sigma_{E_{L_1}} + \Sigma_{B_{M_{L_1}}})^{-1} (\Sigma_{B_{M_{L_{1,2}}}})^{T}.
\] (2.21)
The final predictive density for \( Y_M^{L_2} \mid Y_M^{L_1} \) only conditional on given observations can be obtained by substituting the normal distributions with the mean and variance-covariance matrix from equations (2.18) and (2.19) into equation (2.17):

\[
f_{Y_M^{L_2} \mid Y_M^{L_1}} (y^{L_2}, x) = \frac{1}{\sqrt{2\pi n^{L_2}}} \int \frac{f_{\Theta, \Psi, \Xi \mid Y_M^{L_2}} (\theta, \psi, \xi \mid y^{L_1}, x)}{\sqrt{\text{det} \left( \text{Var} \left[ Y_M^{L_2} \mid Y_M^{L_1}, \Theta, \Psi, \Xi \right] \right)}} \cdot \exp \left( -\frac{1}{2} \left( y^{L_2} - \mathbf{E} \left[ Y_M^{L_2} \mid Y_M^{L_1}, \Theta, \Psi, \Xi \right] \right)^T \cdot \text{Var} \left[ Y_M^{L_2} \mid Y_M^{L_1}, \Theta, \Psi, \Xi \right]^{-1} \cdot \left( y^{L_2} - \mathbf{E} \left[ Y_M^{L_2} \mid Y_M^{L_1}, \Theta, \Psi, \Xi \right] \right) \right) \, d\theta \, d\psi \, d\xi. \tag{2.22}
\]

We may not only be interested in predicting new observations, \( Y_M^{L_2} \), but predicting our knowledge of the true values represented by \( y_M^{L_2} + B_M^{L_2} \). This leads to the same expressions as equations (2.18)–(2.22) with setting all terms \( \Sigma_{E_{L_i}} \) to zero.

**Link to multi-objective calibration**

The statistical bias description technique outlined in the preceding section was then linked to multi-objective model calibration (Yapo et al., 1998; Gupta et al., 1998; Madsen, 2000; Madsen et al., 2002; Gupta et al., 2003a; Boyle et al., 2003; Savić et al., 2011) by Reichert and Schuwirth, 2012 by realizing that the prior of the bias can be used to influence how the overall bias will be “distributed” among different model output variables.

**2.2.2 Transformation**

To account for heteroscedasticity in model output, we can apply the technique described in section 2.2.1 to Box-Cox transformed data and model results (Box and Cox, 1964; Box and Cox, 1982). Forward and backward Box-Cox transformations can be formu-
lated as
\[
g(y) = \begin{cases} 
\frac{(y + \lambda_2)^{\lambda_1} - 1}{\lambda_1} & \lambda_1 \neq 0 \\
\ln(y + \lambda_2) & \lambda_1 = 0 
\end{cases}
\]

\[
g^{-1}(z) = \begin{cases} 
(\lambda_1 z + 1)^{1/\lambda_1} - \lambda_2 & \lambda_1 \neq 0 \\
\exp(z) - \lambda_2 & \lambda_1 = 0 
\end{cases}
\] (2.23)

\[
\frac{dg}{dy} = (y + \lambda_2)^{\lambda_1-1}.
\]

In this equation, \(y\) represents a component of the vector \(y^L\) and the Box-Cox transformation parameters \(\lambda_1\) and \(\lambda_2\) can be chosen to best reduce the heteroscedasticity of \(y^L\). This leads to the following modification of the error model (2.4)

\[
g(Y^L_M(x, \theta, \psi, \xi)) = g(y^L_M(x, \theta)) + \tilde{B}^L_M(x, \xi) + \tilde{E}^L(\psi)\] (2.24)

where \(g\) indicates application of the transformation (2.23) to all components of its argument. \(\tilde{B}\) and \(\tilde{E}\) represent bias and measurement error on a transformed scale at which they are additive to the transformed model results. Through backward transformation of the results we get an equation for non-transformed model results as follows

\[
Y^L_M(x, \theta, \psi, \xi) = g^{-1}\left(g(y^L_M(x, \theta)) + \tilde{B}^L_M(x, \xi) + \tilde{E}^L(\psi)\right).
\] (2.25)

In this equation \(g^{-1}\) indicates application of the inverse Box-Cox transformation to all components of its argument. All equations described in section 2.2.1 are still valid for the transformed case by exchanging the results of the deterministic model and the observations for their Box-Cox transformed values. In addition, probability densities must be multiplied with the derivative of the Box-Cox transformation to account for the back-transformation to original units. For the likelihood function (2.8) this leads to:

\[
f_{Y^L_M|\theta, \psi, \xi}(y^L | \theta, \psi, \xi, x) = \frac{1}{\sqrt{2\pi n_L}} \frac{1}{\sqrt{\det(\Sigma_{\tilde{E}^L} + \Sigma_{\tilde{B}^L_M})}} \cdot \exp\left(-\frac{1}{2} [g(y^L) - g(Y^L_M(x, \theta))]^T (\Sigma_{\tilde{E}^L} + \Sigma_{\tilde{B}^L_M})^{-1} [g(y^L) - g(Y^L_M(x, \theta))]\right) \cdot \prod_{i=1}^{n_L} \left|\frac{dg}{dy}(y^L_i)\right|.\] (2.26)

As the derivatives of the Box-Cox transformation are evaluated at the observations, the last term is only a constant factor that is not relevant when typical numerical schemes for Bayesian inference are applied that do not rely on the normalization of the posterior.
2.2.3 Numerical implementation

Due to the analytical integration over the bias, the calibration technique described in section 2.2.1 is relatively efficient. In comparison to a calibration technique that ignores bias, it only requires to include some more parameters (of the stochastic process characterizing the bias) to be included into the calibration process and to replace the likelihood function. The bias can then be inferred separately without the need for additional simulations. The numerical implementation of this technique as described by Reichert and Schuwirth, 2012 requires an optimization of the posterior followed by a Markov chain simulation of the posterior. This is the most straightforward way of getting a sample from the posterior without further approximation. However, as such a Markov chain sample typically requires tens of thousands of simulation runs, this can still be very difficult to do for computationally demanding models. To provide a solution to this problem, it is the goal of this paper to derive an approximate version of this technique that is computationally less demanding. The underlying idea is to derive a normal approximation at the maximum of the posterior and propagate it through linearized error propagation to the results. The required techniques and the guiding equations will be discussed in this section.

Inference

To infer the parameters, \( \Gamma = (\Theta, \Psi, \Xi) \), we first maximize the posterior given by equation (2.12) for the observation layout \( L_1 \). This leads to a numerical approximation to its mode:

\[
\gamma_0^{(L_1)} \approx \arg\max_{\gamma} \left( \log \left( f_{Y_{M}, L_1}^{L_1}(y, \gamma, x) \right) \right) .
\]

(2.27)

This is a prerequisite for any of the numerical inference techniques (for the Markov chain simulation it increases the numerical efficiency by avoiding a potentially slow convergence during a burn-in phase; for the approximate technique suggested in this paper it defines the point in the environment of which a local approximation is to be found). In order to account for the generally poor knowledge of the position of the maximum and the potentially complicated shape of the posterior (Gupta et al., 2003b), a global optimization algorithm, such as the Shuffled Complex Evolution (SCE) Algorithm (Duan et al., 1993) or the Particle Swarm Optimization (PSO) algorithm (Kennedy and Eberhart, 1995; Trelea, 2003) should be used to get a reasonable approximation of \( \gamma_0 \).

In environmental modeling, the shape of the posterior may be complicated (Gupta et al., 2003b). This makes it challenging to get a good approximation by a normal distribution,
2.2. Methods

For this reason, we suggest to apply different approximation techniques and analyze their behavior for the specific application at hand.

**Finite difference approximation** The most straightforward way of getting a normal approximation at the maximum of the posterior is to derive the variance-covariance matrix from the curvature of the posterior at its maximum by

\[
\text{Var}_\Gamma \approx - \left( \frac{\partial^2 \log(f_{\text{post}}(\gamma, y_{L_1}, x))}{\partial \gamma^T \partial \gamma} \bigg|_{\gamma = \gamma^0} \right)^{-1} = - \left( \frac{\partial^2 \log(f_{Y_{M|L_1} | \Gamma}(y_{L_1}, \gamma, x) f_{\text{pri}}(\gamma))}{\partial \gamma^T \partial \gamma} \bigg|_{\gamma = \gamma^0} \right)^{-1}
\]

(Gelman et al., 1995). The approximate distribution is then given by the normal distribution with this variance-covariance matrix centered at the (approximate) maximum:

\[
N(E_\Gamma = \gamma^0, \text{Var}_\Gamma)
\]

The Hessian in equation (2.28) can be calculated by the finite difference approximation

\[
\frac{\partial^2 \log(f_{\text{post}}(\gamma, y_{L_1}, x))}{\partial \gamma_i \partial \gamma_j} \bigg|_{\gamma = \gamma^0} \approx \frac{\log(f_{Y_{M|L_1} | \Gamma}(y_{L_1}, \gamma^0 + \delta_i e_i + \delta_j e_j, x) f_{\text{pri}}(\gamma^0 + \delta_i e_i + \delta_j e_j)) - \log(f_{Y_{M|L_1} | \Gamma}(y_{L_1}, \gamma^0 - \delta_i e_i - \delta_j e_j, x) f_{\text{pri}}(\gamma^0 + \delta_i e_i - \delta_j e_j))}{4 \delta_i \delta_j}
\]

where \(e_i\) is the unit vector in \(i\)-direction, \(e_i = (0, ..., 0, 1, 0, ..., 0)^T\), and \(\delta_i\) is a “small” increment adapted to the scale of \(\gamma_i\) (Gelman et al., 1995).

Note that the calculation of the derivatives in the equations (2.28) or (2.29) does not depend on an additive constant to the log of the posterior density and thus does not require normalization.

If there are \(n\) parameters to be estimated, this method requires \(2n^2\) simulations to be done in addition to the base simulation at the maximum, \(\gamma^0\), that was already done by the maximization algorithm. It further requires the numerical approximation to the maximum to be accurate enough that the \(2n^2\) simulations all lead to smaller posterior values than the base simulation. Given the potentially difficult shape of the posterior (Gupta et al., 2003b) this may already be challenging. In addition, such a local approximation may underestimate the variances of the parameters as a local maximum may have a substantially higher curvature than would have an approximate distribution that
smoothes the posterior density regionally. Nevertheless, this technique is one of the candidate techniques to be tested for the derivation of the coefficients of a local normal approximation to the posterior.

**Density fitting** To avoid the potential problems of a local approximation mentioned above, it would be an alternative to draw a sample of parameter values in the neighborhood of the maximum, \( \gamma^0 \), and to fit the log of a multivariate normal density to the corresponding (non-normalized) log posterior values of the model. We performed this fit by minimizing the sum of squared deviations between the log posterior values calculated for the sample and the log posterior values approximated by the multivariate normal density at the same points. We used the Nelder-Mead optimization algorithm to execute this task. If the parameter vector of the model, \( \Gamma \), is of length \( n \), this requires us to estimate \( n(n + 3)/2 + 1 \) parameters of the approximate posterior (\( n \) means, \( n \) standard deviations, \( n(n - 1)/2 \) correlation coefficients, and 1 normalization coefficient) from the log posterior values of the sample. If \( n \) is not small, estimating such a large number of distribution parameters may be a challenging problem. This problem becomes even more challenging having the constraints in mind: standard deviations must be positive and the correlation coefficients must be between minus unity and unity and lead to a positive definite correlation matrix. On the other hand, as this technique avoids the problems of the local approximation mentioned above, it may be worth testing. To decrease the number of distribution parameters to be estimated, an alternative might be to keep the mean fixed at the maximum of the sample and recalculate the normalization coefficient in each fitting iteration. This decreases the number of distribution parameters to be estimated to \( n(n + 1)/2 \).

**Importance sampling** Given the potential difficulties of a local normal approximation and those of a regional fit of a normal density mentioned in the previous paragraphs, an alternative concept of getting a normal approximation may be needed. Importance sampling may be such an alternative. Importance sampling (Geweke, 1989; Gelman et al., 1995) is a technique to calculate weights to a sample of a sampling distribution that make it a weighted sample of another distribution, in our case the posterior. In principle, the sampling distribution is arbitrary (except that the density should not be zero within the range of the posterior), but the technique is only efficient if the sampling distribution is a reasonable approximation to the posterior. The weighted sample of the posterior can then be the basis for calculating the posterior mean, \( E_\Gamma \), and variance-covariance matrix, \( \text{Var}_\Gamma \), that can then be used as the parameters of a normal distribu-
tion that approximates the posterior. This leads again to a regional approximation, but the advantage is that the derivation of the mean and variance-covariance matrix does not depend on the assumption of normality. This makes the technique more robust than the previously discussed techniques. Normality is still assumed in the next step of using a normal distribution with this mean and this variance-covariance matrix as a basis for the subsequent step of (linearized) error propagation.

Technically, a sample, \( \{ \gamma^{(i)} \}_{i=1}^{N} \), with sample size \( N \), is drawn from a distribution with density \( f_{\text{samp}} \) centered at \( \gamma^0 \) and then weights are calculated according to

\[
 w_{\text{post},i} = \frac{f_{Y^L_{\mathcal{M}} | \Gamma}(y^{L_1}, \gamma^{(i)}, x) f_{\text{pri}}(\gamma^{(i)})}{\sum_{j=1}^{N} f_{Y^L_{\mathcal{M}} | \Gamma}(y^{L_1}, \gamma^{(j)}, x) f_{\text{pri}}(\gamma^{(j)}) f_{\text{samp}}(\gamma^{(j)})}. \tag{2.30}
\]

Mean and variance-covariance matrix are then calculated using the sample \( \{ \gamma^{(i)} \}_{i=1}^{N} \) with weights \( \{ w_{\text{post},i} \}_{i=1}^{N} \) and the normal distribution with this mean and this variance-covariance matrix is used to approximate the posterior. The procedure may be iterated by sampling from this normal distribution and derive new weights to get new estimates of mean and variance-covariance matrix that are based on this sampling distribution that should be a better approximation to the posterior (Reichert et al., 2002).

We suggest two strategies of finding a reasonable sampling distribution. The first consists of guessing reasonable standard deviations of the parameters (e.g. a constant fraction of the parameter values) and choosing a normal sampling distribution centered at \( \gamma^0 \) with these standard deviations and without correlations as a first sampling distribution. The sampling distribution could then be replaced iteratively by the normal distribution with mean and variance-covariance matrix derived from the weighted sample. If it seems too difficult to guess standard deviations for the first sampling distribution, a uniform distribution in a ball around the maximum could be chosen as a first sampling distribution (see Chen, 2011 for how to sample from a uniform distribution in a ball). The radius of the ball should not be too small to avoid severe bias by the property of the density of this distribution to be zero outside the ball (see comment above about importance sampling). The estimates of mean and variance-covariance matrix could be improved subsequently by proceeding as above with a normal sampling distribution with parameters derived from the previous step. This second procedure can be expected to be less efficient regarding the effective sample size of the first step (based on the uniform sample in the ball), but more robust regarding the potential of identifying a higher posterior density maximum than the one calculated by the optimization algorithm.
Prediction

The different approximation techniques described above result in the parameters $E_\Gamma$ and $\text{Var}_\Gamma$ of a normal distribution approximating the posterior distribution. These parameters can be used for a propagation of this posterior through the model for an uncertainty analysis of the model for a prediction time. Given a conditional probability density of results given parameters, $f_{Y|\Gamma}$, and a probability density of the parameters, $f_\Gamma$, we can calculate the unconditional predictions of the results by marginalization of the joint density as

$$f_Y(y) = \int f_{Y|\Gamma}(y | \gamma) \cdot f_\Gamma(\gamma) \, d\gamma. \quad (2.31)$$

If both distributions, $f_{Y|\Gamma}$, and $f_\Gamma$ are normal and we can neglect the dependence of $\text{det}[\text{Var}_{Y|\Gamma}]$ on the parameter vector $\gamma = (\theta, \psi, \xi)$ in the normalization constant, we get the approximate proportionality

$$f_Y(y) \propto \int \exp \left(-\frac{1}{2} [y - E_{Y|\Gamma}(\gamma)]^T \text{Var}_{Y|\Gamma}(\gamma)^{-1} [y - E_{Y|\Gamma}(\gamma)] - \frac{1}{2} (\gamma - E_\Gamma)^T \text{Var}_\Gamma^{-1} (\gamma - E_\Gamma) \right) \, d\gamma. \quad (2.32)$$

If we further approximate this expression by neglecting the dependence of $\text{Var}_{Y|\Gamma}$ on $\gamma$ and approximate the dependence of $E_{Y|\Gamma}$ on $\gamma$ by the constant and linear terms of the Taylor series, we obtain the approximation

$$f_Y(y) \propto \int \exp \left(-\frac{1}{2} \left[ y - E_{Y|\Gamma}(E_\Gamma) - \frac{\partial E_{Y|\Gamma}}{\partial \gamma^T}(E_\Gamma)(\gamma - E_\Gamma) \right]^T \text{Var}_{Y|\Gamma}(E_\Gamma)^{-1} \cdot \left[ y - E_{Y|\Gamma}(E_\Gamma) - \frac{\partial E_{Y|\Gamma}}{\partial \gamma^T}(E_\Gamma)(\gamma - E_\Gamma) \right] - \frac{1}{2} (\gamma - E_\Gamma)^T \text{Var}_\Gamma^{-1} (\gamma - E_\Gamma) \right) \, d\gamma. \quad (2.33)$$

When completing the square to get a normal density in $\gamma$, integrating over $\gamma$ and completing the square a second time to get a normal density in $y$, we finally get a normal distribution for $Y$ with mean $E_{Y|\Gamma}(E_\Gamma)$ and variance:

$$\text{Var}_Y = \left[ \text{Var}_{Y|\Gamma}(E_\Gamma)^{-1} - \text{Var}_{Y|\Gamma}(E_\Gamma)^{-1} \cdot \frac{\partial E_{Y|\Gamma}}{\partial \gamma^T}(E_\Gamma) \cdot \left( \frac{\partial E_{Y|\Gamma}}{\partial \gamma^T}(E_\Gamma) \right)^T \cdot \text{Var}_{Y|\Gamma}(E_\Gamma)^{-1} \right]^{-1} \quad (2.34)$$
This can further be simplified to

\[ E_Y \approx E_{Y|\Gamma}(E) \]  

and

\[ \text{Var}_Y \approx \text{Var}_{Y|\Gamma}(E) + \frac{\partial E_{Y|\Gamma}}{\partial \gamma^T}(E) \cdot \text{Var}_{\Gamma} \cdot \left( \frac{\partial E_{Y|\Gamma}}{\partial \gamma^T}(E) \right)^T = \text{Var}_{Y|\Gamma}(E) + V \cdot \text{Var}_{\Gamma} \cdot V^T \]  

with the Jacobian

\[ V = \frac{\partial E_{Y|\Gamma}}{\partial \gamma^T}(E) \]

This approximation will now be used to get approximate predictions for new observations, \( Y_{L_i} \), and for our knowledge of the true state, \( y_{L_i} \), for both layouts, \( L_1 \) and \( L_2 \), based on the parameters of the conditional normal distributions discussed in section 2.2.1 and the normal approximation of the posterior discussed in this section. In addition, we are interested in the predictions for the bias, \( B_{L_i} \), and for the posterior of the observation error for layout \( L_1 \), \( E_{L_1} \), to get a probabilistic separation of the residuals into bias and observation error. The Jacobians for these four cases are given by

\[ V_{Y_{L_1,2}} = \frac{\partial E[Y_{M,L_1,2} | Y_{M,L_1,\Gamma}]}{\partial \gamma^T} \bigg|_{\gamma=\gamma^0} \]

using the expected value as presented in equation (2.18),

\[ V_{Y_{M,L_1,2} + B_{M,L_1,2}} = \frac{\partial E[Y_{M,L_1,2} + B_{M,L_1,2} | Y_{M,L_1,\Gamma}]}{\partial \gamma^T} \bigg|_{\gamma=\gamma^0} = \frac{\partial (y_{M,L_1,2} + E[B_{M,L_1,2} | Y_{M,L_1,\Gamma}])}{\partial \gamma^T} \bigg|_{\gamma=\gamma^0} \]

using (2.14) and (2.20),

\[ V_{B_{M,L_1,2}} = \frac{\partial E[B_{M,L_1,2} | Y_{M,L_1,\Gamma}]}{\partial \gamma^T} \bigg|_{\gamma=\gamma^0} \]

using again the equations (2.14) and (2.20), and finally

\[ V_{E_{L_1}} = \frac{\partial E[E_{L_1} | Y_{M,L_1,\Gamma}]}{\partial \gamma^T} \bigg|_{\gamma=\gamma^0} = -\frac{\partial E[Y_{M,L_1} + B_{M,L_1} | Y_{M,L_1,\Gamma}]}{\partial \gamma^T} \bigg|_{\gamma=\gamma^0} \]

which results from equation (2.4). To apply equations (2.35) and (2.36) for calculating the approximate predictions for the four cases \( Y_{L_i}, y_{L_i} + B_{L_i}, B_{L_i}, \) and \( E_{L_1} \), we need the expected values referenced above, the Jacobians given by the equations (2.38) to
Chapter 2. Calibration of computationally demanding models

(2.41), and the variances (2.19), (2.19) with ignoring the error terms \( \Sigma_{E_{L_i}} \), (2.15), and again (2.15), respectively.

To evaluate the expressions (2.38)–(2.41), we need the derivatives of the different expected values. In all four cases, these expected values can be written in the form

\[
E[Y | Y^L_M \Gamma] (\gamma) = a(\gamma) + A(\gamma) \cdot y^L_M(\gamma)
\]  \hspace{1cm} (2.42)

where \( a \) represents a vector and \( A \) a matrix chosen in a way that the three equations (2.14), (2.18) and (2.20) match equation (2.42). The derivatives can be calculated according to

\[
\frac{\partial E[Y | Y^L_M \Gamma]}{\partial \gamma_j} \bigg|_{\gamma = \gamma^0} = \frac{\partial a}{\partial \gamma_j}(\gamma^0) + \frac{\partial A}{\partial \gamma_j}(\gamma^0) \cdot y^L_M(\gamma^0) + A(\gamma^0) \cdot \frac{\partial y^L_M}{\partial \gamma_j}(\gamma^0) . \]  \hspace{1cm} (2.43)

As the derivatives of \( a \) and of \( A \) are computationally much less demanding (no simulations of the deterministic model required) than that of \( y^L_M \), this equation enables to save computation time by using the same Jacobian of \( y^L_M \) for all four cases mentioned above. In addition it allows us to restrict the calculations of the finite difference approximation of this Jacobian to the components \( \theta \) of the parameter vector \( \gamma = (\theta, \psi, \xi) \) as the results of the deterministic model do not depend on the parameters \( \psi \) and \( \xi \). When using the Box-Cox transformation (see section 2.2.2), these equations must be applied to the transformed model results. This requires calculating the Jacobian of transformed results from that of not transformed results

\[
\frac{\partial g(y^L_M)}{\partial \gamma^T} = \frac{\partial y^L_M}{\partial \gamma^T} \cdot \text{diag} \left( \frac{dg}{dy} \left( y^L_M(\gamma^0) \right) \right) , \]  \hspace{1cm} (2.44)

where again \( g \) and \( dg/dy \) return the vector of results of the application of the functions (2.23) to all components.

As an example of the complete predictive distribution, we give the distribution of our knowledge of the real state of the system:

\[
y^L_{M1,2} + B^L_{M1,2} | Y^L_M \sim N \left( Y^L_{M1,2} (E_{\Gamma}) + E[B^L_{M1,2} | Y^L_M \Gamma] (y^L_{M1,2}, E_{\Gamma}) , \right.
\]

\[
V_{y^L_{M1,2} + B^L_{M1,2}} \cdot \text{Var}_{\Gamma} \cdot \left( V_{y^L_{M1,2} + B^L_{M1,2}} \right)^T + \text{Var}[B^L_{M1,2} | Y^L_M, \Gamma] (y^L_{M1,2}, E_{\Gamma}) \right) \]  \hspace{1cm} (2.45)

where \( E_{\Gamma} \) and \( \text{Var}_{\Gamma} \) are the parameters of the normal distribution approximating the posterior derived according to one of the techniques described above. The mean of the bias is calculated according to equations (2.14) and (2.20), its variance according to equations (2.15) and (2.21). \( V_{y^L_{M1,2} + B^L_{M1,2}} \) follows from equation (2.39) using equations (2.42) to (2.44).
2.2. Methods

2.2.4 Summary of the approximate calibration and prediction procedure

The suggested approximate multi-objective calibration technique including uncertainty analysis of model predictions for a deterministic model \( y_L^M(x, \theta) \), based on the concepts presented in sections 2.2.1 to 2.2.3, can be summarized as follows:

1. Specify a prior for the model parameters \( \Theta \) based on present knowledge about (possible) parameter values.

2. Specify a potentially informative prior of the parameters \( \Psi \) of the variance-covariance matrix of the observation process by using model-independent knowledge (equation (2.6)). In the presented case, this step consists of specifying priors for the standard deviations of the observation error of all output variables. In case of heteroscedastic data, a Box-Cox transformation (section 2.2.2) can be performed. The priors should hence represent knowledge about errors for typical concentrations on a transformed scale.

3. Specify a prior for the parameters \( \Xi \) of the Gaussian process characterizing prior knowledge of the bias for an observation layout \( L_1 \) (equations (2.7) and (2.10)). In the presented case, this step consists of choosing a prior for the smoothing parameters \( \beta \) of the correlation structure and for the standard deviations of the bias in different output variables (equation (2.10)). The prior of the standard deviations should reflect the degree of bias we are willing to accept and the desire to avoid bias (for example with an exponential distribution with which probability density increases with decreasing value of the standard deviation), hence our calibration objectives, and the expectations about the bias contribution. When applying a Box-Cox transformation, the specification, again, must be done at the transformed scale.

4. Derive the joint posterior distribution of the parameters \( \Theta, \Psi \) and \( \Xi \) and the model bias \( B \) by an approximate linearized approach described in section (2.2.3) using observations \( y^{L_1} \) of the observation layout \( L_1 \) (equations (2.12), (2.14) and (2.15)). This gives the uncertainty estimation of the model results for times where data provides information about the amount of bias in model results.

5. Quantify the posterior knowledge about the “true” system state, \( y^L_M + B^L_M \mid Y^L_M \), and of new observations for the prediction layout \( L_2 \), according to equation (2.45) without and with considering the observation error and with \( E[B^L_M \mid Y^L_M, \Theta, \Psi, \Xi] \).
Chapter 2. Calibration of computationally demanding models

and $\text{Var}[B_M^{L_2} | Y_M^{L_1}, \Theta, \Psi, \Xi]$ according to (2.20) and (2.21) and $f_{\Theta, \Psi, \Xi | Y_M^{L_1}}(\theta, \psi, \xi | y^{L_1}, x)$ according to (2.12). This gives the uncertainty estimates of the model results at times where data is not available or not used for calibration. If observations are available for layout $L_2$, they can be used for a validation of the predictive power of the model. After conducting the inference and prediction on a Box-Cox transformed scale, for comparison of real data and model output results, the results have to be transformed backward.

2.3 Didactical example

We use the same model of growth of microbes in a continuously operated, stirred tank reactor as Reichert and Schuwirth, 2012, as a simple example for the application of the suggested approach.

2.3.1 Data

In this model, the microbes’ concentration, $C_M$, and the substrate concentration, $C_S$, are described by the following mass balance equations:

$$\frac{dC_M}{dt} = \mu \frac{C_S}{K + C_S} C_M \exp(c(T - T_0)) - b C_M - q C_M$$

$$\frac{dC_S}{dt} = -\frac{\mu Y}{K + C_S} C_M \exp(c(T - T_0)) + q (C_{S,\text{in}} - C_S)$$

(2.46)

with the initial conditions

$$C_M(0) = C_{M,\text{ini}}$$

$$C_S(0) = C_{S,\text{ini}}$$

(2.47)

and the temperature variations

$$T = T_0 + A \sin(2\pi(t - 0.25d))$$

(2.48)

In these equations, $C_M$ is the concentration of microbes in the reactor (ML$^{-3}$), $C_S$ is the concentration of substrate in the reactor (ML$^{-3}$), $t$ is time (T), $\mu$ is the maximum specific growth rate of the microbes (T$^{-1}$), $K$ is the concentration of substrate at which the microbes grow with half of their maximum specific growth rate (ML$^{-3}$), $b$ is the specific death rate of the microbes (T$^{-1}$), $q$ is the volumetric flow rate through the reactor per unit of reactor volume (T$^{-1}$), $Y$ is the yield (produced amount of biomass per consumed...
2.3. Didactical example

amount of substrate) (the concentration of substrate in the inflow \((\text{ML}^{-3})\) (the concentration of microbes in the inflow is assumed to be zero), \(T\) is temperature, \(T_0\) is a reference temperature assumed to be \(20^\circ\text{C}\), \(A\) is the amplitude of temperature variations, \(c\) is a temperature dependence coefficient assumed to be \(0.046\) per degree Celsius, \(C_{M,\text{ini}}\) is the initial concentration of microbes in the reactor \((\text{ML}^{-3})\), and \(C_{S,\text{ini}}\) is the initial concentration of substrate in the reactor \((\text{ML}^{-3})\).

We produced synthetic observations with this model for a simulation period of 4 days, with a temperature amplitude of \(4^\circ\text{C}\) and the following parameter values: \(\mu = 4\), \(K = 10\), \(b = 1\), \(q = 1\), \(Y = 0.6\), \(C_{S,\text{ini}} = 100\), \(C_{M,\text{ini}} = 10\), \(C_{S,\text{ini}} = 40\), and \(\sigma_{E_{CM}} = 0.5\) and \(\sigma_{E_{CS}} = 0.5\) (in model output units) as standard deviations of normally distributed observation errors in both model output variables. The generated data does not contain heteroscedasticity, so no Box-Cox transformation was needed.

2.3.2 Model description

From the first two days out of these data, we will try to estimate the parameters of a model that describes the same system but is biased because of omitting the temperature dependence of substrate and microbial growth described in equations \((2.46)\). This represents a typical problem in environmental modeling resulting from choosing a wrong (i.e. simplified) model structure, for example by choosing a wrong set of model equations. With this example we can discuss parameter estimates and model predictions for the full period of four days under the presence of model bias.

Here, the observation layout as introduced in \((2.2)\) specifies observations of the two variables, \(C_M\) and \(C_S\), at the time points \(\{t_1, t_2, \ldots, t_n\}\) and given inputs \(A\) and parameter \(c\). The deterministic model function resulting from the solution of the system of differential equations \((2.46)\) can be written as

\[
y^L_M(\theta) = (C_M(t_1, \theta), \ldots, C_M(t_n, \theta), C_S(t_1, \theta), \ldots, C_S(t_n, \theta))^T. \tag{2.49}
\]

The vector \(\theta = (\mu, K, b, q, Y, C_{S,\text{ini}}, C_{M,\text{ini}}, C_{S,\text{ini}})^T\) gives the parameter vector of that model. The parameters \(\psi = (\sigma_{E_{CM}}, \sigma_{E_{CS}})^T\) consist of the standard deviations \(\sigma_{E_{CM}}\) and \(\sigma_{E_{CS}}\) of independent normal distributions characterizing the observation error of \(C_M\) and \(C_S\) at all points in time. Finally, the vector \(\xi = (\sigma_{B_{CM}}, \sigma_{B_{CS}}, \tau)^T\) combines the standard deviations \(\sigma_{B_{CM}}\) and \(\sigma_{B_{CS}}\) for the bias in each of the model variables \(C_M\) and \(C_S\) and the correlation time \(\tau\) instead of the parameter \(\beta\) (according to equation \((2.7)\)). The correlation times are kept equal for the two variables in this example but potentially could be different.
2.3.3 Prior distribution

As the microbial growth and substrate consumption are coupled by the yield, $Y$ (see the first terms on the right hand sides of the equations (2.46)) and measured time courses of $C_M$ and $C_S$ might not perfectly match this constraint, the model cannot fit both components equally well. Hence, the quality of fit of each component will depend on the choice of the prior of the bias. Through the choice of the prior of the bias, we can influence how the bias will be “distributed” among the two model variables while an increasing probability density for smaller values of the priors of the standard deviations of the bias should represent the desire to have as little bias as possible. For the priors of $\sigma_{B_{CM}}$ and $\sigma_{B_{CS}}$, we therefore chose normal distributions with mean zero and a standard deviation of 0.5. They were truncated at zero to avoid negative values. For the parameters $\theta$, we used independent lognormal priors with the means at the correct values given above and standard deviations of 50%. For the parameters $\psi$ and the correlation time $\tau$, lognormal priors of the means at the correct values in case of the observation error parameters (given above) and at 0.3 for $\tau$, which does not have a true value, and standard deviations of 10% were used. We calibrated the logarithm of all parameters instead of the parameters themselves, as this appeared to lead to a better approximation of the posterior by a normal distribution. As model parameters are typically positive, this is usually possible. Hence, the posterior marginals of the parameters result in lognormal distributions.

Additionally to this base scenario, we repeated the application of the method for calibration and uncertainty analysis as described in 2.2 for a second scenario choosing wrong priors, which frequently occurs in the context of environmental modeling. We simulated the choice of wrong priors for the model parameters $\theta$ by setting the mean 0.5 times larger than the correct parameter value. For the results we refer to A.1.

2.3.4 Results and discussion

In the following we will show the results of the approximate uncertainty analysis as described in 2.2 for the calibration and the prediction time for the basic prior scenario. Here, we focus on a discussion of the different posterior approximation techniques described in section 2.2.3. For a discussion of different prior choices for the standard deviations of the bias in $C_M$ and $C_S$ and the effect of ignorance of bias, we refer to Reichert and Schuwirth, 2012. The parameter inference procedure described in section 2.2.3 was performed for this case study based on maximization of the posterior followed by all three approximation techniques described in section 2.2.3. A comparison of the com-
2.3. Didactical example

The computational burden of each of those techniques compared to the full MCMC approach is given in Table 2.1. Depending on the sample size and, for the finite difference and the importance sampling approximation, depending on the number of iterations, it shows a 6-70 times smaller computational demand for the approximation techniques compared to the full MCMC approach. To get a local estimate of the Hessian at the maximum of the posterior, the finite difference technique required us to perform 338 (=2·13\(^2\)) simulations in addition to the base simulation at the maximum available from the optimization procedure. Due to the approximate estimation of the position of the maximum, several attempts had to be made to get smaller values of the posterior at these 338 locations than at the numerical “maximum” found so far (by replacing the numerical “maximum” by the position with the higher value and restarting the finite difference approach). This is an indication of a complex posterior shape and/or of a poor performance of the optimization algorithm. Also, the choice of the increment for evaluation of the Hessian is critical, for more information we refer to the supplementary data. The density fitting approach was based on a random sample of 1000 parameter vectors from a uniform distribution in a ball around the numerical approximation to the maximum. Fitting a multivariate normal distribution (with 105 parameters) to these posterior values failed, as the standard deviations of observation error and/or bias tended to diverge. Finally, importance sampling was also done, based on a sample of size 1000, first from a uniform distribution in a ball around the numerical maximum, then improved by iterative sampling from normal distributions based on mean and variance-covariance matrix from the previous iteration step until the effective sample size did no longer increase.
Table 2.1: Number of simulation runs needed for the different approximation techniques and the full Markov chain Monte Carlo method as presented in Reichert and Schuwirth, 2012 for the didactical example. All numbers refer to the simulation runs actually performed for the results shown herein. For further improvement some methods could be conducted with larger sample sizes or several iterations of the same technique.

<table>
<thead>
<tr>
<th>Application Case</th>
<th>Optimization</th>
<th>Finite Difference</th>
<th>Density Fitting</th>
<th>Importance Sampling</th>
<th>MCMC</th>
</tr>
</thead>
</table>

1 we performed 5 iterations of the same optimization technique (algorithm Nelder-Mead) each with 500 as the maximum number of iterations
2 we performed several iterations (maximal 20) until no new maximum was found
3 1 importance sampling from a uniform ball followed by 10 iterations of importance sampling from a normal distribution, sample size 1000
4 the final chain has a sample size of 50000, the first 4 iterations with smaller chains were done for an optimization of the jump distribution

Fig. 2.1 shows the prior marginals for the base scenario of correct priors and the corresponding results for the posterior marginals of the parameter distribution for the finite difference and the importance sampling approximation in comparison with the analogous results of the Markov chain Monte Carlo technique (Reichert and Schuwirth, 2012), which samples the full posterior distribution.

As we apply the approximate technique described in section 2.2.3 to the log of the parameters, all posterior marginals are lognormal distributions under our simplified assumptions. The figure shows a small effect of the choice of the posterior approximation technique on the posterior marginal distribution. For most parameters the results of the finite difference method show slightly wider posterior probability densities compared to the importance sampling results. Potentially, the finite difference method can be problematic as it tends to very local approximations that might underestimate the width of the posterior distribution. This does not seem to be the problem for our didactical example. Only small differences can be found for the standard deviations of the observation errors, the correlation time and the initial concentrations. The results also correspond well with the Markov chain results, although the posteriors of the two approximation methods, especially of the importance sampling technique, tend to be narrower.

Fig. 2.2 shows the model predictions (according to equation (2.22)) of our knowledge of
Figure 2.1: Prior (dotted, supported by light grey shading of the areas below the lines) and posterior (supported by dark grey shading of the areas below the lines) marginals of the model parameters for the posterior approximation by finite difference method (long-dashed), importance sampling (short-dashed) and a full Markov chain Monte Carlo method (solid) (Reichert and Schuwirth, 2012). Vertical lines indicate the true values of the model parameters used for producing the synthetic data (there are no true values for the parameters of the bias).
the ‘true’ results, $Y_M^{L_{1,2}} + B_M^{L_1} | Y_M^{L_1}$, and of observations, $Y_M^{L_{1,2}} | Y_M^{L_1}$, for both posterior approximation techniques shown in the different columns, again in comparison with the MCMC results (Reichert and Schuwirth, 2012), for the basic prior choice.

Figure 2.2: Results of the median and 95% credibility intervals for both variables $C_M$ and $C_S$ (top row; observations used for calibration are marked by solid symbols, those not used for calibration by open symbols; output of the deterministic model (long-dashed line), bias-corrected output (deterministic model plus bias; median as solid line, 95% credibility bounds as dark grey area bounded by dashed lines) and prediction for new observations (including observation error; median as solid line (same as for bias-corrected output), 95% credibility bounds shown as dark and light grey areas bounded by dotted lines)), results of the median and 95% credibility intervals for the bias (solid line and dark grey areas bounded by dashed lines) and of the observation error (dots and vertical line segments) for the variables $C_M$ (middle row) and $C_S$ (bottom row). The left column shows the results for posterior approximation by finite differences, the middle column by importance sampling and the right column the results of the full MCMC method (Reichert and Schuwirth, 2012).
The credibility intervals for the ‘true’ results, $y_{L1,2}^{M} + B_{L1,2}^{M} | Y_{L1}^{M}$, can contain a small or large fraction of future observations, depending on the available knowledge and the observation error. If the parameters would be accurately known, the credibility intervals of observations, $Y_{L1,2}^{M} | Y_{L1}^{M}$, should contain about 95% of future observations. However, due to the uncertainty in parameter values, they can also be considerably wider. Overall, the results demonstrate that our posterior knowledge is much more precise in times where observations are available (and used for inference) than in the extrapolation range. Further results of the posterior estimates of bias and observation error show the probabilistic division of the residuals into bias and observation error. The markers in the plots of the middle and bottom row of Fig. 2.2 indicate approximately normally distributed independent observation errors as required. In comparison to the results of a Markov chain Monte Carlo technique to sample from the posterior shown in the right column of Fig. 2.2, the results of the linearized error propagation technique generally show very similar uncertainty bands, slightly narrower in the case of importance sampling. Furthermore, we see some but small effects of the choice of the posterior approximation technique on the uncertainty estimates.

Figures A.1 and A.2 of the supplementary data show the results of the introduced method for the choice of wrong priors, again in comparison with a Markov chain Monte Carlo approach applied to the same prior choice. Additionally to the density fit, also the finite difference technique failed in this case. Fig. A.1 shows the prior and posterior marginal parameter distributions calculated by the importance sampling approximation and the MCMC method. Fig. A.2 shows the results of the model simulation, the uncertainty estimation and the different contribution of bias and observation error to the uncertainty, calculated by the same methods. We see a negative effect due to the wrong prior choice, but only minor effects of the choice of the calibration and prediction method.

2.4 Application to biogeochemical and ecological lake model

In this section we give a short description of the lake model BELAMO followed by an application of the approximate calibration and prediction technique (described in the sections 2.2.3 and 2.2.4) for calibrating this model to long-term data from Lake Zurich, Switzerland.
2.4.1 BELAMO: model description

The Biogeochemical and Ecological LAke MOdel (BELAMO) aims at a combined calculation of mass balances of nutrients, oxygen, organic particles, phytoplankton and zooplankton. The earliest version of BELAMO is a one-dimensional mechanistic model of nutrient and plankton dynamics developed for Lake Zurich. It models horizontally averaged substance and organism concentrations in the water column and two sediment layers as they are changing with time and the depth of the lake. For a short description of the model see Omlin et al., 2001b; Mieleitner and Reichert, 2006 slightly changed the one-dimensional model and analyzed its transferability to the lakes Greifensee and Walensee.

To reduce simulation time and analyze functional phytoplankton groups, the continuous vertical resolution of the model was simplified to a 4-box (epilimnion, hypolimnion and two sediment boxes) approach by Mieleitner and Reichert, 2008. In these mixed boxes, concentrations of ammonium, nitrate, phosphate, oxygen, degradable and inert dead organic particles and (in the aggregated version used for this study) one group of phytoplankton and one group of zooplankton are modeled.

In the latest version of the model, the following physical processes are considered: inflow into epi- and hypolimnion and outflow from the epilimnion, gas exchange with the atmosphere, mixing of all dissolved substances between epi- and hypolimnion, upwards movement of zooplankton from zones of low oxygen concentration, sedimentation of particles, advection, diffusion of dissolved substances between the pore water within the two sediment layers and the water in the hypolimnion, as well as accumulation of sediment and permanent burial.

The model considers the following biogeochemical and ecological processes: Growth, respiration and death of phyto- and zooplankton, aerobic, anoxic and anaerobic mineralization, nitrification, methane oxidation and phosphate uptake of sinking degradable and inert organic particles.

BELAMO was implemented in AQUASIM (version 2.1f, http://www.aquasim.eawag.ch), a computer program for the identification and simulation of aquatic systems (Reichert, 1994; Reichert, 1998). For a detailed description of the newest version of the model see Dietzel et al., 2012.

As BELAMO describes water quality and plankton dynamics, it can contribute to improving the understanding of the behavior of a lake by linking continuous-time nutrient and plankton fluxes and dynamics to (sparse) observations of concentrations only. An even more ambitious aim would be that the model supports decisions on management alter-
natives for environmental protection. To gain confidence in such a model application, a transparent evaluation of model performance is needed that takes data uncertainty, as well as model structure uncertainty, into account.

Despite the simple spatial structure of the lake model we are already faced with a complex model with 16 state variables, up to 19 processes in each box and around 70 parameters, of which at least 10 strongly influence the model results. There is not much prior knowledge available for several of these parameters. Examples are the mineralization rates of aerobic, anoxic and anaerobic mineralization processes that represent a combination of bacterial densities and their growth rates. Only indirect information is available through their influence on observed concentrations of oxygen, nitrate and organic material in the water column of the lake.

On the one hand, the complexity of the model results in simulation times of at least 1 min per year. This makes usual Bayesian numerical techniques, such as Markov chain Monte Carlo simulations, difficult to apply for long-term simulations. On the other hand, the coarse description of plankton communities with an aggregation into one group of phyto- and one group of zooplankton, as well as the spatial aggregation of the whole lake depth into two boxes and process simplifications, cause significant biases. This is particularly the case as this model closes mass balances by explicitly considering mineralization in the sediment. This couples the model variables much stronger than it is the case in more widespread lake models that use sediment fluxes as independent parameters (Mooij et al., 2010). Without making the model even more complex, this bias cannot be significantly reduced by extending the model structure. Furthermore, as we have model outputs and data for evaluation of both nutrients and plankton, we are confronted with a multi-objective calibration problem. Manual calibrations showed the difficulty of calibrating the model evenly well for all output variables during the whole simulation time. Because of these problems, an uncertainty estimation, as the one described above, became necessary. As a first calibration step, the model was applied to long-term observations of Lake Zurich. As BELAMO should describe the processes in lakes as universally as possible, we tend to a joint calibration of the model parameters for lakes of different trophic states but similar climate and mixing conditions. In Dietzel et al., 2012 we show the results of a joint calibration of three Swiss lakes with different trophic states with the suggested calibration and prediction technique.
2.4.2 Study area

Lake Zurich is located in the north-eastern part of the Swiss plateau. At its deepest point the lake is 136 m deep. We apply the model to the lower part of Lake Zurich only, which has a length of around 28 km. The catchment of Lake Zurich is mainly influenced by urban areas and less by agriculture. As measured by prevailing phosphorus concentrations, Lake Zurich is a mesotrophic lake. For a detailed description of the main lake attributes see Mieleitner and Reichert, 2006, for a location map see Dietzel et al., 2012.

2.4.3 Data

Monthly measured profiles of physical, chemical and biological variables for Lake Zurich were obtained from 1976 to 2005 (zooplankton only since 1985). These data were collected by the Water Supply Authority of Zurich (Wasserversorgung Zürich, WVZ). Information on inflowing rivers and contributing waste water treatment plants (physical and chemical parameters) and meteorological data were received from federal and cantonal agencies and from technical reports.

Data from input, concentrations in the lake and meteorological forcing were converted to AQUASIM format. Data compilation was done with S-Plus. Accounting for the availability of input and validation data we accomplished model calibrations with the data from 1976 up to and including 1995. Hence, our observation layout $L_1$ indicates combinations of monthly sampling dates, the five calibrated output variables phytoplankton, zooplankton, nitrate, oxygen and phosphate and the two compartments epilimnion and hypolimnion. Simulations and data of the following 10 years up to the end of 2005 were used for model validation. The monthly sampling dates of these years in combination with each of the variables and compartments compose the prediction layout $L_2$. For a detailed description of the available data and its compilation see Dietzel et al., 2012.

We assume that observation errors as well as bias are smaller for smaller concentrations of chemical and biological variables. To account for this heteroscedasticity of the data, we conducted a Box-Cox transformation of data and model results as described in section 2.2.2 (Box and Cox, 1964). For the application to BELAMO, the transformation parameters $\lambda_1$ and $\lambda_2$ were chosen as 0.5 and 0, respectively.
2.4. Application to biogeochemical and ecological lake model

2.4.4 Prior distribution

The parameters included in the parameter estimation procedure comprise a selection of model parameters and the standard deviations of the observation errors and the bias for all considered output variables.

Concerning the model parameters of the deterministic model, $\theta$, we concentrated on $k_{gro,ALG,To}$, the growth rate of phytoplankton at reference temperature and saturating light intensity, $k_{death,ALG,To}$, the death rate of phytoplankton at reference temperature, $k_{gro,ZOO,To}$ and $k_{death,ZOO,To}$, the growth and death rate of zooplankton at reference temperature. Furthermore, the aerobic, anaerobic and anoxic mineralization rates at reference temperature, $k_{miner,aero,sed,To}$, $k_{miner,anae,sed,To}$ and $k_{miner,anox,sed,To}$ are considered for calibration, as well as the threshold concentration of phytoplankton when zooplankton feeding switches from quadratic to Monod-type limitation, $K_{Feed}$, and $f_{X_I,rivers}$, the fraction of inert organic particles of allochthonous input. We chose these parameters because of uncertain knowledge about their value in nature. Good estimates are neither available for the average growth and death rates of the aggregated phyto- and zooplankton communities, nor for the mineralization rates that depend on the composition and abundances of the bacterial communities. Moreover, these parameters are the most influential ones, as they describe main aspects of the plankton kinetics and control the nutrient dynamics to some extent. For a more detailed description of the parameter choice and the results of a sensitivity analysis we refer to Dietzel et al., 2012. The prior marginals, reflecting our vague prior knowledge about the parameter values, are given in Table 2.2. Except for $f_{X_I,rivers}$, lognormal distributions, which avoid biologically impossible negative values, were chosen as prior marginals for the model parameters with a mode representing our best estimate of the parameter value. As $f_{X_I,rivers}$ defines a fraction, this parameter is restricted to values between 0 and 1. We performed a transformation of this parameter by means of an arctangent function. For the prior marginal of the transformed parameter, a normal distribution was assumed. The prior knowledge of all calibrated model parameters was formulated by relatively wide marginal distributions accounting for the high parameter uncertainty. The assumed values of the remaining parameters of $\theta$ can be found in Omlin et al., 2001b, Mieleitner and Reichert, 2008 and Dietzel et al., 2012. The results of the calibration and prediction method depend only slightly on the choice of these remaining, not calibrated model parameters, as the most influential and less known parameters were chosen for calibration.

Because the Box-Cox transformation accounts for heteroscedasticity, we can choose the same standard deviation of the observation error over the whole range of (Box-Cox transformed) measured concentrations for each output variable. The same is true for
Table 2.2: Prior marginals for the calibrated parameters of $\theta$ of the deterministic model.

<table>
<thead>
<tr>
<th>Name</th>
<th>Unit</th>
<th>Distribution</th>
<th>Mean</th>
<th>StDev</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_{\text{gro},\text{ALG},T_0,I_0}$</td>
<td>d$^{-1}$</td>
<td>Lognormal</td>
<td>1.79</td>
<td>0.5</td>
</tr>
<tr>
<td>$k_{\text{death},\text{ALG},T_0}$</td>
<td>d$^{-1}$</td>
<td>Lognormal</td>
<td>0.0721</td>
<td>0.05</td>
</tr>
<tr>
<td>$k_{\text{gro},\text{ZOO},T_0}$</td>
<td>d$^{-1}$</td>
<td>Lognormal</td>
<td>0.397</td>
<td>0.25</td>
</tr>
<tr>
<td>$k_{\text{death},\text{ZOO},T_0,zh}$</td>
<td>d$^{-1}$</td>
<td>Lognormal</td>
<td>0.0675</td>
<td>0.05</td>
</tr>
<tr>
<td>$K_{\text{Feed}}$</td>
<td>gDM m$^{-3}$</td>
<td>Lognormal</td>
<td>0.295</td>
<td>0.3</td>
</tr>
<tr>
<td>$\hat{f}_{\lambda_1,\text{rivers}}$ (transformed)</td>
<td>–</td>
<td>Normal</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>$k_{\text{miner},\text{aero},\text{sed},T_0,zh}$</td>
<td>d$^{-1}$</td>
<td>Lognormal</td>
<td>0.295</td>
<td>0.3</td>
</tr>
<tr>
<td>$k_{\text{miner},\text{anae},\text{sed},T_0,zh}$</td>
<td>d$^{-1}$</td>
<td>Lognormal</td>
<td>0.295</td>
<td>0.3</td>
</tr>
<tr>
<td>$k_{\text{miner},\text{anox},\text{sed},T_0,zh}$</td>
<td>d$^{-1}$</td>
<td>Lognormal</td>
<td>0.295</td>
<td>0.3</td>
</tr>
</tbody>
</table>

The bias. We assumed lognormal priors for the standard deviations of the observation errors representing the parameter vector $\psi$. These values are listed in Table 2.3 in Box-Cox transformed units. The chosen means reflect our knowledge that, due to spatial heterogeneity (“patchiness”) and higher measurement error, plankton data are prone to larger observation errors than nutrient data. The values were chosen by assuming a smaller relative error compared to the typical (average) concentrations of the different variables and transforming the respective absolute values which have the same units as the measured variables by a Box-Cox transformation to the shown means.

Last, Table 2.4 contains the priors for the standard deviations of the bias in Box-Cox transformed units. We chose exponential distributions for which the probability density

Table 2.3: Prior marginals for the parameters $\psi$ of the error model (on a Box-Cox transformed scale).

<table>
<thead>
<tr>
<th>Name</th>
<th>Unit</th>
<th>Distribution</th>
<th>Mean</th>
<th>StDev</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_{\varepsilon,\text{ALG}}$</td>
<td>gWM$^{1/2}$ m$^{-3/2}$</td>
<td>Lognormal</td>
<td>0.07</td>
<td>0.02</td>
</tr>
<tr>
<td>$\sigma_{\varepsilon,\text{ZOO}}$</td>
<td>gWM$^{1/2}$ m$^{-3/2}$</td>
<td>Lognormal</td>
<td>0.08</td>
<td>0.02</td>
</tr>
<tr>
<td>$\sigma_{\varepsilon,\text{NO}_3}$</td>
<td>gN$^{1/2}$ m$^{-3/2}$</td>
<td>Lognormal</td>
<td>0.022</td>
<td>0.002</td>
</tr>
<tr>
<td>$\sigma_{\varepsilon,\text{O}_2}$</td>
<td>gO$^{1/2}$ m$^{-3/2}$</td>
<td>Lognormal</td>
<td>0.11</td>
<td>0.01</td>
</tr>
<tr>
<td>$\sigma_{\varepsilon,\text{HPO}_4}$</td>
<td>gP$^{1/2}$ m$^{-3/2}$</td>
<td>Lognormal</td>
<td>0.007</td>
<td>0.0002</td>
</tr>
</tbody>
</table>
increases with decreasing value of the standard deviation. This reflects our desire to avoid bias if possible. The differences between the priors for the different output variables indicate rather a relative weighting than an assumption of the absolute magnitude of the bias, they coarsely represent around 25% bias (compared to typical concentrations) for the plankton and around 15% for the chemical variables. In general, we assume the bias to be larger than the observation error. The parameter $\beta$ that determines the correlation structure was fixed to $0.00005 \, \text{d}^{-2}$, representing a correlation time of about four months. As this parameter would be difficult to identify and is coarsely known for a lake strongly affected by annual mixing processes, it was not included in the calibration procedure. The standard deviations of the bias and the parameter $\beta$ represent the parameter vector $\xi$.

For the parameter estimation, the latest version of UNCSIM [Reichert, 2005] a program package for statistical inference, identifiability analysis and uncertainty analysis, was used by coupling with AQUASIM. Further calculations for the prediction layout and the approximation of the posterior distribution of model results at its maximum for estimation of the prediction uncertainty were done with the statistics software R (http://www.r-project.org/).

### 2.4.5 Results and discussion

A calibration period of 20 years (ending in 1995) was chosen, the description of the observations of those years served as layout $L_1$. The prediction uncertainty was estimated for the validation period of the following 10 years defined by $L_2$. Observations are also available for layout $L_2$. They were not used for calibration, but serve as validation data for comparison with model results. The deterministic model function, $y_M^L(\theta)$, consists of

<table>
<thead>
<tr>
<th>Name</th>
<th>Unit</th>
<th>Distribution</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_{B,ALG}$</td>
<td>gWM$^{1/2}$ m$^{-3/2}$</td>
<td>Exponential</td>
<td>0.3</td>
</tr>
<tr>
<td>$\sigma_{B,ZOO}$</td>
<td>gWM$^{1/2}$ m$^{-3/2}$</td>
<td>Exponential</td>
<td>0.4</td>
</tr>
<tr>
<td>$\sigma_{B,NO3}$</td>
<td>gN$^{1/2}$ m$^{-3/2}$</td>
<td>Exponential</td>
<td>0.11</td>
</tr>
<tr>
<td>$\sigma_{B,O2}$</td>
<td>gO$^{1/2}$ m$^{-3/2}$</td>
<td>Exponential</td>
<td>0.5</td>
</tr>
<tr>
<td>$\sigma_{B,HPO4}$</td>
<td>gP$^{1/2}$ m$^{-3/2}$</td>
<td>Exponential</td>
<td>0.035</td>
</tr>
</tbody>
</table>
model outputs for the different variables oxygen, phosphate, nitrate, phytoplankton and zooplankton in the epilimnion and hypolimnion at discrete simulation times specified by the same layouts.

Inference was again done by applying all three options for deriving a normal distribution as an approximation to the posterior at its maximum described in section 2.2.3. The maximum was obtained before by applying the shuffled complex evolution algorithm (Duan et al., 1993) with UNCSIM. Application of the finite difference algorithm was not successful as we could not achieve that all incremented simulations led to a smaller posterior than the base simulation found by the optimization algorithm. Due to the computational demand we could not redo the analysis very frequently at the newly found maximum. The complex shape of the posterior, the inaccuracy of the optimization algorithm, and the small increments required for the finite difference approximation may all have contributed to this problem. In contrast to the didactical example, for this application, it was possible to fit a multivariate normal distribution by density fitting (we chose the simplified approach described in section 2.2.3 that keeps the mean at the parameter values corresponding to the maximal posterior pdf value). Finally, we did importance sampling from a uniform distribution in a ball (large enough to approximately cover the whole posterior) which was refined by an importance sampling using the normal distribution with mean and variance-covariance matrix from the first importance sampling as a sampling distribution. Table 2.5 summarizes the computational demand of the applied techniques.

Table 2.5: Number of simulation runs needed for the optimization and the different approximation techniques for the application case of BELAMO. All numbers refer to the simulation runs actually performed for the results shown herein.

<table>
<thead>
<tr>
<th>Application Case</th>
<th>Optimization</th>
<th>Finite Difference</th>
<th>Density Fitting</th>
<th>Importance Sampling</th>
<th>MCMC</th>
</tr>
</thead>
</table>

1 some preliminary optimization runs for parameter estimation were done before
2 we performed several iterations but still new “maxima” were found
3 1 importance sampling from a uniform ball followed by 1 importance sampling from a normal distribution, sample size 200

The comparison of prior and posterior marginal distributions of the model parameters
and the parameters of the error model are shown for both successful approaches in Fig. 2.3. The results presented in this figure show a dependence on the chosen approximation technique, but for most parameters the marginals are similar. For some parameters the position of the maximum depends on the chosen approximation technique, for others just the width of the distribution. When applying the technique of importance sampling from a uniform distribution in a ball around the maximum, it must be guaranteed that the ball is large enough to cover the relevant part of the posterior to get unbiased results. This can be difficult as the efficiency of the procedure decreases with increasing radius of the ball. The method is more robust when followed by an importance sampling from the resulting normal distribution as shown herein. For some parameters the differences between prior and posterior distribution are relatively large. Due to unknown bacterial densities, the mineralization rates are not well known and wide priors were chosen. The shift to the posterior shows that we learn something about the values of these parameters. For the standard deviations of the observation errors, the deviations between priors and posteriors show that the priors might have been chosen slightly too small or too narrow or that some of the natural variability is “interpreted” as observation error by the model. For the standard deviations of the bias, the posterior distributions show how few bias is possible, which is the only requirement formulated by the exponential priors, here the shift is not an indication for a prior-data conflict.

The long-dashed lines in Figs. 2.4–2.6 show the simulation results of BELAMO for the fitted output variables at the maximum of the posterior for both the calibration and the prediction time. They represent the median of the distribution of model results, $y_{M}^{L_{1,2}}$, due to the posterior distribution of model parameters, $\theta$, conditional on the model results of layout $L_1$ without contribution of the bias. The dark grey areas in Figs. 2.4–2.6 depict our knowledge about the true state of the system ($y_{M}^{L_{1,2}} + B_{M}^{L_{1,2}}$, model results plus bias without measurement error), the light grey areas that of potential new observations ($Y_{M}^{L_{1,2}}$). In the first part, observations were used to enhance our knowledge. In the second part, observations were not used for inference and thus our results can be seen as a prediction of the “future” state of the system. Observations within layout $L_2$ are shown for validation of the predictive power of the model. Depending on the correlation time, the gained information about the bias for layout $L_1$ gives information also for the beginning of the validation time. After the correlation between the bias of layout $L_1$ and $L_2$ decreased to zero, the median of $y_{M}^{L_{1,2}} + B_{M}^{L_{1,2}} | Y_{M}^{L_1}$ (solid line in Fig. 2.4–2.6) equals the median of $y_{M}^{L_{1,2}}$ (long-dashed line). The shown results represent the results from the density fitting technique. However, the results from the importance sampling technique, conducted for this case study, show large similarities.
Figure 2.3: Prior (dotted, supported by light grey shading of the areas below the lines) and posterior (supported by dark grey shading of the areas below the lines) marginals of the model parameters from the posterior approximation by combined ball/normal importance sampling (dashed) and a density fit (dash-dotted).
2.4. Application to biogeochemical and ecological lake model

Figure 2.4: Phytoplankton ((a),(b)) and zooplankton ((c),(d)) concentrations in the epilimnion ((a),(c)) and hypolimnion ((b),(d)) of Lake Zurich. Data points (markers), output of the deterministic model (long-dashed), median (solid) and 95% credibility bounds (dark grey area with dashed boundaries) of bias-corrected output and median (solid; same as for bias-corrected output) and 95% credibility bounds (dark and light grey areas with dotted boundaries) of predictions of new observations (including observation error) for the whole simulation time.
Figure 2.5: Oxygen concentrations in the epilimnion (a) and hypolimnion (b) of Lake Zurich. Data points (markers), output of the deterministic model (long-dashed), median (solid) and 95% credibility bounds (dark grey area with dashed boundaries) of bias-corrected output and median (solid; same as for bias-corrected output) and 95% credibility bounds (dark and light grey areas with dotted boundaries) of predictions of new observations (including observation error) for the whole simulation time.

The uncertainty due to parameter uncertainty and bias (width of the dark grey band in Figs. 2.4b,2.6) is generally quite narrow during the calibration period (where the data points constrain the predictions) and increases considerably during the prediction phase. For all chemicals (dissolved oxygen, nitrate and phosphate) the observation error can be neglected in the prediction period compared to the bias, whereas there is still considerable observation error for the biological variables. As mentioned before, this reflects the larger sampling error due to “patchiness” and the larger measurement error of the biological variables compared to the chemical variables. Generally, a comparison of the results indicates that annual patterns of both phyto- and zooplankton seem to be most difficult to be represented by the model. The deviations most likely result from simplifications by spatial and functional aggregation in the model. These simplifications are less realistic for biological than for chemical variables.

The good representation of data by the model results plus bias compared to the poorer
Figure 2.6: Nitrate ((a),(b)) and phosphate ((c),(d)) concentrations in the epilimnion ((a),(c)) and hypolimnion ((b),(d)) of Lake Zurich. Data points (markers), output of the deterministic model (long-dashed), median (solid) and 95% credibility bounds (dark grey area with dashed boundaries) of bias-corrected output and median (solid; same as for bias-corrected output) and 95% credibility bounds (dark and light grey areas with dotted boundaries) of predictions of new observations (including observation error) for the whole simulation time.
deterministic model results demonstrates the large contribution of bias in the model results. A more detailed insight into the importance of bias for the different output variables can be given by Fig. 2.7, which shows the median and the 95% credibility interval of the bias and the observation error of all output variables. All graphs indicate that the structural error of the model is mainly captured by our bias description, whereas the measurement error seems to be normally distributed and independent as required.

**Figure 2.7:** Median and 95% credibility interval of posterior estimate of bias (solid line and shaded area bounded by dashed lines) and median posterior estimate and uncertainty of observation error (markers and vertical lines) of observed output variables in Lake Zurich. The units refer to the units of respective output variables on a Box-Cox transformed scale.

In general, the simulation and uncertainty results show the ability of the model to describe the most important processes in the studied lake in a meaningful way. But the
deviations between data and model results also demonstrate the difficulty of a multi-objective calibration of different output variables over a long-term period. The large uncertainties indicate the limited prediction ability of such a deterministic model.

## 2.5 Conclusions

Previous experiences in the literature, the technical development in this paper and the practical experiences with a didactical example and an application to a lake water quality model in this paper lead us to the following conclusions:

1. The normal approximation to the posterior at its maximum and the linearized uncertainty propagation technique developed in this paper for the bias description technique described above considerably reduces the computational requirements of the method introduced by Reichert and Schuwirth, 2012. Starting with the same optimization procedure (that is hard to avoid for any model calibration procedure), the tens of thousands of simulation runs for the Markov chain are replaced by a few hundred (or few thousand) simulation runs required to achieve the normal approximation to the maximum of the posterior and the error propagation. This makes the approximate version of the important technique mentioned above applicable to a much larger class of computationally demanding simulation models, specifically to the large group of environmental models.

2. The shape of the posterior is often complicated (Gupta et al., 2003b) which can make its approximation by a normal distribution challenging. We proposed three options to get such an approximation: A finite difference approximation at the maximum, a fit of a normal density to a parameter sample with corresponding posterior values, and derivation of the mean and variance-covariance matrix from such a sample using importance sampling. The finite difference approach may be difficult to apply due to local maxima of the posterior density and an inaccurate determination of the maximum by the numerical maximization algorithm. The density fitting approach may have difficulties due to the large number of parameters and the constraints they have to fulfill. Indeed, each of these approaches failed in one of the two case studies of this paper. The importance sampling approach seems to be more robust and could be applied for both cases.

3. The experiences with the lake water quality model BELAMO indicate strong difficulties in calibration. Unlike most other lake water quality models (Mooij et al., 2012),
Chapter 2. Calibration of computationally demanding models

BELAMO (Omlin et al., 2001b; Mieleitner and Reichert, 2006; Mieleitner and Reichert, 2008) uses closed material balances by including a sediment compartment with two zones and coupling diagenetic processes in the sediment to the nutrient dynamics of the water column. This leads to a much stronger coupling between different model variables and makes unbiased calibration much more difficult than with independent parameters for release of nutrients from the sediment. For this reason, this model is particularly well suited to test a bias modeling approach. Our results indicate that calibration without significant amount of bias seems not to be possible for this model. The bias description technique used in this paper seems to address these problems in an adequate way.

4. It is remarkable that the resulting separation of the deviation between model and observations into observation error and bias leads to a significant observation error and bias for phyto- and zooplankton whereas the bias dominates strongly over the observation error for dissolved oxygen, nitrate and phosphate. This result is in qualitative accordance with our prior expectation, but the observation errors are identified to be larger than our priors.

2.6 Outlook

Both, the original bias description technique (Kennedy and O’Hagan, 2001; Bayarri et al., 2007; Reichert and Schuwirth, 2012), and the simplified, efficient version developed in this paper, are useful techniques to consider the presence of bias in model calibration. The approximate technique has considerable advantages for computationally demanding models where the original technique can not be applied due to long simulation times. Another approach to reduce computation times is emulation. With a dynamic emulator, a statistical description of the behavior of the model, the original bias description technique would be applicable to computationally expensive models as well. Some attempts have been made towards dynamic emulators (Bhattacharya, 2007; Conti and O’Hagan, 2010; Conti et al., 2009; Liu and West, 2009; Reichert et al., 2011). This area is still ongoing research. A hybrid approach, where the inference is done by emulating only the posterior of the model and then propagating this posterior through the original model, is an intermediate stage and subject of future work. It is important to note that the suggested technique is not able to reduce the bias. In addition to the option of considering bias, we would be interested in techniques that support the identification of causes and reduction of bias such as those described in Vrugt et al., 2005; Vrugt and Robinson, 2007; Lin and Beck, 2007; Reichert and Mieleitner, 2009.
2.7 Acknowledgments

Both the complexity of the bias or structural uncertainty problem and the variation in computational demands of different simulation models require a toolbox of techniques to be available to the modeler. It was the intent of this paper to add a relatively simple and computationally efficient technique to this toolbox.

2.7 Acknowledgments

We thank Richard Forster and Oliver Köster from WVZ, Barbara Känel, Pius Niederhauser, Irene Purtschert and Peter Spohn from AWEL and Daniel Streit and Christian Leu from BAFU for kindly providing their data on chemistry, biology and physics of Lake Zurich, discharging waste water treatment plants and inflowing rivers. We thank the Swiss National Science Foundation (SNF) for the financial support of the project and three anonymous reviewers for their helpful comments to the manuscript.
Appendix A

Supplementary data to: Calibration of computationally demanding and structurally uncertain models with an application to a lake water quality model

A.1 Didactical example: wrong priors

For the scenario of wrong priors, which is frequently encountered in environmental modeling, we chose the means of the priors of the model parameters to be 0.5 times larger than the true values and again a standard deviation of 50\% of these wrong values. The observation error and bias parameter priors were chosen as in the base scenario.

A.1.1 Results and discussion

Fig. A.1 shows the prior marginals for the scenario of wrong priors and the corresponding results for the posterior marginals of the parameter distribution for the importance sampling approximation in comparison with the analogous results of a Markov chain Monte Carlo technique which samples the full posterior distribution. The finite difference approach failed in this example, as the resulting inverse of the negative Hessian was not positive definite. This technique seems to be sensitive to the choice of prior distributions and the choice of the increment for evaluation of the Hessian. The problem
shows again, that the importance sampling approximation is the most robust method out of the three suggested approximation methods.

The figure shows the effect of the choice of the prior. In many cases the two techniques show problems to meet the true parameter value and uncertainties are underestimated in some cases. For wide priors, the maxima of the posterior distributions are at least close to the true parameter values. This is meaningful, as the choice of a wrong value as the mean of the prior reflects a larger uncertainty about the parameter, which should be represented by relatively wide prior distributions. In general, the difference between the approximated and the full MCMC technique is not large in the sense that both techniques show similar failures. Solely, the posterior distributions derived from the MCMC technique tend to be wider than the importance sampling ones.

Fig. A.2 shows the model predictions of our knowledge of the ‘true’ results, $y_{M_1} + B_{M_i}$ | $Y_{M_1}$, and of observations, $Y_{M_1}$ | $Y_{M_1}$, for the one successful posterior approximation technique, again in comparison with the MCMC results, for the wrong prior choice.

The results presented in Fig. A.2 show some differences to the base scenario with priors at the true parameter value, mainly slightly underestimated uncertainties. However, the results indicate large similarities between the importance sampling and the MCMC method, which was desired by our approximation approach for computationally demanding and structurally uncertain models.
A.1. Didactical example: wrong priors

Figure A.1: Prior (dotted, supported by light gray shading of the areas below the lines) and posterior (supported by dark gray shading of the areas below the lines) marginals of the model parameters for the posterior approximation by finite difference method (long-dashed), importance sampling (short-dashed) and a full Markov chain Monte Carlo method (solid). Vertical lines indicate the true values of the model parameters used for producing the synthetic data (there are no true values of the parameters of the bias).
Figure A.2: Results of the median and 95% credibility intervals for both variables $C_M$ and $C_S$ (top row; observations used for calibration are marked by solid symbols, those not used for calibration by open symbols; output of the deterministic model (long-dashed line), bias-corrected output (deterministic model plus bias; median as solid line, 95% credibility bounds as dark gray area bounded by dashed lines), and prediction for new observations (including observation error; median as solid line (same as for bias-corrected output), 95% credibility bounds shown as dark and light gray areas bounded by dotted lines)), results of the median and 95% credibility intervals for the bias (solid line and dark gray areas bounded by dashed lines) and of the observation error (dots and vertical line segments) for the variables $C_M$ (middle row) and $C_S$ (bottom row). The left column shows results for posterior approximation by importance sampling and the right column the results of the full MCMC method.
A.1. Didactical example: wrong priors
Chapter 3

Effects of changes in the driving forces on water quality and plankton dynamics in three Swiss lakes – long-term simulations with BELAMO

Anne Dietzel, Johanna Mieleitner, Sascha Kardaetz and Peter Reichert
Published in Freshwater Biology (2012), DOI: 10.1111 /fwb.12031
Chapter 3. Long-term simulations with BELAMO

Summary

1. With a modified version of the lake model BELAMO, we were able to describe the essential features of the dynamics of nutrients, dissolved oxygen, phyto- and zooplankton in three lakes of different trophic status over periods of 19–30 years, with essentially the same model parameters for all three lakes. This is remarkable, as the measured nutrient inputs decreased considerably during the simulated time period.

2. Despite having done this before for a period of 4 years with an earlier version of the model, a considerable effort was required that led to a series of model modifications without which the data could not be matched. This demonstrates that long-term calibration of a model that combines processes in the water column with mineralization in the sediment can be difficult.

3. Due to the necessarily simplified processes within the model, there is a bias in its output. We applied a recently developed technique for model calibration and uncertainty analysis to address bias and multiple calibration criteria. To account for the demanding long-term simulations, a simplified numerical implementation of this technique was used.

4. Our results demonstrate good understanding of the chemical state of the lake during the calibration period but less of the biological variables. The credibility intervals used to visualize this knowledge widen substantially during the prediction period (consisting of the last 10 years of the simulation).

5. The joint calibration of the model with long-term data from lakes of different trophic status is possible but only with considerable prediction uncertainty. Due to the explicit consideration of bias in our calibration technique, we are able to estimate quantitatively the uncertainty of our knowledge about chemical and biological variables in the lake.

Keywords: Input loads; Lake modeling; Long term; Model bias; Universality.

3.1 Introduction

Water quality and plankton dynamics are important indicators of the ecological status of a lake. In addition to a general responsibility to conserve ecosystems, deterioration
3.1. Introduction

in lake water quality affects human interests directly. However, water quality and plankton dynamics are affected by many factors. For example, wastewater is discharged into lakes, agriculture produces diffuse loads of herbicides and nutrients, which either run off directly into lakes or enter indirectly via polluted rivers. Due to the installation and improvement of wastewater treatment plants and because of changes in the use of chemicals in households and industry, in agricultural practice and in the human population and the climate, these external driving forces shift significantly over time.

To understand the behaviour of a lake under changing driving forces (such as nutrient loads or climate change), modeling is an important tool. A model used for this purpose should represent our quantitative understanding of the main biogeochemical and ecological processes within the lake, and it should be as universal as possible. As stated by Mieleitner and Reichert, 2006, a high degree of universality of ecological models is important for two main reasons. First, from a fundamental research perspective, it indicates that the model captures the dominant underlying mechanisms. Second, from an applied perspective, universality gives a model predictive power even when driving forces are changing. To evaluate and check the universality of biogeochemical-ecological lake models, case studies should be performed for different lakes and over time periods during which the external driving forces may shift considerably.

Several lake models have been developed and applied over the past few decades (for reviews about lake modeling see Arhonditsis and Brett, 2004; Jørgensen, 2010; Mooij et al., 2010). These include SALMO (Benndorf and Recknagel, 1982), its further developments SALMO-1D and SALMO-HR (Baumert and Benndorf, 2005; Petzoldt et al., 2005), (DYRESM-)CAEDYM (Bruce et al., 2006; Hamilton and Schladow, 1997; Rinke et al., 2009; Romero et al., 2004; Schladow and Hamilton, 1997; Tanentzap et al., 2007; Trolle et al., 2008), PROTECH (Elliott et al., 1999b; Elliott et al., 1999a; Elliott et al., 2000; Elliott et al., 2005; Elliott et al., 2006; Elliott et al., 2007; Elliott et al., 2010; Elliott and Thackeray, 2004; Reynolds et al., 2001) and BELAMO (Omlin et al., 2001b; Omlin et al., 2001a; Mieleitner and Reichert, 2006; Mieleitner and Reichert, 2008). We used the Biogeochemical and Ecological Lake MOdel (BELAMO) and assessed its universality by applying it to lakes of different trophic status over a long period. The main reason for this choice (and for the development of BELAMO) was the emphasis of the model on closing the element cycles by coupling a lake water column model with a sediment model and explicitly modeling mineralization of organic particles in the sediment rather than parameterizing only the effect of the sediment on the water column by source or sink terms of nutrients and dissolved oxygen. This was identified as a major deficit of most lake models by Mooij et al., 2010, and it is of relevance for investigations of the effect of changes in nutrient inputs.
Chapter 3. Long-term simulations with BELAMO

The model BELAMO aims at a joint calculation of mass balances of nutrients, oxygen, organic particles, phytoplankton and zooplankton in the lake water column and two sediment layers. The results of applying BELAMO, originally implemented for Lake Zurich only (Omlin et al., 2001b; Omlin et al., 2001a), to the lakes Greifensee and Walensee indicated that it is possible to devise a more universal model applicable to three lakes that are similar in many properties but have different trophic status (Mieleitner and Reichert, 2006). Later, Mieleitner et al., 2008 and Mieleitner and Reichert, 2008 analyzed the influence of taxonomic plankton aggregation on the universality of the lake model by extending the model to consider different functional groups of phytoplankton. To refrain the computational burden, the continuous vertical resolution of the original model was reduced to four mixed compartments representing the epilimnion and the hypolimnion of the lake and two sediment layers that were already considered in the one-dimensional model (Mieleitner and Reichert, 2008). The results indicated that the universality of the relatively simple plankton sub-model was not improved by introducing functional groups, as the predictive capability was reduced. The main reason for this was that the additional state variables of the functional groups model were less constrained by mass-balance equations than the state variables of the aggregated (i.e. total phytoplankton) model (the division of phosphate conversion by phytoplankton into contributions by different functional groups is not determined by mass-balance constraints). Additionally, total biomass was not assessed more accurately by the sum of the functional groups than by the aggregated model, and the representation of the separate concentrations of the different functional groups was poor (Mieleitner and Reichert, 2008).

The simulations reported by Mieleitner and Reichert, 2006 and Mieleitner and Reichert, 2008 were conducted for a period of four years only, during which the driving forces did not change significantly. Longer-term simulations covering a period of change would provide a further test of the universality of the model. In the present study, therefore, we performed simulations over the last 19 or 30 years (depending on the data available for particular lakes) of Greifensee (19 years), Lake Zurich (30 years) and Walensee (30 years), to test the model for time-independency of parameters and process formulations and for predictive power under changing environmental conditions (in particular decreasing phosphate inputs).

Our focus was to find model formulations and parameter values that led to model outputs consistent with empirical data over a long period, to reduce the systematic model errors and to estimate all the remaining uncertainties in the model, including model bias. This required the development of new, computationally efficient techniques based on work on statistical bias description (Craig et al., 1996; Craig et al., 2001; Kennedy and O’Hagan, 2001; Higdon et al., 2004; Bayarri et al., 2007). Using the calibration
3.2 Methods

3.2.1 Study area

Greifensee and Lake Zurich are both located in the north-eastern part of the Swiss plateau (Fig. 3.1). The area around Greifensee is mainly agricultural, while the catchment of Lake Zurich is more urban. Walensee is south-east of the other two lakes, in the foothills of the Swiss Alps. Linthkanal, the outflow of Walensee, flows into Lake Zurich. Lake Zurich is separated into two parts by a natural dam, Seedamm. As vertical profiles were taken in the lower basin, for modelling we therefore concentrated on lower Lake Zurich. As measured by prevailing phosphate concentrations, Greifensee is still eutrophic, Lake Zurich mesotrophic and Walensee oligotrophic.

3.2.2 BELAMO: model description

Basic model version

The first version of BELAMO was a one-dimensional (vertical) mechanistic model of the nutrient and plankton dynamics in Lake Zurich. It calculates horizontally averaged concentrations changing with the depth of the lake. The concentrations of biochemical variables and particles in the sediment, divided into two layers, are also considered. For an introduction to the model see Omlin et al., 2001b Mieleitner and Reichert, 2006 modified the one-dimensional model and analyzed its transferability to the lakes Greifensee and Walensee.

To reduce simulation time, and to be able to conduct many simulations over several years to analyze functional phytoplankton groups, the continuous vertical resolution of the model was simplified to a box approach (Mieleitner and Reichert, 2008). It was difficult with this model version to ensure the persistence of all functional groups and to represent all groups reasonably well. At the level of total biomass, the results did
not appear to be significantly better than when modeling only the total phytoplankton (Mieleitner and Reichert, 2008). As it showed a higher predictive capability in the preceding project, we elected to use the aggregated version of the model, describing the total biomass of phyto- and zooplankton. We used the box version of the model to save computational time.

The box model describes the lake as four “boxes”, each with constant volume: epilimnion, hypolimnion and two sediment boxes. In these boxes, concentrations of ammonium, nitrate, phosphate, oxygen, degradable and inert dead organic particles and (in the aggregated version) the total biomass of phytoplankton and zooplankton are calculated.

The following biogeochemical and ecological processes are considered: growth, respiration and death of phyto- and zooplankton, aerobic, anoxic and anaerobic mineralization, nitrification and phosphate uptake by sinking particles. Physical processes that are included are: inflows into epi- and hypolimnion and outflow from the epilimnion, gas exchange with the atmosphere, mixing of all dissolved substances between epi- and
3.2. Methods

hypolimnion (mixing coefficients were estimated from temperature data), vertical migration of zooplankton in the water column, sedimentation of particles, advection (if the inflow to the lake is into the hypolimnion), diffusion of dissolved substances between the two sediment layers and the water in the hypolimnion, as well as accumulation of sediment and permanent burial.

Here, we concentrated on an application of one model to several lakes that cover a range of trophic conditions. However, the three study lakes are all pre-alpine European lakes subjected to similar climatic conditions.

For a detailed description of the original box version of BELAMO see Mieleitner and Reischert, 2008. All structural model extensions discussed in the next section are included in Figs. 3.2 and 3.3.

Figure 3.2: Structure and physical processes of the four-box version of BELAMO.
Figure 3.3: Biological and chemical processes taken into account in each of the model compartments of BELAMO. Some of them do not play a role in the sediment boxes and hence are not activated there.

Model extensions

Using the model for long-term simulations required considerable adaptations to the model used by [Mieletner and Reichert, 2008]. We assumed that not all phosphorus released by sloppy feeding and excretion during zooplankton growth is dissolved, but only a fraction $f_{sol}$. The fraction $1 - f_{sol}$ remains as organic particles.

To account for phytoplankton that grows when phosphate is limiting, we assumed the fraction of phosphorus in organic particles entering the lakes from another lake to be smaller than from rivers ($a_{P,inflow,lake} < a_{P,inflow}$). This was used where inflowing particles had to be assessed from measurements of particulate organic phosphorus or vice versa. The fraction $a_{P,inflow,lake}$ accounts for the inflow from the upper basin of
Lake Zurich at Seedamm and from Aa Uster, a river flowing from Lake Pfäffikon into Greifensee (“Aa (U.)” in Fig. 3.1).

**Initial conditions in the sediment** One main feature of our model is the joint mechanistic description of the processes in the sediment. As stated by Mooij et al., 2010, sediment nutrient cycling is rarely dynamically accounted for in lake modeling but can have a strong influence on the ecosystem. To describe this feature, initial conditions for the different kinds of particles considered are needed that ensure both mass and volume balance. As we made some modifications to the version applied by Mieleitner and Reichert, 2008, we give a short description here. The initial conditions of degradable organic ($X_S$), inert organic ($X_I$) and inorganic ($X_{II}$) particles in the sediment ($sed$), modeled as concentrations of particulate matter (symbolized by $X$) in gDM m$^{-3}$, were parameterized as follows. As the organic phosphorus content of degradable and inert organic particles were modeled as separate state variables in BELAMO, and the state variables of organic particles ($X_S$ and $X_I$) describe the mass without the phosphorus content, we introduced the mass of organic fractions including phosphorus as $X_{SP}$ and $X_{IP}$. In so doing, the first important constraint was derived from the volume balance of overall particles in the sediment:

\[
1 - \theta = \frac{X_{SP,sed}}{\rho_{XS}} + \frac{X_{IP,sed}}{\rho_{XI}} + \frac{X_{II,sed}}{\rho_{XII}} ,
\]

where $\theta$ is the volumetric water content of the sediment and $\rho$ the density of the respective particles. Second, the mass fraction (symbolized by $a$) of organic material in total sediment material, $a_{org,sed}$, and the mass fraction of degradable material in organic sediment material, $a_{deg,sed}$ provided two more equations:

\[
X_{SP,sed} + X_{IP,sed} = a_{org,sed} \cdot (X_{SP,sed} + X_{IP,sed} + X_{II,sed})
\]

and:

\[
X_{SP,sed} = a_{deg,sed} \cdot (X_{SP,sed} + X_{IP,sed}) .
\]

By assuming $a_{org,sed_{ini}}$ and $a_{deg,sed_{ini}}$ as initial conditions of organic and degradable material mass fractions, conversion of these three equations led to the first initial condition for inorganic particles, $X_{II_{ini}}$:

\[
X_{II_{sed_{ini}}} = \frac{1 - \theta}{\frac{1}{\rho_{XII}} \frac{a_{deg,sed_{ini}} a_{org,sed_{ini}}}{\rho_{XS}} + \frac{(1 - a_{deg,sed_{ini}}) a_{org,sed_{ini}}}{\rho_{XI}}} .
\]

The parameters $a_{org,sed_{ini}}$ and $a_{deg,sed_{ini}}$ are input parameters of the model and were set to 0.1 and 0.3, respectively. By introducing the organic fractions without phosphorus
content, $X_S$ and $X_I$, it follows:

$$X_P^S = \frac{X_S}{(1 - a_{P,S})}, \quad X_P^I = \frac{X_I}{(1 - a_{P,I})}$$ (3.5)

with the phosphorus contents, $a_{P,S}$ and $a_{P,I}$, which were assumed to conform to the Redfield ratio ($a_{P,\text{red}}$) initially (Redfield, 1958). Hence, the other two initial conditions resulted in:

$$X_{S,\text{sed ini}} = \frac{a_{\text{deg,sed ini}} \cdot a_{\text{org,sed ini}} \cdot (1 - a_{P,\text{red}})}{1 - a_{\text{org,sed ini}}} \cdot X_{\text{II, sed ini}}$$ (3.6)

and

$$X_{I, \text{sed ini}} = \frac{(1 - a_{\text{deg,sed ini}}) \cdot a_{\text{org,sed ini}} \cdot (1 - a_{P,\text{red}})}{1 - a_{\text{org,sed ini}}} \cdot X_{\text{II, sed ini}}.$$ (3.7)

Equations (3.4), (3.6) and (3.7) served as initial conditions of inorganic ($X_{\text{II}}$), degradable organic ($X_S$) and inert organic ($X_I$) particles in the sediment. Note that the notation for the phosphorus mass fractions $a_{P,S}$ and $a_{P,I}$ follows the equations:

$$a_{P,S} = \frac{X_P^S - X_S}{X_P^S}, \quad a_{P,I} = \frac{X_P^I - X_I}{X_P^I}.$$ (3.8)

This notation differs from the one described in Omlin et al., 2001b and Mieleitner and Reichert, 2006, where the mass fractions of phosphorus in organic particles were related to the organic fractions without phosphorus content. As the phosphorus content is small, this change did not have a significant effect on the model results, but is in better agreement with common notation. The calculation of phytoplankton phosphorus content was changed accordingly.

**Active movement of zooplankton** To make the zooplankton sub-model of BELAMO more realistic, we decided not to compare with model results the volume-weighted averages of zooplankton concentration over the whole lake depth, but to use the two depth-integrated samples for an estimate of epilimnial and hypolimnial concentrations. This is described in more detail in the Supporting Information B.1 and was possible only for Lake Zurich and Walensee. With the zooplankton mobility component of the model, implemented in addition to normal turbulent diffusion, the model tended to overestimate the zooplankton concentrations in the hypolimnion and to underestimate the concentrations in the epilimnion. This was avoided by switching off this mobility component, which was originally implemented as a diffusive exchange process between the two compartments, hence balancing the concentration of zooplankton in the epi- and hypolimnion. The data gave hints that zooplankton moves more actively towards food. Therefore, we introduced an advection process that represents the movement of zooplankton to the
3.2. Methods

epilimnion, where phytoplankton is present. The mass flux of zooplankton from hypo-
to epilimnion was formulated as follows:

\[ F_{\text{up,ZOO}} = -v_{\text{up,ZOO}} \cdot A_{\text{surf}} \cdot X_{\text{ZOO,hypo}}, \quad (3.9) \]

where \( A_{\text{surf}} \) is the surface area of the specific lake and \( X_{\text{ZOO,hypo}} \) is the zooplankton concentration in the hypolimnion. The upwards velocity of zooplankton, \( v_{\text{up,ZOO}} \), was set to 5 m d\(^{-1}\).

Zooplankton grazing Compared to Omlin et al., 2001b and Mieleitner and Reichert, 2008, the implementation of zooplankton growth by feeding on phytoplankton was modified. Zooplankton growth rate was formulated as:

\[ r_{\text{gro,ZOO}} = k_{\text{gro,ZOO},T_0} \cdot e^{\beta_{\text{ZOO}} \cdot (T - T_0)} \cdot \frac{S_{O_2}}{K_{O_2,ZOO} + S_{O_2}} \cdot f_{\text{ALG}} \cdot X_{\text{ZOO}}, \quad (3.10) \]

where \( k_{\text{gro,ZOO},T_0} \) is the specific maximum zooplankton growth rate at the reference temperature \( T_0 \) (in our case 20°C), \( \beta_{\text{ZOO}} \) is the temperature dependence coefficient of zooplankton and \( T \) the actual temperature. Furthermore, the growth rate includes a Monod-type limitation term for \( O_2 \) (\( S \) refers to a concentration of a dissolved substance, in this case measured in gO m\(^{-3}\)) with the half-saturation concentration for zooplankton growth with respect to oxygen, \( K_{O_2,ZOO} \), and depends on the actual zooplankton concentration, \( X_{\text{ZOO}} \). The term \( f_{\text{ALG}} \) describes the dependence of the zooplankton growth on the phytoplankton concentration. Rather than using a simple linear approach \( f_{\text{ALG}} = X_{\text{ALG}} \), as did Omlin et al., 2001b and Mieleitner and Reichert, 2008, we chose the following non-linear approach:

\[ f_{\text{ALG}} = \begin{cases} \frac{X_{\text{ALG}} - \lim_a}{K_{\text{ALG,ZOO}} + X_{\text{ALG}} - \lim_a} & \text{if } X_{\text{ALG}} > K_{\text{Feed}} \\ \text{lim}_b \cdot X_{\text{ALG}}^2 & \text{else} \end{cases} \quad (3.11) \]

with

\[ \lim_a = \frac{K_{\text{ALG,ZOO}} + 2 \cdot K_{\text{Feed}} - \sqrt{K_{\text{ALG,ZOO}}^2 + 2 \cdot K_{\text{ALG,ZOO}} \cdot K_{\text{Feed}}}}{2} \quad (3.12) \]

and

\[ \text{lim}_b = \frac{K_{\text{Feed}} - \lim_a}{K_{\text{Feed}} \cdot (K_{\text{ALG,ZOO}} + K_{\text{Feed}} - \lim_a)}. \quad (3.13) \]

For small phytoplankton concentrations below the threshold for feeding \( (K_{\text{Feed}}) \), calibrated to 0.95 gDM m\(^{-3}\) for all three lakes, the dependence on phytoplankton is quadratic; above that threshold it has a Monod-type limitation. This accounts for the assumption that zooplankton consume disproportionately less phytoplankton when food
concentration is low. Test simulations showed that this modified growth rate solved the problem of very low modeled phytoplankton concentrations (in data not shown) as described by Mieleitner and Reichert, 2006 and Mieleitner and Reichert, 2008. The Monod-type limitation guarantees that zooplankton do not grow too much at high phytoplankton concentrations. For the Monod limitation, $K_{ALG,ZOO}$ is the half-saturation concentration for the growth of zooplankton on phytoplankton and was set to $0.8 \text{ gDM m}^{-3}$.

The two factors, $\lim_a$ and $\lim_b$, were chosen such that $f_{ALG}$ and its derivative are continuous for every value of $X_{ALG}$. For a visualization of the function $f_{ALG}$, see Fig. B.1 in the Appendix. It is apparent that the visualized response is similar to the more common ecological modeling approach of a Holling-type III functional response (Holling, 1959a; Holling, 1959b). One reason for the choice of such a functional response is the spatially heterogeneous distribution of prey, which is also the reason for the choice of the approach here. The main difference between the two approaches is that the point for switching from a quadratic to a Monod-type limitation is not fixed to half of the maximal rate, as in the Holling-type III definition. The threshold $K_{Feed}$ was calibrated to a value corresponding to 30% (and not 50%) of the maximal rate in our application.

**Feeding by Diptera**: Mieleitner and Reichert, 2008 mentioned that modeled zooplankton concentrations are also often too low, especially in the Greifensee. After discussions with a plankton specialist (H.R. Bürgi, Eawag), it was decided to change the formulation of the zooplankton death rate by introducing the predation by Diptera larvae (genus Chaoborus) most abundant in the shallow Greifensee (Mieleitner and Reichert, 2006). To account explicitly for that, we introduced a Diptera model part by splitting the death rate of zooplankton into two processes. “Normal” mortality has the rate $r_{death,ZOO} = k_{death,ZOO,T_0} \cdot e^{\beta_{ZOO}(T-T_0)} \cdot X_{ZOO}$, which is linearly dependent on the zooplankton concentration and includes predation by fish. $k_{death,ZOO,T_0}$ is the specific death rate and $\beta_{ZOO}$ describes the temperature dependence. The specific zooplankton death rate was set to the same value for all three lakes before calibration. An additional source of mortality due to predation of Chaoborus, only active in Greifensee, was specified and implemented in a similar way as zooplankton grazing on phytoplankton described above to account for lower grazing efficiency at low concentrations. Equation (3.14) shows the formulation of the process rate

$$r_{death,ZOO,Dipt} = k_{death,ZOO,Dipt,T_0} \cdot e^{\beta_{ZOO}(T-T_0)} \cdot f_{ZOO}.$$  

(3.14)

The dependence on zooplankton, $f_{ZOO}$, is described as

$$f_{ZOO} = \begin{cases} 
\frac{X_{ZOO}^2}{2K_{Feed,ZOO,Dipt}} & \text{if } X_{ZOO} < K_{Feed,ZOO,Dipt} \\
X_{ZOO} - \frac{K_{Feed,ZOO,Dipt}}{2} & \text{else}
\end{cases}.$$  

(3.15)
At the threshold zooplankton concentration, $K_{\text{Feed,ZOO,Dipt}}$, predation by Chaoborus switches from a quadratic to a linear dependency on zooplankton concentration, rather than switching to a Monod-type limitation as for zooplankton grazing. Test simulations showed that this new implementation no longer allows the concentration of zooplankton to decrease dramatically, in better agreement with the data. The specific death rate due to Diptera feeding was set to 0.085 $d^{-1}$, and was not included in the calibration, but the parameter calibration resulted in different values of the specific zooplankton death rate, $k_{\text{death,ZOO,T}}$, for the three lakes. This could be due to differences in the fish population in the three lakes and could be represented by a separated fish model component.

**Nutrient “preference” of phytoplankton**  Another model change resulted in a more realistic description of the nutrient preference of phytoplankton. To ensure that phytoplankton favors ammonium over nitrate in a way that is still valid at low concentrations of both ammonium and nitrate, we used the following rate descriptions for phytoplankton growth implemented as two processes (one for growth on ammonium and one for growth on nitrate):

$$r_{\text{gro,ALG,NH}_4} = k_{\text{gro,ALG,T},I_0} \cdot f_{\text{lim,T}} \cdot f_{\text{lim,I}} \cdot \text{pref}_{\text{NH}_4,\text{NO}_3} \cdot \min \left( \frac{\text{S}_{\text{NO}_3} + \text{S}_{\text{NH}_4}}{K_{\text{N,ALG}} + \text{S}_{\text{NO}_3} + \text{S}_{\text{NH}_4}}, \frac{\text{S}_{\text{HPO}_4}}{K_{\text{HPO}_4,\text{ZOO}} + \text{S}_{\text{HPO}_4}} \right) \cdot X_{\text{ALG}}$$  

(3.16)

and

$$r_{\text{gro,ALG,NO}_3} = k_{\text{gro,ALG,T},I_0} \cdot f_{\text{lim,T}} \cdot f_{\text{lim,I}} \cdot (1 - \text{pref}_{\text{NH}_4,\text{NO}_3}) \cdot \min \left( \frac{\text{S}_{\text{NO}_3} + \text{S}_{\text{NH}_4}}{K_{\text{N,ALG}} + \text{S}_{\text{NO}_3} + \text{S}_{\text{NH}_4}}, \frac{\text{S}_{\text{HPO}_4}}{K_{\text{HPO}_4,\text{ZOO}} + \text{S}_{\text{HPO}_4}} \right) \cdot X_{\text{ALG}}$$  

(3.17)

The terms $f_{\text{lim,T}}$ and $f_{\text{lim,I}}$ account for the limitations of phytoplankton growth due to temperature, $T$ and light intensity, $I$. The temperature dependence is formulated as an exponential limitation, and the dependence on light intensity as a Monod-type limitation. Hence, the specific growth rate, $k_{\text{gro,ALG,T},I_0}$, refers to reference temperature and saturating light intensity. Furthermore, both growth processes are limited by phosphate ($S_{\text{HPO}_4}$) and total nitrogen concentration ($S_{\text{NO}_3} + S_{\text{NH}_4}$). For nutrient concentrations below their specific half-saturation constant, $K_{\text{N/P,ALG}}$, the nutrient concentration at which the difference to the half-saturation constant is larger, controls the limitation, following Liebig’s law of the minimum. The preference factor, $\text{pref}_{\text{NH}_4,\text{NO}_3}$, was newly introduced into the equations as:

$$\text{pref}_{\text{NH}_4,\text{NO}_3} = \frac{p_{\text{NH}_4,\text{ALG}} \cdot S_{\text{NH}_4}}{p_{\text{NH}_4,\text{ALG}} \cdot S_{\text{NH}_4} + S_{\text{NO}_3}},$$  

(3.18)
where the preference coefficient, \( p_{NH_4,ALG} \), which determines the magnitude of preference for ammonium over nitrate, was set to 10. As the concentration of ammonium is much lower than that of nitrate, and also because phytoplankton is more limited by phosphate than by nitrogen in these lakes, this change in the growth formulation changes the model results only slightly. Hence, neither is the absolute value of the parameter crucial. However, this description is also valid for low nitrate concentrations.

**Anaerobic mineralization and methane** In the first version of BELAMO, anaerobic mineralization of organic matter was not considered (Omlin et al., 2001b). As it seemed to be relevant, especially for the sediments of the eutrophic Greifensee, it was introduced in the subsequent study (Mieleitner and Reichert, 2006) and also transferred to the box version of the model (Mieleitner and Reichert, 2008). The process of anaerobic mineralization was implemented with a specific anaerobic mineralization rate, \( k_{miner,anae,sed,T_0} \), a bacterial temperature dependency, \( \beta_{BAC} \), inhibition terms for nitrate and oxygen, and a dependence on the concentration of organic particles, \( X_S \):

\[
\begin{align*}
    r_{miner,anae,sed} &= k_{miner,anae,sed,T_0} \cdot e^{\beta_{BAC} \cdot (T-T_0)} \\
                         &\cdot \left( 1 - \frac{S_{NO_3}}{K_{NO_3,miner} + S_{NO_3}} \right) \cdot \left( 1 - \frac{S_{O_2}}{K_{O_2,miner} + S_{O_2}} \right) \cdot X_S.
\end{align*}
\] (3.19)

This process combines anaerobic mineralization by reduction of manganese oxide \((MnO_2 \text{ to } Mn^{2+})\), iron hydroxide \((FeOOH \text{ to } Fe^{2+})\), sulfate \((SO_2^{2-} \text{ to } HS^-)\) and methanogenesis \((C_{org} \text{ to } CH_4)\). It was assumed that at least one of these processes occurs in the absence of oxygen and nitrate and that there is no limitation due to a lack of reducible substances.

The re-oxidation of the reduced substances \(Mn^{2+}, Fe^{2+}, HS^-\) and \(CH_4\) that are released into the hypolimnion after anaerobic mineralization in the sediment and the subsequent depletion of oxygen needed for this oxidation, was not considered. This oxidation can have a significant effect on the oxygen balance in lakes. As there are no data available for any of the four reduced substances, however we cannot compare any modeled concentrations with data. Furthermore, the stoichiometry of all four oxidation processes is the same with respect to the amount of oxygen needed for oxidation per mole of mineralized organic carbon in the corresponding reduction process. For this reason, we can describe anaerobic mineralization and subsequent oxidation of the reduced compound as methanogenesis. Therefore we explicitly considered the production of methane \((CH_4)\) during anaerobic mineralization and introduced oxidation of methane as a new process.
3.2. Methods

The process rate of re-oxidation of methane in the hypolimnion (after anaerobic reduction of particulate organic carbon and diffusion of methane into the hypolimnion) is described as

$$ r_{\text{oxid,CH}_4} = k_{\text{oxid,CH}_4} \cdot S_{\text{CH}_4} \cdot S_{\text{O}_2} $$ (3.20)

with an oxidation rate constant, $k_{\text{oxid,CH}_4}$. For the oxidation of 1 gC m$^{-3}$ methane 5.33 gO m$^{-3}$ oxygen is needed. The concentration of methane increases with a total net transformation rate dependent on the actual anaerobic mineralization rate referred to equation (3.19). According to the stoichiometry of the process of methanogenesis (Reichert and Schuwirth, 2010), resulting from Redfield composition, per 1 gDM m$^{-3}$ organic matter, $1/2 \cdot a_C$ gC m$^{-3}$ methane is produced, where $a_C$ is the fraction of carbon ($a_C$ equals 0.358 gC gDM$^{-1}$ as for Redfield composition) in degradable particulate organic matter ($X_S$). Diffusion of methane between all boxes and gas exchange with the atmosphere are considered as well. According to Cussler, 2009, molecular diffusivity of methane was set to 0.000129 m$^2$ d$^{-1}$. By a comparison with the similar diffusivity of oxygen, the methane exchange velocity was set to the same as the oxygen exchange velocity, according to Schwarzenbach et al., 2003.

**Inert fraction of organic particles** Sobek et al., 2009 found that the negative correlation of organic carbon burial efficiency with oxygen exposure time is stronger for lakes with significant allochthonous input, for example from land erosion. It follows that the rate of permanent sedimentation of organic carbon is greater for allochthonous input than for autochthonous material. This leads to the implication that allochthonous organic particles are less degradable than autochthonous ones. As possible reasons, Sobek et al., 2009 implicated the greater age of allochthonous material, its higher degree of degradation and transformation and its richness in molecules resistant to anaerobic mineralization. These findings are supported by the fact that the phosphate concentration of Walensee is decreasing in the long term, which is not true of the inflow of organic particles, which are the main source of annual phosphorus load. We concluded that a large fraction of these particles are only slowly degradable, because they mainly originate from allochthonous input discharged by the river Linth. Hence, the fraction of slowly degradable organic particles within allochthonous input, $f_{X_I,\text{rivers}}$, was assumed to be large (calibration resulted in a value of 0.93), whereas that of inputs from lakes (Lake Zurich, upper basin and Lake Pfäffikon), $f_{X_I,\text{lake}}$, was set to 0.2, close to the fraction of the cadavers of organisms that turns into slowly degradable material in the death process, $f_p$, which has the value 0.1. In the model, the slowly degradable particles are assumed not to be degraded within the observed time span and hence were considered as inert.
Sorption of inorganic phosphorus to inert particles  The higher inert fraction of organic particles that reaches Greifensee and Walensee could not solely explain the marked decrease in phosphate concentration in Walensee over the past few decades. We assumed that inorganic phosphorus, in addition to the sorption to degradable organic particles introduced by Omlin et al., [2001b] as the state variable $X_{PI,S}$, also adsorbs to slowly degradable organic particles while those are settling down to the sediment. Another state variable, $X_{PI,I}$, was included into the model which represents the concentration of inorganic phosphorus adsorbed to organic particles that are considered as inert and represented by the state variable $X_I$. The phosphorus uptake rate of inert organic particles follows the subsequent equation:

$$r_{uptake,X_I} = k_{upt,X_I} \cdot (a_{P,max,X_I} - a_{PI,I}) \cdot \frac{S_{O_2}}{K_{O_2,ads} + S_{O_2}} \cdot S_{HPO_4} \cdot X_I.$$  \hspace{1cm} (3.21)

In this equation, $k_{upt,X_I}$ represents the phosphorus uptake rate constant of $X_I$. The uptake rate includes a Monod-type limitation to the concentration of oxygen, $S_{O_2}$. It depends on the concentration of inert organic particles, $X_I$, the concentration of dissolved phosphate, $S_{HPO_4}$, and the difference between the actual fraction of inorganic phosphorus on inert organic particles, $a_{PI,I}$, and the maximal fraction, $a_{P,max,X_I}$. This maximal fraction was set to 0.001 gP gDM$^{-1}$, three times smaller than the maximal phosphorus fraction of degradable organic particles, $a_{P,max,X_S}$.

Additionally to these model extensions, we also considered the possibility of internal nutrient storage capacity of plankton as potential reason for observed lake nutrient concentrations lower than predicted by the model. We studied the Droop model (Droop, [1973]) accounting for internal storages as a possible alternative to the Monod model. It was concluded that the two model equations are equivalent (Burmaster, [1979]) for time steps larger than several days. Hence, in the considered time steps given by the monthly averaged driving forces, these processes do not play a role and the Monod model seems to be appropriate.

Table 3.1 summarizes all new parameters and those that have a new meaning compared to the most recent published version of the lake model (Omlin et al., [2001b]; Mieleitner and Reichert, [2008]) and their actual values. For the description of parameters that changed their values due to automatic calibration, see Table 3.4. For their assumed prior distributions, see Table 3.2, where the parameters of the new error model are also listed.
Table 3.1: Parameters that were newly introduced into the model or have different value compared to the most recently published version of BELAMO (in alphabetical order). Lake-specific parameters adjusted to temperature measurements (WVZ, AWEL Zürich) are indicated by "***".

<table>
<thead>
<tr>
<th>Name (Unit)</th>
<th>Value</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_{P,\text{inflow,lake}}$</td>
<td>0.0056</td>
<td>Mass fraction of phosphorus in organic material from lakes; &lt; Redfield, accounting for phosphate limitation of phytopl. in lakes</td>
</tr>
<tr>
<td>$a_{P,\text{max},X_{I}}$</td>
<td>0.001</td>
<td>Maximum mass fraction of phosphate adsorbed to inert org. mat.</td>
</tr>
<tr>
<td>$a_{P,\text{max},X_{S}}$</td>
<td>0.004</td>
<td>Maximum mass fraction of phosphate adsorbed to degradable org. mat.</td>
</tr>
<tr>
<td>$b_{P,\text{min}}$</td>
<td>0.0025</td>
<td>Minimum phosphorus content of newly produced phytoplankton</td>
</tr>
<tr>
<td>$D_{\text{CH}_4}$ (m$^2$d$^{-1}$)</td>
<td>0.000129</td>
<td>Molecular diffusivity of methane</td>
</tr>
<tr>
<td>$\mathcal{J}_{\text{sol}}$</td>
<td>0.5</td>
<td>Dissolved fraction of phosphorus from excretion and sloppy feeding</td>
</tr>
<tr>
<td>$h_{\text{epi}}$ (m (gre/wal/zh))</td>
<td>5/10/10</td>
<td>Thickness of epilimnion (hypolimnion changed accordingly)</td>
</tr>
<tr>
<td>$K_{\text{ALG,ZOO}}$ (gDM m$^{-3}$)</td>
<td>0.8</td>
<td>Half-saturation phytoplankton concentration for zooplankton growth</td>
</tr>
<tr>
<td>$K_{\text{Feed}}$ (gDM m$^{-3}$)</td>
<td>0.945</td>
<td>Threshold phytopl. conc. when zoopl. feeding switches to Monod limit.</td>
</tr>
<tr>
<td>$K_{\text{Feed,ZOO,Dipt}}$ (gDM m$^{-3}$)</td>
<td>0.15</td>
<td>Threshold zoopl. conc. when diptera feed. switches to lin. dependence</td>
</tr>
<tr>
<td>$K_{\text{NO}_3,\text{miner}}$ (gN m$^{-3}$)</td>
<td>0.02</td>
<td>Half-saturation nitrate concentration for anoxic mineralization; Leads to a sharper transition between anaerobic and anaerobic miner.</td>
</tr>
<tr>
<td>$K_{\text{O}_2,\text{miner}}$ (gO m$^{-3}$)</td>
<td>0.1</td>
<td>Half-saturation oxygen concentration for aerobic mineralization; Leads to a sharper transition between aerobic and anaerobic miner.</td>
</tr>
<tr>
<td>$k_{\text{oxid,CH}_4}$ (m$^3$gO$^{-1}$d$^{-1}$)</td>
<td>0.04</td>
<td>Reaction velocity of methane oxidation</td>
</tr>
<tr>
<td>$k_{\text{resp,ZOO}}$ (d$^{-1}$)</td>
<td>0.01</td>
<td>Respiration rate of zoopl.; more realistic including Planktothrix rubescens</td>
</tr>
<tr>
<td>$K_{\text{upt},X_{I}}$ (m$^3$gDM$^{-1}$d$^{-1}$)</td>
<td>30</td>
<td>Maximum uptake rate of phosphate uptake on inert particles</td>
</tr>
<tr>
<td>$K_{\text{summer}}/K_{\text{winter}}$ (m$^2$d$^{-1}$)</td>
<td>*</td>
<td>Coefficient of vertical turbulent diffusion in the metalimnion</td>
</tr>
<tr>
<td>$P_{N_H_4,\text{ALG}}$</td>
<td>10</td>
<td>Preference factor of phytoplankton growth on ammonium</td>
</tr>
<tr>
<td>$\text{tempdif}$ (°C)</td>
<td>*</td>
<td>Temperature difference between epi- and hypolimnion; Switching between summer and winter mixing</td>
</tr>
<tr>
<td>$P_{\text{CH}_4,\text{atm}}$ (m d$^{-1}$) (gre/wal/zh)</td>
<td>0.5/3/1</td>
<td>Methane exchange velocity with atmosphere, according to oxygen</td>
</tr>
<tr>
<td>$P_{\text{O}_2,\text{atm}}$ (m d$^{-1}$) (gre/wal/zh)</td>
<td>0.5/3/1</td>
<td>Oxygen exchange velocity with atm., differing by specific wind speed</td>
</tr>
<tr>
<td>$P_{\text{sed,ALG}}$ (m d$^{-1}$)</td>
<td>0.1</td>
<td>Sedimentation velocity of phytoplankton; better approximation for aggregated plankton model</td>
</tr>
<tr>
<td>$V_{u_{P,ZOO}}$ (m d$^{-1}$)</td>
<td>5</td>
<td>Upwards velocity of zooplankton</td>
</tr>
</tbody>
</table>

### 3.2.3 Data

Monthly measured profiles of physical, chemical and biological variables for Lake Zurich and Walensee were obtained from the Water Supply Authority of Zurich (Wasserversorgung Zürich (WVZ)) from 1976 to the spring of 2006. For Greifensee, monthly to daily measurements of physical, chemical and biological variables were obtained from the Aquatic Ecology Department of Eawag (Swiss Federal Institute of Aquatic Science and Technology) and the environmental agency of the canton of Zürich (Amt für Abfall, Wasser, Energie und Luft der Baudirektion des Kantons Zürich (AWEL Zürich)). For chemical and physical variables for Greifensee, data were collected from 1985-2006.
Phytoplankton data of Greifensee were available from 1987-2004, while zooplankton data were available for the years 1987 to 2006. Information on inflows to the lakes (physical and chemical variables) and meteorological data were received directly from federal (Bundesamt für Umwelt (BAFU)) and cantonal agencies (Amt für Umweltschutz des Kantons St. Gallen (AFU St. Gallen) and AWEL Zürich) and from WVZ. Technical reports from these institutions provided insights into sampling techniques and methodological details (Gammeter et al., 1996; Gammeter et al., 1997; Gammeter and Forster, 2002). A detailed description of the data availability, compilation and processing can be found in the Appendix B.1 to this paper.

Figure 3.4: Input loads of bioavailable nitrogen and phosphorus per lake surface area. Dashed, dotted and dash-dotted lines indicate measured input data, bold lines their moving averages over 12 months.

A comparison of the nutrient input loads (bioavailable nitrogen and phosphorus) per lake surface area from inflowing rivers, waste water treatment plants (WWTPs) and wet deposition is shown in Fig. 3.4. In our model formulation, bioavailable nitrogen and phosphorus represent the sum of the dissolved compounds (nitrate, ammonium and phosphate, respectively) and the N- and P-fraction of the non-inert part \(1 - f_{X_1,rivers}\) or \(1 - f_{X_1,lake}\) of the inputs of organic material. It became obvious that for all three lakes the input load of bioavailable phosphorus decreased considerably during the measurement period. This is particularly true of Lake Zurich and Walensee and for the period 1976 to 1987. This is due to the introduction of phosphate-free detergents, restrictions in the use of phosphate fertilizers and the connection of more households to WWTPs. The introduction of phosphate precipitation in WWTPs and the ultimate prohibition of detergents containing phosphate in 1986 also resulted in a decline. Exceptions were the years around 2001, when some WWTPs exported higher phosphate loads to the
3.2. Methods

Over the same time period the mean nitrogen load did not change significantly. A comparison of the inputs to the three lakes is broadly consistent with their trophic status. For most years, the bioavailable phosphorus and nitrogen loads per surface area of Greifensee were higher than for Lake Zurich and Walensee, for which the loads were similar. After the significant decrease in phosphorus input into Walensee, the phosphorus load to Lake Zurich exceeded that for the Walensee in most years.

3.2.4 Sensitivity analysis

As stated by Mieleitner and Reichert, 2006, the calibration of complex models with a large number of parameters is done best by an iterative procedure of sensitivity analysis, parameter estimation and identifiability analysis. A sensitivity analysis measures how strongly parameter changes influence the model output. A preliminary sensitivity analysis can help to identify potential parameters that must be estimated. After estimation, identifiability analysis can be used to check for problems in producing a unique estimate. The set of parameters to be estimated can then be updated according to a new sensitivity analysis. Finally, a posterior sensitivity analysis can serve as a reassessment of the subset of fitted parameters and model properties.

We conducted sensitivity analyses for the model results of the three lakes separately, as well as jointly for all lakes. The analyses were done according to Brun et al., 2001 and Omlin et al., 2001a as local sensitivity analyses using linear error propagation as a measure of the contribution of each parameter to the model output error. The overall sensitivity measure $\delta_{\theta_j}^{msqr}$ is calculated as

$$
\delta_{\theta_j}^{msqr}(\theta) = \sqrt{\frac{1}{n_L} \sum_{L} \left( \frac{\Delta \theta_j}{s_{c_y_{LM}}} \cdot \frac{\partial y_{LM}^L(x, \theta)}{\partial \theta_j} \right)^2}
$$

(3.22)

using the square root of the average of the squared derivatives of all model outputs, $y_{LM}^L$, of the layout L (of length $n_L$) with respect to the parameter $\theta_j$ times the ratio of the uncertainty range, $\Delta \theta_j$, of the parameter $\theta_j$ and the scaling factor, $s_{c_y_{LM}}$. The layout L characterizes the model output, i.e. which output variables are considered at which point of the output dimension (e.g. time/space). In our case the layout consists of combinations of the five output variables phytoplankton, zooplankton, oxygen, nitrate and phosphate, a (monthly) date and a specification of the compartment (epilimnion or hypolimnion). The chosen layout mostly equals the description of available measurements.
For local sensitivity analyses, the results differ for each combination of parameter values, $\theta$. We will show the results of the posterior sensitivity analysis conducted for the best parameter estimate found during model calibration in the next section. The model results for this point in the parameter space served as scaling factor, $s_{Cy_{e}}$, to get non-dimensional sensitivity measures and to make the influences of parameters on different output variables with different units comparable to each other. The uncertainty of the parameters was chosen according to [Omlin et al., 2001a] by classifying all parameters into three groups of relative uncertainty (accurately known parameters = class 1 = 5%; stoichiometric parameters/specific growth rates = class 2 = 20%; most kinetic/poorly known parameters = class 3 = 50%). We did not group the parameters into those able to be fitted or not in order to compare the sensitivity of all parameters. Nevertheless, this distinction influenced our choice of calibration parameters. We also analyzed the sensitivity of the model to the parameters $a_{P_{inflow,lake}}$, $f_{X_{I_{rivers}}}$ and $f_{X_{I_{lakes}}}$ which are related to input fluxes, but are new or have a new meaning due to model extensions. Beside those, we did not include input fluxes in the sensitivity analysis, as inflow parameters did not seem to be meaningful to calibrate and to distinguish these parameters from model parameters.

### 3.2.5 Model calibration

Manual calibration of simulation models by experts in the presence of bias in model outputs has often been judged superior to automatic calibration, as experts take a more “holistic” view of the quality of the fit and emphasize certain characteristic patterns rather than merely minimizing the average deviation of model results from observations ([Boyle et al., 2000]). However, for computationally demanding models with high-dimensional parameter and output spaces, manual calibration is almost impossible and a more systematic procedure is required. As it is difficult to formalize and automate the weighting process of multiple criteria carried out by experts during manual calibration, the use of multi-objective optimization techniques has been suggested to determine a Pareto-set of good solutions rather than trying to find the optimal one ([Gupta et al., 1998; Yapo et al., 1998; Madsen, 2000; Madsen et al., 2002; Gupta et al., 2003a; Boyle et al., 2003]). [Reichert and Schuwirth, 2012] recently adapted a statistical bias description technique ([Craig et al., 1996; Craig et al., 2001; Kennedy and O’Hagan, 2001; Higdon et al., 2004; Bayarri et al., 2007]) to provide a statistical basis for an estimation procedure that attempts to imitate and systematize expert calibration while considering systematic model errors, a particularly typical phenomenon in environmental modeling. This technique is based on prior specification of the magnitude of “acceptable bias” in
different model variables to take into account the expert’s choice of relative importance of fit objectives. An approximation to the technique of [15] for computationally demanding models has been developed by [16] and was applied in this paper.

This technique is based on describing system observations as the sum of the output of the deterministic model, bias (systematic errors) and observation error to account explicitly for the contribution of bias in model results:

\[
Y^L_M(x, \theta, \psi, \xi) = y^L_M(x, \theta) + B^L_M(x, \xi) + E^L(\psi),
\]

where \( Y^L_M \) is the vector of random variables representing the observations as described by the model, \( M \), at the layout, \( L \), that defines the output variables and the time points and locations at which they are observed or evaluated. \( Y^L_M \) depends on the external influence factors, \( \xi \), unknown model parameters, \( \theta \), and additional parameters, \( \psi \) and \( \xi \), of the error terms. It is composed of the deterministic function \( y^L_M(x, \theta) \), representing our knowledge of the system response (i.e. the output of our model), of the random process \( B^L_M(x, \xi) \), expressing our knowledge of model bias depending on external inputs and additional parameters and of \( E^L(\psi) \), a vector of random variables that represents the observation error, which may also depend on additional parameters. Note that when propagating the appropriate parameter distribution (either prior or posterior, depending on which output distribution we are interested in) the representation (3.23) allows us to distinguish the random variables representing our knowledge of the true state of the system, \( y^L_M + B^L_M \), from the random variable representing observations, \( Y^L_M \), including observation errors.

Under normality assumptions for both the bias and the observation error, and after integrating out the bias, we obtain the likelihood function:

\[
f_{y^L_M|\Theta, \Psi, \Xi}(y^L | \theta, \psi, \xi, x) = \frac{1}{\sqrt{2\pi}^n} \frac{1}{\sqrt{\det(\Sigma_E + \Sigma_B^L)}} \cdot \exp \left( -\frac{1}{2} [y^L - y^L_M(x, \theta)]^T (\Sigma_E + \Sigma_B^L)^{-1} [y^L - y^L_M(x, \theta)] \right),
\]

where \( \Sigma_E^L \) and \( \Sigma_B^L \) are the variance-covariance matrices of observation error and bias, respectively. The posterior of the parameters is then given by:

\[
f_{\Theta, \Psi, \Xi|y^L_M}(\theta, \psi, \xi | y^L_M, x) \propto \frac{f_{\Theta, \Psi, \Xi}(\theta, \psi, \xi)}{\sqrt{\det(\Sigma_E + \Sigma_B^L)}} \cdot \exp \left( -\frac{1}{2} [y^L_M - y^L_M(x, \theta)]^T (\Sigma_E^L + \Sigma_B^L)^{-1} [y^L_M - y^L_M(x, \theta)] \right),
\]

where \( \Sigma_E^L \) and \( \Sigma_B^L \) are the variance-covariance matrices of observation error and bias, respectively. The posterior of the parameters is then given by:

\[
f_{\Theta, \Psi, \Xi|y^L_M}(\theta, \psi, \xi | y^L_M, x) \propto \frac{f_{\Theta, \Psi, \Xi}(\theta, \psi, \xi)}{\sqrt{\det(\Sigma_E + \Sigma_B^L)}} \cdot \exp \left( -\frac{1}{2} [y^L_M - y^L_M(x, \theta)]^T (\Sigma_E^L + \Sigma_B^L)^{-1} [y^L_M - y^L_M(x, \theta)] \right),
\]
where \( f_{\Theta,\Psi,\Xi}(\theta,\psi,\xi) \) represents the prior distribution of the parameters \( \theta, \psi \) and \( \xi \) (see Reichert and Schuwirth, 2012 for more details). This posterior is not much more difficult to maximize than another objective function used in other calibration techniques (there are a few additional parameters in \( \Sigma_{\mathbf{B}_{\mathbf{L}}^{M}} \)). As described in Dietzel and Reichert, 2012, we use a Box-Cox transformation (Box and Cox, 1964) of model results and data (with \( \lambda_1 = 0.5 \) and \( \lambda_2 = 0 \)) to account for heteroscedasticity, we derive a normal approximation to the posterior and we use linearized error propagation for deriving the posteriors of parameters and model results.

After identifying major calibration problems by manual calibration and improving the model as described above (section 2.2.2) to reduce these problems, this technique was applied to solve the calibration problem and to estimate uncertainties in model output due to parameter uncertainty, structural uncertainty and observation error. The identifiability problem between deterministic model output and bias was addressed by specifying a prior with a mean of zero for the bias and a prior for its standard deviation which favors a small value. This prior was parameterized by an exponential distribution with different means for different observed substances or organisms. This formalizes the preference for a small bias, but allows different magnitudes of bias for different state variables. For the measurement error, a lognormal prior distribution for its standard deviation was chosen, representing our uncertainty about the amount of observation error in each output variable. The prior distributions of all calibrated model parameters, as well as the parameters of the error model, are summarized in Table 3.2. The modes of the distributions were fixed approximately at the previously assumed model parameter value. The correlation length of the bias was fixed to a little under 2 months. Due to the use of Box-Cox transformation, the parameters of the error model are expressed in transformed units (as shown in Table 3.2). The parameter \( f_{X_1,\text{rivers}} \), referring to a fraction, was transformed by an arctan function to keep it within the interval \([0,1]\). In general, most influential and least known parameters were chosen to be calibrated, while trying to keep the calibration and identifiability problem as small as possible. The mixing parameters were not included in the automatic calibration, because they were already manually calibrated to the aggregated values of the temperatures in epi- and hypolimnion. The respiration rate constants were not included, as it had been found previously that growth and respiration rate constants show a large collinearity and are hence non-identifiable when calibrated jointly (Omlin et al., 2001a). The very influential parameter \( K_{\text{Feed}} \) was used for calibration, whereas the parameter \( K_{I,\text{ALG}} \) was excluded, because it was found to be non-identifiable in earlier studies (Mieleitner and Reichert, 2006). Stoichiometric parameters were not included either, as it was assumed that prior knowledge was the more reliable source of information than the fit of a lake model that
### Table 3.2: Description and prior marginals of the calibrated parameters of the deterministic model and the error model.

<table>
<thead>
<tr>
<th>Name</th>
<th>Unit</th>
<th>Distribution</th>
<th>Mean</th>
<th>StDev</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$f_{X_{1, \text{rivers}}}$(transformed)</td>
<td></td>
<td>Normal</td>
<td>2</td>
<td>1</td>
<td>Inert fraction of allochthonous organic particles</td>
</tr>
<tr>
<td>$k_{\text{death, ALG}, T_{0}}$</td>
<td>d$^{-1}$</td>
<td>Lognormal</td>
<td>0.0721</td>
<td>0.05</td>
<td>Death rate of phytoplankton at reference temperature</td>
</tr>
<tr>
<td>$k_{\text{death, ZOO, T_{0, gre}}}$</td>
<td>d$^{-1}$</td>
<td>Lognormal</td>
<td>0.0675</td>
<td>0.05</td>
<td>Death rate of zooplankton at reference temperature (Greifensee)</td>
</tr>
<tr>
<td>$k_{\text{death, ZOO, T_{0, wal}}}$</td>
<td>d$^{-1}$</td>
<td>Lognormal</td>
<td>0.0432</td>
<td>0.05</td>
<td>Death rate of zooplankton at reference temperature (Walensee)</td>
</tr>
<tr>
<td>$k_{\text{death, ZOO, T_{0, zh}}}$</td>
<td>d$^{-1}$</td>
<td>Lognormal</td>
<td>0.0675</td>
<td>0.05</td>
<td>Death rate of zooplankton at reference temperature (Lake Zurich)</td>
</tr>
<tr>
<td>$K_{\text{Feed}}$</td>
<td>gDM m$^{-3}$</td>
<td>Lognormal</td>
<td>0.295</td>
<td>0.3</td>
<td>Threshold phytopl. conc. when zoopl. feeding switches to Monod limit.</td>
</tr>
<tr>
<td>$k_{\text{gro, ALG, T_{0, l0}}}$</td>
<td>d$^{-1}$</td>
<td>Lognormal</td>
<td>1.79</td>
<td>0.5</td>
<td>Growth rate of phytoplankton at reference temp./sat. light intensity</td>
</tr>
<tr>
<td>$k_{\text{gro, ZOO, T}_{0}}$</td>
<td>d$^{-1}$</td>
<td>Lognormal</td>
<td>0.397</td>
<td>0.25</td>
<td>Growth rate of zooplankton at reference temperature</td>
</tr>
<tr>
<td>$k_{\text{miner,aero, sed, T_{0, gre}}}$</td>
<td>d$^{-1}$</td>
<td>Lognormal</td>
<td>0.365</td>
<td>0.25</td>
<td>Aerobic mineralization rate at reference temperature (Greifensee)</td>
</tr>
<tr>
<td>$k_{\text{miner,aero, sed, T_{0, wal}}}$</td>
<td>d$^{-1}$</td>
<td>Lognormal</td>
<td>0.063</td>
<td>0.05</td>
<td>Aerobic mineralization rate at reference temperature (Walensee)</td>
</tr>
<tr>
<td>$k_{\text{miner,anae, sed, T_{0, gre}}}$</td>
<td>d$^{-1}$</td>
<td>Lognormal</td>
<td>0.295</td>
<td>0.3</td>
<td>Aerobic mineralization rate at reference temperature (Lake Zurich)</td>
</tr>
<tr>
<td>$k_{\text{miner,anae, sed, T_{0, wal}}}$</td>
<td>d$^{-1}$</td>
<td>Lognormal</td>
<td>0.365</td>
<td>0.25</td>
<td>Anaerobic mineralization rate at reference temperature (Greifensee)</td>
</tr>
<tr>
<td>$k_{\text{miner,anae, sed, T_{0, zh}}}$</td>
<td>d$^{-1}$</td>
<td>Lognormal</td>
<td>0.295</td>
<td>0.3</td>
<td>Anaerobic mineralization rate at reference temperature (Walensee)</td>
</tr>
<tr>
<td>$k_{\text{miner,anae, sed, T_{0, wall}}}$</td>
<td>d$^{-1}$</td>
<td>Lognormal</td>
<td>0.365</td>
<td>0.25</td>
<td>Anaerobic mineralization rate at reference temperature (Lake Zurich)</td>
</tr>
<tr>
<td>$k_{\text{miner,anae, sed, T_{0, zh, wall}}}$</td>
<td>d$^{-1}$</td>
<td>Lognormal</td>
<td>0.063</td>
<td>0.05</td>
<td>Anaerobic mineralization rate at reference temperature (Walensee)</td>
</tr>
<tr>
<td>$\sigma_{B, \text{ALG}}$</td>
<td>gWM$^{1/2}$ m$^{-3/2}$</td>
<td>Exponential</td>
<td>0.295</td>
<td>0.3</td>
<td>Anaerobic mineralization rate at reference temperature (Lake Zurich)</td>
</tr>
<tr>
<td>$\sigma_{B, \text{HPO4}}$</td>
<td>gP$^{1/2}$ m$^{-3/2}$</td>
<td>Exponential</td>
<td>0.035</td>
<td></td>
<td>Standard deviation of bias in phytoplankton</td>
</tr>
<tr>
<td>$\sigma_{B, \text{NO3}}$</td>
<td>gN$^{1/2}$ m$^{-3/2}$</td>
<td>Exponential</td>
<td>0.05</td>
<td></td>
<td>Standard deviation of bias in phosphate</td>
</tr>
<tr>
<td>$\sigma_{B, \text{O2}}$</td>
<td>gO$^{1/2}$ m$^{-3/2}$</td>
<td>Exponential</td>
<td>0.5</td>
<td></td>
<td>Standard deviation of bias in oxygen</td>
</tr>
<tr>
<td>$\sigma_{B, \text{ZOO}}$</td>
<td>gWM$^{1/2}$ m$^{-3/2}$</td>
<td>Exponential</td>
<td>0.4</td>
<td></td>
<td>Standard deviation of bias in zooplankton</td>
</tr>
<tr>
<td>$\sigma_{E, \text{ALG}}$</td>
<td>gWM$^{1/2}$ m$^{-3/2}$</td>
<td>Lognormal</td>
<td>0.07</td>
<td>0.02</td>
<td>Standard deviation of measurement error in phytoplankton</td>
</tr>
<tr>
<td>$\sigma_{E, \text{HPO4}}$</td>
<td>gP$^{1/2}$ m$^{-3/2}$</td>
<td>Lognormal</td>
<td>0.007</td>
<td>0.0002</td>
<td>Standard deviation of measurement error in phosphate</td>
</tr>
<tr>
<td>$\sigma_{E, \text{NO3}}$</td>
<td>gN$^{1/2}$ m$^{-3/2}$</td>
<td>Lognormal</td>
<td>0.022</td>
<td>0.002</td>
<td>Standard deviation of measurement error in nitrate</td>
</tr>
<tr>
<td>$\sigma_{E, \text{O2}}$</td>
<td>gO$^{1/2}$ m$^{-3/2}$</td>
<td>Lognormal</td>
<td>0.11</td>
<td>0.01</td>
<td>Standard deviation of measurement error in oxygen</td>
</tr>
<tr>
<td>$\sigma_{E, \text{ZOO}}$</td>
<td>gWM$^{1/2}$ m$^{-3/2}$</td>
<td>Lognormal</td>
<td>0.08</td>
<td>0.02</td>
<td>Standard deviation of measurement error in zooplankton</td>
</tr>
</tbody>
</table>
could lead to unrealistic values. A normal approximation to the posterior was calculated by first estimating the mean and covariance matrix by importance sampling from a uniform ball around the maximum, followed by importance sampling using the normal distribution with this mean and covariance matrix as a new sampling distribution. The mean and covariance matrix of this second importance sampling are unbiased estimates of mean and covariance matrix of the posterior and were used to characterize the approximating normal distribution. This technique has proved to be the most robust for addressing the challenging problem of approximating a potentially complex shape of the posterior by a normal distribution in Dietzel and Reichert, 2012.

Parameter estimation was done for 9 years for Greifensee (1987-1995) and 20 years for Lake Zurich and Walensee (1976-1995). To assess the predictive capability of the model, simulations were done also for the following 10 years without re-calibration.

3.2.6 Model implementation

As previously, the model was implemented using the computer program AQUASIM for simulation and identification of aquatic systems [Reichert, 1994](http://www.aquasim.eawag.ch). Initial model calibration was done manually; sensitivity analyses and calibrations in later stages were performed by coupling AQUASIM to UNCSIM, a tool for statistical inference and sensitivity, identifiability and uncertainty analysis with arbitrary simulation programs [Reichert, 2005](http://www.r-project.org/).

3.3 Results

3.3.1 Sensitivity analysis

Table 3.3 shows the sensitivity ranking of all model parameters including the input parameters $a_{P,inflow,lake}$, $f_{X_i,rivers}$ and $f_{X_i,lake}$. It reflects, in linear approximation, the amount of influence of each parameter, $\theta_j$, within its uncertainty range, $\Delta \theta_j$, (see above section 3.2.4 for the choice of the uncertainty ranges of the parameters) on all the model results of the five output variables phytoplankton, zooplankton, oxygen, nitrate and phosphate in the two water compartments of the three lakes. Larger values of $\delta_{msqr}^{\theta_j}$ (equation (3.22)) indicate a larger change in the model results due to a change in the respective parameter. An explanation of the parameter names can be found in Tables
3.3. Results

3.1 and 3.2, as well as in Omlin et al., 2001b and Mieleitner and Reichert, 2006.

Table 3.3: Results of the joint sensitivity analysis for all three lakes

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$\delta_{j}\hat{y}_i$</th>
<th>Parameter</th>
<th>$\delta_{j}\hat{y}_i$</th>
<th>Parameter</th>
<th>$\delta_{j}\hat{y}_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_{z,summer,gre}$</td>
<td>10.520</td>
<td>$h_{sed_1, zh}$</td>
<td>0.888</td>
<td>$K_{z,summer,wal}$</td>
<td>0.116</td>
</tr>
<tr>
<td>$K_{Feed}$</td>
<td>9.562</td>
<td>$h_{sed_1, gre}$</td>
<td>0.774</td>
<td>$K_{z,summer,gre}$</td>
<td>0.105</td>
</tr>
<tr>
<td>$K_{ Feed,ALG,T_0,T_0}$</td>
<td>7.967</td>
<td>$h_{sed_2, gre}$</td>
<td>0.756</td>
<td>$K_{z,summer,wal}$</td>
<td>0.104</td>
</tr>
<tr>
<td>$K_{z,winter, zh}$</td>
<td>7.024</td>
<td>$b_{P,min}$</td>
<td>0.707</td>
<td>$DO_2$</td>
<td>0.101</td>
</tr>
<tr>
<td>$w_{ZOO}$</td>
<td>5.701</td>
<td>$a_{P,max,X_I}$</td>
<td>0.625</td>
<td>$k_{miner, anox, sed, T_0, wal}$</td>
<td>0.093</td>
</tr>
<tr>
<td>$K_{1,ALG}$</td>
<td>4.450</td>
<td>$\nu_{sed, ORG}$</td>
<td>0.571</td>
<td>$k_{upt,X_2, wal}$</td>
<td>0.066</td>
</tr>
<tr>
<td>$K_{z,summer, zh}$</td>
<td>4.236</td>
<td>$c_c$</td>
<td>0.548</td>
<td>$\nu_{O_2, atm, zh}$</td>
<td>0.062</td>
</tr>
<tr>
<td>$k_{death, ZOO, T_0, zh}$</td>
<td>4.159</td>
<td>$f_{X_I, rivers}$</td>
<td>0.544</td>
<td>$k_{miner, anox, sed, T_0, zh}$</td>
<td>0.057</td>
</tr>
<tr>
<td>tempdif$_{zh}$</td>
<td>4.059</td>
<td>$a_{P, inflow}$</td>
<td>0.463</td>
<td>$R_{NO_3, miner}$</td>
<td>0.048</td>
</tr>
<tr>
<td>$K_{z,summer, ALG, T_0}$</td>
<td>3.891</td>
<td>$w_{ORG}$</td>
<td>0.427</td>
<td>$k_{nitri, w_0}$</td>
<td>0.044</td>
</tr>
<tr>
<td>$K_{z, summer, ALG}$</td>
<td>3.747</td>
<td>$k_{miner, aero, sed, T_0, w_0}$</td>
<td>0.360</td>
<td>$K_{O_2, atm, w_0}$</td>
<td>0.043</td>
</tr>
<tr>
<td>$K_{ALG, ZOO}$</td>
<td>3.725</td>
<td>$k_{upt,X_2, gre}$</td>
<td>0.350</td>
<td>$k_{death, ZOO, T_0, w_0}$</td>
<td>0.040</td>
</tr>
<tr>
<td>$b_{P, max}$</td>
<td>3.541</td>
<td>$h_{sed_1, wal}$</td>
<td>0.287</td>
<td>$k_{miner, aero, w_0, T_0}$</td>
<td>0.035</td>
</tr>
<tr>
<td>$k_{resp, ALG, T_0}$</td>
<td>3.210</td>
<td>$\beta_{BAC}$</td>
<td>0.287</td>
<td>$ac$</td>
<td>0.033</td>
</tr>
<tr>
<td>$f_{sol}$</td>
<td>3.143</td>
<td>$h_{sed_2, wal}$</td>
<td>0.265</td>
<td>$p_{ALG, NH_4}$</td>
<td>0.032</td>
</tr>
<tr>
<td>$K_{g, summer, T_0}$</td>
<td>2.985</td>
<td>$K_{N, ALG}$</td>
<td>0.2600</td>
<td>$k_{upt, X_1, gre}$</td>
<td>0.031</td>
</tr>
<tr>
<td>$K_{death, ZOO, T_0, w_0}$</td>
<td>2.818</td>
<td>$K_{O_2, resp}$</td>
<td>0.247</td>
<td>$K_{O_2, ads}$</td>
<td>0.024</td>
</tr>
<tr>
<td>$f_p$</td>
<td>2.648</td>
<td>$k_{miner, anox, sed, T_0, gre}$</td>
<td>0.241</td>
<td>$k_{upt, X_1, zh}$</td>
<td>0.022</td>
</tr>
<tr>
<td>$\beta_{ZO0}$</td>
<td>2.534</td>
<td>$k_{upt, X_2, zh}$</td>
<td>0.235</td>
<td>$k_{ozid, CH_4}$</td>
<td>0.022</td>
</tr>
<tr>
<td>tempdif$_{gre}$</td>
<td>2.404</td>
<td>$K_{O_2, miner}$</td>
<td>0.230</td>
<td>$k_{nitri, sed, T_0}$</td>
<td>0.017</td>
</tr>
<tr>
<td>$\beta_{ALG}$</td>
<td>2.151</td>
<td>$K_{z, winter, w_0}$</td>
<td>0.217</td>
<td>$k_{miner, anox, w_0, T_0}$</td>
<td>0.012</td>
</tr>
<tr>
<td>$k_{resp, ZOO, T_0}$</td>
<td>2.100</td>
<td>$k_{miner, aero, sed, T_0, zh}$</td>
<td>0.207</td>
<td>$K_{O_2, nitri}$</td>
<td>0.012</td>
</tr>
<tr>
<td>$K_{HPO_4, ALG}$</td>
<td>1.965</td>
<td>$a_{P, max, X_I}$</td>
<td>0.206</td>
<td>$\nu_{O_2, atm, w_0}$</td>
<td>0.012</td>
</tr>
<tr>
<td>$\nu_{sed, ALG}$</td>
<td>1.614</td>
<td>$k_{miner, anox, sed, T_0, gre}$</td>
<td>0.197</td>
<td>$k_{miner, anox, w_0, T_0}$</td>
<td>0.011</td>
</tr>
<tr>
<td>$f_{porp, part}$</td>
<td>1.597</td>
<td>$k_{miner, aero, sed, T_0, gre}$</td>
<td>0.197</td>
<td>$\nu_{sed, INORG}$</td>
<td>0.009</td>
</tr>
<tr>
<td>$\theta$</td>
<td>1.518</td>
<td>$h_{sed_2, zh}$</td>
<td>0.178</td>
<td>$DO_{CH_4}$</td>
<td>0.009</td>
</tr>
<tr>
<td>$f_{X_I, lake}$</td>
<td>1.472</td>
<td>$a_N$</td>
<td>0.157</td>
<td>$v_{CH_4, atm, zh}$</td>
<td>0.007</td>
</tr>
<tr>
<td>$HPO_4, crit$</td>
<td>1.100</td>
<td>$D_{NO_3}$</td>
<td>0.154</td>
<td>$v_{CH_4, atm, w_0}$</td>
<td>0.004</td>
</tr>
<tr>
<td>tempdif$_{wal}$</td>
<td>1.078</td>
<td>$k_{miner, anox, sed, T_0, zh}$</td>
<td>0.152</td>
<td>$v_{CH_4, atm, gre}$</td>
<td>0.003</td>
</tr>
<tr>
<td>$a_{P, inflow, lake}$</td>
<td>0.902</td>
<td>$k_{miner, anox, sed, T_0, wal}$</td>
<td>0.144</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$w_{ALG}$</td>
<td>0.894</td>
<td>$K_{O_2, ZOO}$</td>
<td>0.137</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

It becomes obvious that the parameters describing the mixing of the three lakes in summer and winter (for example $K_{z,summer, gre}$ and $K_{z, winter, zh}$, as well as the temperature difference at which the mixing switches from winter to summer conditions, tempdif$_{gre}$, tempdif$_{zh}$ and tempdif$_{wal}$) are very influential. This is a meaningful result as those parameters influence the distribution of the concentrations of all output variables between the two lake model compartments and the onset of stratification in spring is a very crucial determinant of the initiation of phytoplankton growth. As mentioned above, these parameters were calibrated manually to the measured temperatures. Also influential were kinetic parameters describing the growth, death and respiration of the phyto- and
zooplankton. The respiration rate coefficients contributed less to model output changes than did the growth and death rates of the same plankton community. The parameters $K_{\text{Feed}}$ and $K_{I,\text{ALG}}$ were strongly influential, as well as some of the stoichiometric parameters, such as the factor converting zooplankton dry mass to wet mass, $w_{\text{ZOO}}$, and the maximum phosphorus content of newly produced phytoplankton, $b_{P,\text{max}}$. The results of the joint sensitivity analysis (Table 3.3) were mostly in agreement with the results of the lake-specific analyses. More information about the lake-specific sensitivity analyses can be found in the Appendix B.2 and Table B.1 therein.

### 3.3.2 Model calibration

In the following, the model outputs for phyto-, zooplankton, nitrate, phosphate and oxygen are compared to data from the three lakes over the whole simulation period (19/30 years). For each output variable we chose to depict the results of the compartment that showed the strongest dynamics (hypolimnion for oxygen, epilimnion for the others). The model results are the medians of the posterior (see below a discussion of uncertainty) calculated by linearized propagation of a normal approximation to the posterior of the parameters as summarized above in section 3.2.5 and described in detail by Dietzel and Reichert, 2012. The posterior of the parameters was calculated based on a calibration period of 9 and 20 years. In general, we tried to calibrate as few parameters as possible, to not increase the identifiability problem, but as many parameters as needed to get reasonable results. Table 3.4 summarizes all parameter values changed due to automatic calibration, with the best estimates of their values representing the maximum of the (non-approximated) posterior distribution. For the remaining parameter values, we refer to Omlin et al., 2001b Mieleitner and Reichert, 2006 Mieleitner and Reichert, 2008 and Table 3.1. Except for the mixing coefficients, the thicknesses of the sediment boxes and sediment mineralization rate constants, as well as lake-specific zooplankton death rates and gas exchange velocities differing by wind speeds, these parameter values were kept the same for all three lakes.

Compared to Mieleitner and Reichert, 2008 some parameter values changed during calibration. The growth rate coefficient of zooplankton took a new meaning (and unit) due to the changes in model formulation (see section 3.2.2). For this reason, the change from $0.4 \text{ gDM}^{-1}\text{m}^3\text{d}^{-1}$ to $1.27 \text{ d}^{-1}$ only reflects the change in model structure. The specific growth rate of phytoplankton decreased slightly from $1.6 \text{ d}^{-1}$ to $1.51 \text{ d}^{-1}$. Except for Greifensee, the death rate coefficients of zooplankton decreased somewhat, while the order of magnitude stayed the same. The lowest death rate coefficient was still
Table 3.4: Estimated values of parameters used for automatic calibration.

<table>
<thead>
<tr>
<th>Name (Unit)</th>
<th>Value</th>
<th>Name (Unit)</th>
<th>Value</th>
<th>Name (Unit)</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$f_{X_1,\text{rivers}}$ (-)</td>
<td>0.932</td>
<td>$k_{\text{miner,aero,sed,}T_0,\text{gre}}$ (d$^{-1}$)</td>
<td>0.161</td>
<td>$g_{B,\text{ALG}}$ (gWM$^{1/2}$ m$^{-3/2}$)</td>
<td>0.622</td>
</tr>
<tr>
<td>$k_{\text{death,ALG},T_0}$ (d$^{-1}$)</td>
<td>0.046</td>
<td>$k_{\text{miner,aero,sed,}T_0,\text{wal}}$ (d$^{-1}$)</td>
<td>0.018</td>
<td>$g_{B,\text{HPO}_4}$ (gP$^{1/2}$ m$^{-3/2}$)</td>
<td>0.126</td>
</tr>
<tr>
<td>$k_{\text{death,ZOO},T_0,\text{gre}}$ (d$^{-1}$)</td>
<td>0.119</td>
<td>$k_{\text{miner,aero,sed,}T_0,\text{zh}}$ (d$^{-1}$)</td>
<td>0.116</td>
<td>$g_{B,\text{NO}_3}$ (gN$^{1/2}$ m$^{-3/2}$)</td>
<td>0.499</td>
</tr>
<tr>
<td>$k_{\text{death,ZOO},T_0,\text{wal}}$ (d$^{-1}$)</td>
<td>0.001</td>
<td>$k_{\text{miner,anae,sed,}T_0,\text{gre}}$ (d$^{-1}$)</td>
<td>0.668</td>
<td>$g_{B,\text{O}_2}$ (gO$^{1/2}$ m$^{-3/2}$)</td>
<td>1.093</td>
</tr>
<tr>
<td>$K_{\text{Feed}}$ (gDM m$^{-3}$)</td>
<td>0.945</td>
<td>$k_{\text{miner,anae,sed,}T_0,\text{zh}}$ (d$^{-1}$)</td>
<td>0.866</td>
<td>$g_{E,\text{ALG}}$ (gWM$^{1/2}$ m$^{-3/2}$)</td>
<td>0.637</td>
</tr>
<tr>
<td>$k_{\text{gro,ALG},T_0,\text{I}}$ (d$^{-1}$)</td>
<td>1.509</td>
<td>$k_{\text{miner,anae,sed,}T_0,\text{wal}}$ (d$^{-1}$)</td>
<td>2.146</td>
<td>$g_{E,\text{HPO}_4}$ (gP$^{1/2}$ m$^{-3/2}$)</td>
<td>0.013</td>
</tr>
<tr>
<td>$k_{\text{gro,ZOO},T_0}$ (d$^{-1}$)</td>
<td>1.265</td>
<td>$k_{\text{miner,anae,sed,}T_0,\text{wal}}$ (d$^{-1}$)</td>
<td>2.146</td>
<td>$g_{E,\text{NO}_3}$ (gN$^{1/2}$ m$^{-3/2}$)</td>
<td>0.043</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$k_{\text{miner,anae,sed,}T_0,\text{zh}}$ (d$^{-1}$)</td>
<td>1.880</td>
<td>$g_{E,\text{O}_2}$ (gO$^{1/2}$ m$^{-3/2}$)</td>
<td>0.201</td>
</tr>
</tbody>
</table>

found for Walensee and the highest for Greifensee. This agrees with expected fish densities across the eutrophication gradient of these lakes. Mineralization rate coefficients in the sediment are parameters that can be expected to vary because they aggregate effects of bacterial density, growth rate and geometric resolution of the sediment. Except for the aerobic mineralization rate coefficient of Greifensee, all mineralization rate constants increased by 1-2 orders of magnitude. Largest increases were found for the anaerobic mineralization and for Walensee. As before, Greifensee showed the greatest mineralization rate coefficients and Walensee the smallest. This is in agreement with our perception of the bacterial abundance at the different trophic levels. Compared to the model calibration for Lake Zurich only (see (Dietzel and Reichert, 2012)), the results of the calibrated parameter values showed some similarities. The parameters of the error model in particular were in good agreement with the joint calibration results. For the mineralization rates of Lake Zurich, the lake-specific calibration led to smaller values, but to the same order of magnitude and the same succession of the three mineralization processes. As those parameters are lake-specific parameters, they should result in very similar values during the different calibrations. The divergences demonstrate the difficulty and complexity of the calibration problem. During calibration of Lake Zurich only, the growth rate of phytoplankton was found to be smaller than during joint calibration of all lakes, while the zooplankton growth rate was larger. This can result from differences in the composition of plankton communities between lakes that effect growth rates. Also, the divergences can again be an indication for the complexity of the calibration problem or for an identifiability problem.
Greifensee

For a comparison of data and model results, data points (markers) only have to be compared with the dashed line during the calibration period, which becomes a bold line during validation and represents the output of the deterministic model at the maximum of the approximate posterior distribution (Fig. 3.5). The other lines and areas are discussed below. For a more direct comparison of data and best simulation results gained by calibration, as well as results for compartments not shown herein, see Fig. B.2 in Appendix B.3.

For most state variables the modeled concentrations were in good agreement with the measured concentrations. The long-term trends in plankton and phosphate concentrations were well represented by the model, whereas the modeled annual patterns of plankton showed a less dynamic behavior than the measurements. Year-to-year variability of plankton was less pronounced in the calculations. Additionally, calculated zooplankton concentrations were still too low, as well as phosphate concentrations in some years. Nitrate and oxygen concentrations in the compartment shown tended to be too high; the model was not always able to show the anoxic conditions in summer.

Lake Zurich

Even better results are found for the mesotrophic Lake Zurich (Fig. 3.6). The long-term trends of plankton, nutrients and oxygen were well represented by the model. Best results were achieved for oxygen and phosphate. For the latter, the model results captured the decreasing long-term trend very well. Nitrate matched the data less well, as the model indicated higher concentrations and less annual dynamics in the epilimnion than the measurements. More results for Lake Zurich can be found in Fig. B.3 in Appendix B.3.

Walensee

In the progress of calibration, the results for Walensee showed the greatest discrepancies with the data. This was especially true for phosphate concentration that did not represent the significant decline apparent in the data. Introducing the additional sorption process, which decreased the actual phosphate load, reduced the problem. After intensive calibration studies, the calculations agreed with the data about as well as for the other lakes (Fig. 3.7). In general, the model captured the long-term trends.
3.3. Results

Figure 3.5: Phytoplankton (entire lake), zooplankton (entire lake), nitrate (epilimnion), phosphate (epilimnion) and oxygen (epilimnion) concentrations in Greifensee. Data points (markers), output of the deterministic model (long-dashed), median (solid) and 95% credibility bounds (dark grey area with dashed boundaries) of bias-corrected output and median (solid; same as for bias-corrected output) and 95% credibility bounds (dark and light grey areas with dotted boundaries) of predictions of new observations (including observation grey error).
Figure 3.6: Phytoplankton, zooplankton, nitrate and phosphate concentrations in the epilimnion and oxygen concentrations in the hypolimnion of Lake Zurich. Data points (markers), output of the deterministic model (long-dashed), median (solid) and 95% credibility bounds (dark grey area with dashed boundaries) of bias-corrected output and median (solid; same as for bias-corrected output) and 95% credibility bounds (dark and light grey areas with dotted boundaries) of predictions of new observations (including observation error).
However, in Walensee the oxygen concentration in the hypolimnion also tended to be too high. Modeled nitrate concentrations showed similar problems as for Lake Zurich, concentrations were too high and not so dynamic as the data. Phosphate concentration showed the decreasing trend over time, but the decline was rather too rapid in the first years. See Fig. B.4 in Appendix B.3 for a comparison of epilimnion and hypolimnion results of Walensee.

**All lakes**

In general, most output variables showed a good agreement with the measured concentrations. The long-term trends, where present, of all output variables were reproduced well by the model. This was especially obvious in the temporal changes in phosphate and plankton, in particular for Lake Zurich and Walensee. Recurring problems for all three lakes included annual dynamics of some of the output variables that were too smooth or even out-of-phase, as for the zooplankton in the hypolimnion of Lake Zurich (Appendix B.3). Furthermore, concentrations of nitrogen and oxygen were too high and plankton too low (especially zooplankton). These are persistent problems while applying BELAMO. Overall, the results indicate only a coarse representation of the plankton dynamics in these lakes but a more detailed representation of the biogeochemistry.

### 3.3.3 Uncertainty analysis

The dark shaded areas in the plots shown in Figs. 3.5–3.7 represent the 95% credibility intervals of our posterior knowledge of the true states (marginal posterior distributions of the components of $y_{LM}^\text{b} + B_{LM}^b$ in equation (3.23)) of the variables shown in the different plots. The light shaded extensions of these areas represent our uncertainty about new observations (including observation error).

The results demonstrate that our knowledge of the true state was much more accurate during the calibration period (up to 1995) than during the prediction period (after 1995) (Figs. 3.5–3.7). This follows from the information about the bias gained during the calibration period due to comparison with data. For the prediction period, we lost information about the direction and the absolute amount of bias at each time point, but we kept knowledge about the uncertainty due to bias. This led to larger credibility intervals and to merged results of the deterministic model and the median of the posterior of model results plus bias during the prediction period. The model was significantly more accurate for the chemical variables than for phyto- and zooplankton. A comparison of
Figure 3.7: Phytoplankton, zooplankton, nitrate and phosphate concentrations in the epilimnion and oxygen concentrations in the hypolimnion of Walensee. Data points (markers), output of the deterministic model (long-dashed), median (solid) and 95% credibility bounds (dark grey area with dashed boundaries) of bias-corrected output and median (solid; same as for bias-corrected output) and 95% credibility bounds (dark and light grey areas with dotted boundaries) of predictions of new observations (including observation error).
the light shaded extensions with the dark shaded areas indicates that the observation error can be neglected compared to the bias for all chemical variables, whereas it led to increased uncertainty for observations of phyto- and zooplankton.

As the objective of this study was a joint calibration of the same model to all investigated lakes, the same parameters for the bias were applied to all lakes. This implies that the bias considers not only bias within the time series of a single lake, but also across lakes.

### 3.3.4 Nutrient mass fluxes

Nutrient fluxes for the three lakes were calculated with the calibrated version of BE-LAMO and encompass the dissolved nutrients as well as their fractions in degradable organic material (Figs. 3.8 and 3.9). Within the model, the bioavailable parts of the nutrients were represented by the state variables of their dissolved compounds ($S_{\text{HPO}_4}; S_{\text{NO}_3} + S_{\text{NH}_4}$) and their fractions in degradable organic material ($X_{\text{P,S}}$ (the variable fraction of phosphorus within degradable organic particles $X_S$), $X_{\text{P,LS}}$ (see end of section 3.2.2), $X_{\text{P,ALG}}$ (the variable fraction of phosphorus in phytoplankton $X_{\text{ALG}}$), the constant fraction of phosphorus in $X_{\text{ZOO}}$ and the constant fractions of nitrogen in $X_{\text{ALG}}$, $X_{\text{ZOO}}$ and $X_S$). The Figs. 3.8 and 3.9 depict the main transfer processes as input, output, sedimentation, mineralization and subsequent release of the dissolved compounds to the water phase, diffusion between water and sediment and denitrification in case of nitrogen. The accumulation in the water phase shows the calculated accumulation in the model. It did not completely match the difference between input and output fluxes in all cases. Those errors are below 5% and within the range of numerical inaccuracies due to non-continuous model outputs used for the calculation of fluxes. Fig. 3.8 visualizes the importance of denitrification that increases with increasing nutrient richness. For oligotrophic Walensee only slightly more than 10% of the nitrogen input was denitrified, for mesotrophic Lake Zurich this fraction amounted to around 35% and for eutrophic Greifensee to more than 55%.

### 3.4 Discussion

We discuss two major aspects of this study. In the first section, we discuss what we have learned about causes, reduction and dealing with remaining systematic errors. In the second, we elaborate on relevant ecological processes in the studied lakes.
Figure 3.8: Mass fluxes of bioavailable nitrogen in the water column and the sediment of Greifensee (left numbers), Lake Zurich (middle numbers) and Walensee (right numbers) in t/a (averaged over the 5 years 2001-2005) calculated with BELAMO. The errors for the overall mass balances are mostly around 1%, at most 5%.

3.4.1 Bias

As systematic errors are the main concern in any predictive modeling, we discuss the following three aspects of systematic errors in the following section: causes of bias, reduction of bias, and consideration of bias for model calibration and prediction. We then discuss biogeochemical implications and findings resulting from the model presented.
Figure 3.9: Mass fluxes of bioavailable phosphorus in the water column and the sediment of Greifensee (left numbers), Lake Zurich (middle numbers) and Walensee (right numbers) in t/a (averaged over the 5 years 2001-2005) calculated with BELAMO. The errors for the overall mass balances are mostly around 1%, at most 5%.

Causes of bias

Systematic deviations of model outputs from observations are primarily caused by the simplified description of reality by the model. Main problems are neglected or unknown external influence factors and the use of a simplified model structure including potential simplifications by aggregation (in time, space or state variables; by the model or by sampling procedures) and by the choice and parameterization of processes considered in the model.
A first major problem in our current simulations was spatial aggregation. For computational reasons, the model divides the lake water column into two mixed boxes (with constant volumes) only, the epilimnion and the hypolimnion. This leads to a poorer representation of vertical mixing processes than in the 1d model used earlier for shorter simulation periods (Omlin et al., 2001b; Mieleitner and Reichert, 2006) and it neglects variations in mixing depth. In addition, both model versions neglect horizontal inhomogeneity. This problem becomes even worse, as the depth-integrated samples of zooplankton (and phytoplankton, in case of Greifensee) do not fit directly to the resolution of compartments in the lake model. Hence, the aggregated and processed data might not represent the average concentrations in epilimnion and hypolimnion as described by the model compartments. The amount of bias in these cases supports the finding. The problems caused by the high spatial aggregation, might be solved by simulations with the one-dimensional model version and comparison with concentration profiles where possible.

A second problem was the aggregation of more than 100 species to only single groups of phytoplankton and zooplankton. On the one hand, this may simplify the description of nutrient fluxes (see Mieleitner and Reichert, 2008 for a discussion of potential problems), but on the other hand it makes it necessary to use average kinetic characteristics and behaviors for the whole group. This does not account for the variability in species attributes and composition of the ecological community, and in particular not for the succession of dominant species within each year. Least of all, genetic adaptation or acclimatization processes of the plankton community are considered, except for changes in phosphorus content.

A further important potential source of bias was our attempt to fit jointly three lakes of different trophic status with a single model and with a minimum of lake-specific parameters over a long simulation period during which the phosphorus input changed considerably. This requires a model structure and parameterization that is considerably more universal than for describing a single lake over a shorter period. In particular, the addition of potential bias across lakes to the bias within time series of each lake obviously increases the bias. Our task of jointly modeling all lakes over such a long period of time obviously challenges the model and it is a significant result of our study to demonstrate that this is possible.

Another challenge was the use of a model with closed mass balances that describes mineralization of sedimented organic particles rather than using sediment oxygen demand and ammonia and phosphate release as external parameters, as is still done in the majority of models (Mooij et al., 2010). In addition to limiting the degrees of freedom
of exchange processes, this required a mineralization model for the sediment. As the different mineralization processes were described by just a single overall lake-specific mineralization rate, the microbial community was (implicitly) assumed to be constant. For example, the problems in representing the anoxia in Greifensee in some years could result from difficulties in finding adequate mineralization rates for the whole simulation time due to changes in the bacterial community which the model does not describe. The dynamics of microbial communities, changes in bacterial abundance and adaptation processes to nutrient loads and availability of organic material, were not explicitly taken into account. The introduction of biomass of mineralizing bacteria as an additional state variable would be a meaningful extension to tackle this problem. But one has to be aware that this would also cause additional difficulties, as it increases the complexity of the model. The increased number of parameters would also increase the calibration problem, especially because the number of model output variables that can be fitted to data cannot be increased, as there are no bacteria data available.

Reduction of bias

Although the model used in previous studies led to good simulations of all three lakes over a period of four years without significant changes in input, the model required important changes in order to be able to describe correctly the observed trends in phosphate and plankton concentrations, while still being able to represent correctly nitrate and dissolved oxygen concentration. The most important changes were (see section 2.2.2 for more details):

- introduction of active movement of zooplankton to the epilimnion;
- improvement of the formulation of zooplankton growth to decrease the growth rate for small phytoplankton concentrations more strongly than linearly;
- improvement of the formulation of predation on zooplankton by Chaoborus similarly to zooplankton growth on phytoplankton;
- improvement of the dissolved oxygen mass balance of anaerobic mineralization by introducing methane production and subsequent re-oxidation in the water column (this process also represents the effect of mineralization by sulfate, iron oxide and manganese oxide reduction and subsequent re-oxidation);
- reduction of the biodegradability of organic particles in the inflow to decrease their release of phosphate into the lake;
• addition of sorption of phosphate to sedimenting “inert” organic particles addition-ally to uptake by biologically degradable organic particles.

These modifications led to a considerable reduction in bias and a reasonable representation of key features of the data. Nevertheless significant discrepancies between model results and data remained. These discrepancies were larger and more systematic than expected for observational errors and must be considered during the model calibration process to make it possible to get adequate uncertainty estimates of model parameters and predictions.

**Consideration/Description of bias**

The chosen calibration method used a bias description approach that was originally published in the statistical literature (Craig et al., 1996; Craig et al., 2001; Kennedy and O’Hagan, 2001; Higdon et al., 2004; Bayarri et al., 2007), subsequently transferred to environmental applications and linked to multi-objective model calibration (Reichert and Schuwirth, 2012), and finally approximated to make it applicable for computationally demanding models (Dietzel and Reichert, 2012). Application of this technique to our long-term lake simulations led to the consideration of important aspects that are not yet commonly addressed in environmental modeling:

• *Separation of bias and observation error*: While the bias component shows autocorrelated behavior characteristic of structural errors, the observation error is approximately independently and normally distributed (in our application on a Box-Cox transformed scale, see section 2.5).

• *Consideration of bias in uncertainty estimates of model parameters*: The description of the bias as an autocorrelated process (partially) solves the problem that the uncertainty of model parameters is underestimated with an independent error model that does not account for bias.

• *Consideration of bias for model predictions*: The chosen technique allowed us to predict the current state of knowledge of the true model variables as well as about (future) observations. It naturally leads to narrow prediction credibility intervals during the calibration period, where the observations constrain the model output, and to (much) wider intervals during the prediction period. This is an important aspect that is not included in most calibration techniques currently applied in environmental modeling. However, consideration of bias for predictions is based on
the assumption that model and bias structure do not change. This is obviously a strong assumption that can hardly be avoided for the bias component, as it describes the deviations that cannot easily be described by a mechanistic model.

- **Bias of a joint simulation:** In our application, systematic errors were even increased by the objective of jointly describing three lakes of different trophic status and varying input loads over a long time period. This requires a very high degree of universality of the model. Such a high degree of universality is a desired property of a model as it increases our confidence of model predictions under changes in driving forces; but it makes calibration even more challenging.

### 3.4.2 Biogeochemical and ecological processes

This project with the lake model BELAMO provided some insights into the biogeochemistry and ecology of the Swiss lakes Greifensee, Lake Zurich and Walensee. The comparison of model results with measurements indicated a more complex than linear dependence of zooplankton grazing on phytoplankton concentration when phytoplankton concentrations are low. This functional description has similarities to a Holling-type III response \cite{Holling1959a,Holling1959b}. The heterogeneity of plankton abundance is a possible explanation for that. Furthermore, the upwards movement of zooplankton seems to be an important process impacting the grazing of zooplankton on phytoplankton and the distribution of zooplankton between epi- and hypolimnion. A comparison of the three lakes indicates the influence of insect larvae on zooplankton concentration in Greifensee, not present in Lake Zurich and Walensee. Additionally, the different calibration results (see Table 3.4 for calibrated parameter values) for the death rates of zooplankton in the three lakes give evidence for an increasing predation pressure by fish on zooplankton with increasing nutrient richness across the trophic gradient. The trophic status of the lakes is also the reason for the different importance of the three mineralization processes. When comparing the lake-specific sensitivity analyses (see results of lake-specific sensitivity analyses, Appendix B.2), it becomes obvious that, for Walensee, aerobic mineralization is the dominant mineralization process. For Lake Zurich, the importance of anoxic mineralization is greater than in Walensee, whereas for Greifensee anaerobic mineralization also has an increased influence. These findings were supported by the calculated nitrogen mass balances, which indicated that the fraction of denitrification was largest for eutrophic Greifensee and smallest for oligotrophic Walensee. Moreover, the model study supports the assumptions that allochthonous organic particles are less degradable than autochthonous and that the sorption of in-
organic phosphorus to slowly degrading organic particles can be an important loss process for nutrients.

In conclusion, despite the large simplifications in system description, BELAMO represented the main aspects of the biogeochemistry and, rather more coarsely, the ecological dynamics of plankton in the study lakes. With that we are able to apply the model also to other lakes with similar characteristics as the application cases shown herein. Further work could concentrate on comparing different models applied to the same lakes as well as the same model applied to an extended range of lakes, including lakes from other climate zones. This would be possible by changing the mixing parameters in the model, according to climate. In case that the climate is not expected to change significantly in the future, predictions about the future state of a lake can also be performed. To model the effects of climate change, BELAMO would have to be coupled to a hydrodynamic model.

### 3.5 Acknowledgments

We thank Richard Forster (WVZ), Hans-Rudolf Bürgi (Eawag), Barbara Känel, Pius Niederhauser, Irene Purtschert and Peter Spohn from AWEL, Margot Blaser and Markus Faden from AFU St. Gallen and Daniel Streit and the department for water from BAFU for kindly providing their data on chemistry, biology and physics of the lakes, WWTPs and inflowing rivers. Karsten Rinke, Sebastian Sobek and Bernhard Wehrli are acknowledged for helpful comments on different aspects of the lake model and Rosi Siber for GIS support and the preparation of Fig. 3.1. We thank the Polyfaktor Designbüro (http://www.polyfaktor.com/) for the layout of Figs. 3.2, 3.3, 3.8 and 3.9. The Swiss National Science Foundation (SNSF) is acknowledged for the financial support of the project.
Appendix B

Supporting information to: Effects of changes in the driving forces on water quality and plankton dynamics in three Swiss lakes - long-term simulations with BELAMO

Fig. B.1: Zooplankton grazing

Figure B.1: The dependency of zooplankton grazing on phytoplankton concentration according to equation (3.11) describing the limitation term $f_{ALG}$ for zooplankton growth on phytoplankton. The parameter $K_{Feed}$ was calibrated to 0.95 gDM m$^{-3}$. At this value, the function switches from a quadratic dependence on phytoplankton concentration to a Monod-type limitation.
B.1 Data compilation and processing

B.1.1 Chemical and physical variables

Lakes In Lake Zurich, profiles of temperature, oxygen, nitrate, ammonium and phosphate were taken by the Water Supply Authority of Zurich (WVZ) at the deepest location (136 m). The profiles consist of measurements at 19 different depths. In Walensee, the same variables were measured by the WVZ at the deepest location (151 m) and in Greifensee these substances were measured by Eawag up to 2000 and afterwards by the AWEL Zürich at its deepest location (32 m) in several depths.

Since we are interested in data comparable with the results of the box model, we had to aggregate the depth-specific values over the depths of the corresponding boxes. In order to receive results which relate either to the epilimnion, the hypolimnion or the entire lake, we calculated the volume-weighted averages of concentrations for the epilimnion (first 5 m in Greifensee and first 10 m in Lake Zurich and Walensee) and the hypolimnion (from 5 m and 10 m, respectively, to the lake bed) by multiplying each concentration measurement by the ratio of the water volume corresponding to the lake area in the depth in which the measurement was taken and the water volume of the whole lake model compartment (epilimnion, hypolimnion or the whole lake) and summing up over all measurements belonging to one lake model compartment. This processing differs from that done by Mieleitner and Reichert, 2008 as they treated only the first 2.5 m as the epilimnion in all three lakes. The larger epilimnion box accounts for measured temperature profiles and the problem that an epilimnion that is too small can lead to unrealistic mixing coefficients fitted to the aggregated temperature values.

An approximation of the light intensity in W m$^{-2}$ at the lake surface was obtained using solar radiation data from measurement sites of the Swiss Meteorological Service (MeteoSchweiz). For Lake Zurich and Greifensee the measurement site in Zurich, for Walensee the one in Glarus (see Fig. 3.1) was used. Hourly measured values were converted to monthly averages. The day lengths were also estimated from these data.

Inflows Upper Lake Zurich is the main input to lower Lake Zurich. At the dam separating the lake into two parts, among others oxygen, nitrate, ammonium, phosphate and total and particulate phosphorus concentrations were measured monthly for the years 1976-2005 by the WVZ. For the estimation of the discharge at the dam we compared the discharge data at the outflow of Walensee (Linthkanal) that flows into upper Lake Zurich with the river Limmat, main outflow of lower Lake Zurich, where data were avail-
B.1. Data compilation and processing

able. On average, the Limmat discharge is larger by a factor of 1.6 than the discharge of Linthkanal. As the amount of water that flows into and out of whole Lake Zurich should be the same, and the other inflowing rivers are comparably small, we could assess the discharge at Seedamm by multiplying the time series of Linthkanal discharge at Weesen with a factor of 1.6. This factor is confirmed by the fact that the averaged discharge at the dam calculated from sparse discharge measurements is larger by a factor of about 1.6 than the averaged Linthkanal discharge (Gammeter and Forster, 2002). The estimated inflow into lower Lake Zurich was used to calculate the nutrient input loads from the concentration data at the dam.

To estimate the input loads of the less important inflowing rivers Feldbach, Hornbach Zürichhorn, Dorfbach (Küschnacht), Dorfbach (Erlenbach), Dorfbach (Meilen), Dollikerbach, Reidbach and Aabach (Horgen) we could not use data measured by AWEL Zürich directly at these rivers, because monthly measurements took place only in some years and for each river in different years. But the nutrient concentrations of all rivers are in the same order of magnitude as in the river Jona that flows into upper Lake Zurich. For the river Jona, monthly data of oxygen, nitrate, ammonium, phosphate and total phosphorus concentrations were available from 1994 to 2006. The land use in this catchment is similar to that in the catchments of the rivers around lower Lake Zurich. The catchment of lower Lake Zurich, excluding the one of upper Lake Zurich, is 4.13 times larger than the Jona catchment. Hence, we could assess the load of the smaller rivers contributing to the inflow of lower Lake Zurich by multiplying the load of the river Jona by a factor of 4.13. For the years before 1994 the same nutrient load as at the beginning of 1994 was assumed, as there was no information available supporting a much higher (or much lower) nutrient load or a specific annual load dynamic.

The 10 waste water treatment plants (WWTPs) Hombrechtikon, Horgen, Küsnacht, Männedorf, Meilen, Richterswil, Stäfa Oetikon, Stäfa Uerikon, Thalwil, and Wädenswil discharge directly into Lake Zurich and represent the second important input source. Monthly measured data of all waste water treatment plants could be obtained from AWEL Zürich for the variables discharge, nitrate, ammonium, phosphate and total phosphorus. The data covers the time range 1976-2005. Input loads were calculated from this data.

For Walensee, monthly measured data of the four main discharging rivers could be obtained from the AFU St. Gallen for the variables discharge, nitrogen, ammonium, oxygen, and phosphate, particulate and total phosphorus. For the river Linth data was available from 1976-2005, for the river Seez from 1988-2005 and for the rivers Tscherlerbach and Fabrikkanal monthly data were only available from 1990-1995, some ran-
dom samples were also taken by AFU St. Gallen from 1987 to 1999 (river Fabrikkanal is an artificial channel and was closed in 1999). In order to complete the time range for river Seez, Linth data available over the whole time range were compared with Seez data where possible. The calculated ratio of the two data sets was extrapolated to 1976-1988. Monthly averages of river Seez were estimated from Linth data. For the small brooks Fabrikkanal and Tscherlerbach information was very sparse. As they are of minor importance for the total input and their nutrient loads are far less dynamic than of the rivers Linth and Seez, the average value of the data in the years 1990 to 1995 was calculated and taken as constant for the missing years. The input loads of the remaining, not monitored brooks, was estimated in the same way as for Lake Zurich. For Walensee, river Linth was taken as a reference river and its input loads and discharge multiplied by a catchment correction factor. As the monthly measurements of the river Linth are very erratic, we found it better to use smoothed data.

The five waste water treatment plants Seez, Mittensee/Murg, Walenstadt, Amden and Unterterzen discharge directly into the Walensee. The WWTP Unterterzen was closed in 1999 and the water consecutively was directed to the WWTP Mittensee. The WWTP Amden was closed in 1992 and connected to a WWTP not discharging into Walensee. Data of discharge, nitrogen, ammonium, phosphate and total phosphorus of the waste water treatment plants could be obtained from the AFU St. Gallen for the years 1978-2005. The data were extended until 1976 by taking the mean value of the data from 1978-1981 for the missing years, as it was assumed that measurements are more similar among years of close temporal proximity.

For Greifensee, daily measured data of the rivers Aa Uster and Aa Mönchaltorf were obtained from AWEL Zürich for discharge, nitrogen, ammonium, phosphate and total phosphorus (only weekly). The data cover the time range 1988-2005. Monthly averages were calculated from these data. There are only few nutrient measurements of the remaining smaller rivers. Their input loads and discharge were estimated by a correction factor related to the comparison of the catchment sizes. This factor was multiplied by the data of Aa Mönchaltorf, as the land use of this catchment is more similar to the remaining Greifensee catchment.

The three waste water treatment plants Uster, Maur and Mönchaltorf discharge directly into the Greifensee. Monthly measured data of discharge, nitrogen, ammonium and phosphate were obtained from AWEL Zürich for the years 1988-2005.

According to the temperature of the inflowing rivers compared to the lake temperature, the loads were treated as inputs to the epilimnion or the hypolimnion. The loads from WWTPs were always implemented as an input into the epilimnion.
Wet deposition of phosphate and nitrogen was considered as an additional input source. Inputs into the lakes from combined sewer overflows were omitted, as there were no reliable estimations available. At least for Greifensee, the mixed 24 h sampling technique of the two main inflows should cover most of these contributions.

B.1.2 Biological variables

Lakes  Biological variables were measured at the same locations as the physical and chemical variables in all three lakes. In Lake Zurich and Walensee, concentrations of phytoplankton and zooplankton were measured by the WVZ for the years 1976 (zooplankton: 1977) to 2005. Phytoplankton was measured as profiles, samples at several depths. Zooplankton concentrations were measured by two depth-integrated samples (collection of particles with a net from the bed to 20 m depth and from 20 m depth up to the surface). For the years 1977-1982 only one zooplankton sample integrated over the whole lake depth was collected, for the two years 1983/1984 there is a gap in the data. Data for Lake Zurich and Walensee consist of counts of 100-140 different species. For the biogeochemical lake model, the total volume of all phytoplankton and zooplankton species was used. The volumes were calculated by multiplying the counts of each species by the typical volume of one cell of the species. The volume was converted to wet mass using the density of water. For the conversion from wet mass to dry mass, which is modeled, a factor of 5 was used for phytoplankton [Bürgi, 1994; Jørgensen et al., 1991] and a factor of 10 was used for zooplankton [based on measurements by WVZ, the factor of 5 used by Omlin et al., 2001b was replaced by a factor of 10].

In order to receive values which either represent the concentration in the epilimnion, the hypolimnion or the entire lake, the volume-weighted averages of phytoplankton was calculated (see above). For zooplankton, it was assumed that it is mostly present in the epilimnion (above 10 m depth), where also its food source is available. Based on this assumption and because the hypolimnions of Lake Zurich and Walensee are large (116 m and 131 m depths, respectively), we concluded that the zooplankton concentration between 10 and 20 m does not differ enough from the concentration in a depth below 20 m to be a significant contribution to the overall hypolimnion concentration. Hence, the concentration of zooplankton in the hypolimnion was estimated as the concentration in the sample between 20 m and the bed. The epilimnion concentration was estimated by subtracting the mass of zooplankton between 10 and 20 m (calculated from the depth-integrated sample below 20 m) from the mass of zooplankton measured in 0-20 m.
In Greifensee, phytoplankton and zooplankton were measured by Eawag in terms of depth-integrated samples over the whole lake depth in the case of phytoplankton and over the top 20 m in the case of zooplankton for the time range 1987-2004 and -2005, respectively. Hence, we could not distinguish the epilimnion and hypolimnion concentrations; only mean lake concentrations could be calculated. Remaining processing of the plankton data was done as described for Lake Zurich and Walensee.

**Inflows**  
The model does not account for inflowing rivers as a plankton source, as this contribution seems to be relatively small compared to the nutrient and plankton turnover within the lakes. However, a small colonization flux for phyto- and zooplankton is considered in the model, which also summarizes zooplankton colonization by ephippia or diapausing organisms.

### B.2 Lake-specific sensitivity analysis

Table B.1 shows the results for the sensitivity analyses separately for each lake. Here, it becomes obvious that the zooplankton death rate for the Walensee does not seem to be very important for the model results. This might result from the fact that the calibrated value for which the sensitivity analysis was done, is already very small. The rate was included in the calibration as the ones of the other two lakes were included as well. The same is true for the different types of mineralization rates in the different lakes, which only partly result to be influential. The parameter $f_{X_1,rivers}$ is more important for Walensee, as this lake is mainly influenced by input from the river Linth, where Lake Zurich and Greifensee are influenced by lake inputs by a considerable amount. Hence, the parameters $a_{P,inflow,lake}$ and $f_{X_1,lake}$ do not play a role for the Walensee model results (parameters with no sensitivity are not shown in Table B.1). For Greifensee, the upwards velocity of zooplankton, $v_{up,ZOO}$, does not influence the model results averaged over the whole lake depth.

In general, the lake-specific sensitivity analyses give similar results as the joint sensitivity analysis. Again, the parameters determining the mixing between the two lake compartments, the parameters $K_{Feed}$, $K_{ALG,ZOO}$, $K_{I,ALG}$, and the growth and death rates of phyto- and zooplankton are among the most influential parameters. We see some differences in the importance of the mineralization rates for the three lakes. First, we detect a larger importance of the mineralization rates (especially of aerobic mineralization, $k_{miner,aero,sed,T_0,wal}$ for Walensee than for the other two lakes. This could be due
B.2. Lake-specific sensitivity analysis

to the smaller values of calibrated mineralization rates for which the model might be more sensitive to changes, than for the other two lakes, where the mineralization rates were already estimated large and where a change might have a smaller effect. Second, we see a deviation in the order of influences of the different mineralization processes. For Walensee, the more relevant mineralization process seems to be aerobic mineralization, followed by anoxic and last by anaerobic. The order of the processes is the same for Lake Zurich, but in this case aerobic and anoxic mineralization rates are closer together, only later followed by the anaerobic mineralization rate. For Walensee the aerobic rate is much more influential, followed only later by anoxic and anaerobic rates. For Greifensee even the order of the influence of the different mineralization processes is completely reverse. Here, the anaerobic mineralization rate is the most influential one.

These findings are in accordance to our expectation of the importance of the different mineralization processes due to the trophic states of the three lakes. In general, because of the small amount of organic material in the water column and also lower concentrations of bacteria, the mineralization rates in the water column, $k_{\text{miner,aero,} \text{wat},T}$, $k_{\text{miner,anox,} \text{wat},T}$ and $k_{\text{miner,anae,} \text{wat},T}$, play a minor role compared to the mineralization in the sediment. The nitrogen half-saturation concentration of phytoplankton growth, $K_{N,\text{ALG}}$, shows the largest effect in Walensee, followed by Lake Zurich and last by Greifensee. This is plausible due to different nitrogen levels in the three lakes. Although the phytoplankton in all lakes is mainly limited by phosphorus, the potential for nitrogen limitation is largest in Walensee and smallest in Greifensee. The smaller the actual concentration, the more sensitive the Monod model is to the half-saturation concentration, visible in the different sensitivities to $K_{N,\text{ALG}}$. The same is true for the phosphate half-saturation concentration of phytoplankton growth, $K_{HPO_4,\text{ALG}}$. However, the effect is less pronounced than for nitrogen, the sensitivities are more similar between lakes, as phosphate is limiting in all lakes. Additionally, as phosphate is more limiting than nitrogen, $K_{HPO_4,\text{ALG}}$ is in general more influential than $K_{N,\text{ALG}}$.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$\delta_{\text{msqr}}$</th>
<th>$\delta_{\text{g}}$</th>
<th>Parameter</th>
<th>$\delta_{\text{msqr}}$</th>
<th>$\delta_{\text{g}}$</th>
<th>Parameter</th>
<th>$\delta_{\text{msqr}}$</th>
<th>$\delta_{\text{g}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_{2,\text{summer,gre}}$</td>
<td>24.681</td>
<td></td>
<td>$K_{2,\text{winter,zh}}$</td>
<td>10.995</td>
<td></td>
<td>$K_{\text{feed}}$</td>
<td>1.685</td>
<td></td>
</tr>
<tr>
<td>$K_{\text{feed}}$</td>
<td>21.513</td>
<td></td>
<td>$w_{\text{ZOO}}$</td>
<td>8.920</td>
<td></td>
<td>$\text{tempdif}_{\text{zh}}$</td>
<td>1.683</td>
<td></td>
</tr>
<tr>
<td>$k_{\text{gro,ALG,T0,T0}}$</td>
<td>14.086</td>
<td></td>
<td>$k_{\text{gro,ALG,T0,T0}}$</td>
<td>8.038</td>
<td></td>
<td>$K_{\text{gro,ALG,T0}}$</td>
<td>1.614</td>
<td></td>
</tr>
<tr>
<td>$K_{\text{ALG,ZOO}}$</td>
<td>8.370</td>
<td></td>
<td>$K_{2,\text{summer,zh}}$</td>
<td>6.631</td>
<td></td>
<td>$K_{\text{resp,ALG,T0}}$</td>
<td>0.905</td>
<td></td>
</tr>
<tr>
<td>$b_{\text{p,max}}$</td>
<td>7.566</td>
<td></td>
<td>$k_{\text{death,ZOO,T0,T0}}$</td>
<td>6.511</td>
<td></td>
<td>$\theta$</td>
<td>0.903</td>
<td></td>
</tr>
<tr>
<td>$k_{\text{death,ZOO,T0,gre}}$</td>
<td>6.607</td>
<td></td>
<td>$\text{tempdif}_{\text{zh}}$</td>
<td>6.354</td>
<td></td>
<td>$k_{\text{death,ALG,T0}}$</td>
<td>0.856</td>
<td></td>
</tr>
<tr>
<td>$f_{\text{sol}}$</td>
<td>5.699</td>
<td></td>
<td>$v_{\text{up,ZOO}}$</td>
<td>5.864</td>
<td></td>
<td>$K_{\text{ALG,ZOO}}$</td>
<td>0.677</td>
<td></td>
</tr>
<tr>
<td>$k_{\text{gro,ZOO,T0}}$</td>
<td>6.500</td>
<td></td>
<td>$K_{1,\text{ALG}}$</td>
<td>5.719</td>
<td></td>
<td>$K_{HPO_4,\text{ALG}}$</td>
<td>0.651</td>
<td></td>
</tr>
<tr>
<td>$K_{1,\text{ALG}}$</td>
<td>5.904</td>
<td></td>
<td>$k_{\text{death,ALG,T0}}$</td>
<td>5.718</td>
<td></td>
<td>$S_{\text{HPO}_4,\text{crit}}$</td>
<td>0.639</td>
<td></td>
</tr>
</tbody>
</table>

Continued on next page
Table B.1 – continued from previous page

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Greifensee $\delta_{\text{meq}}^{\text{meq}}$</th>
<th>Lake Zurich $\delta_{\text{meq}}^{\text{meq}}$</th>
<th>Walensee $\delta_{\text{meq}}^{\text{meq}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>tempdif,gre</td>
<td>5.641</td>
<td>3.595</td>
<td>3.874</td>
</tr>
<tr>
<td>$k_{\text{exp,ALG,T0}}$</td>
<td>5.349</td>
<td>3.874</td>
<td>0.562</td>
</tr>
<tr>
<td>$f_p$</td>
<td>5.194</td>
<td>3.420</td>
<td>0.561</td>
</tr>
<tr>
<td>$f_{\text{porgpart}}$</td>
<td>3.742</td>
<td>3.276</td>
<td>0.549</td>
</tr>
<tr>
<td>$\theta$</td>
<td>3.214</td>
<td>2.634</td>
<td>0.539</td>
</tr>
<tr>
<td>$f_{\text{X}_1,\text{Lake}}$</td>
<td>3.099</td>
<td>2.496</td>
<td>0.484</td>
</tr>
<tr>
<td>$\beta_{\text{ALG}}$</td>
<td>3.070</td>
<td>2.280</td>
<td>0.449</td>
</tr>
<tr>
<td>$k_{\text{death,ALG,T0}}$</td>
<td>2.686</td>
<td>2.274</td>
<td>0.448</td>
</tr>
<tr>
<td>$S_{\text{HPO}_4,\text{crit}}$</td>
<td>1.880</td>
<td>2.263</td>
<td>0.438</td>
</tr>
<tr>
<td>$K_{\text{HPO}_4,\text{ALG}}$</td>
<td>1.831</td>
<td>2.110</td>
<td>0.416</td>
</tr>
<tr>
<td>$h_{\text{sed},\text{alc}}$</td>
<td>1.817</td>
<td>1.536</td>
<td>0.414</td>
</tr>
<tr>
<td>$h_{\text{sed},\text{alg}}$</td>
<td>1.774</td>
<td>1.390</td>
<td>0.392</td>
</tr>
<tr>
<td>$w_{\text{ALG}}$</td>
<td>1.434</td>
<td>1.356</td>
<td>0.340</td>
</tr>
<tr>
<td>$\beta_{\text{ZOO}}$</td>
<td>1.195</td>
<td>1.194</td>
<td>0.300</td>
</tr>
<tr>
<td>$c_e$</td>
<td>1.080</td>
<td>1.019</td>
<td>0.299</td>
</tr>
<tr>
<td>$b_{\text{p,min}}$</td>
<td>1.023</td>
<td>0.991</td>
<td>0.282</td>
</tr>
<tr>
<td>$w_{\text{ORG}}$</td>
<td>0.859</td>
<td>0.786</td>
<td>0.276</td>
</tr>
<tr>
<td>$k_{\text{upt},\text{X}_1}$</td>
<td>0.822</td>
<td>0.764</td>
<td>0.229</td>
</tr>
<tr>
<td>$f_{\text{X}_1,\text{rivers}}$</td>
<td>0.716</td>
<td>0.595</td>
<td>0.225</td>
</tr>
<tr>
<td>$w_{\text{sed,ORG}}$</td>
<td>0.618</td>
<td>0.512</td>
<td>0.195</td>
</tr>
<tr>
<td>$a_{\text{p,flow,flux}}$</td>
<td>0.588</td>
<td>0.478</td>
<td>0.187</td>
</tr>
<tr>
<td>$k_{\text{miner,anox,sed,T0,gre}}$</td>
<td>0.566</td>
<td>0.404</td>
<td>0.187</td>
</tr>
<tr>
<td>$w_{\text{sed,ALG}}$</td>
<td>0.546</td>
<td>0.368</td>
<td>0.187</td>
</tr>
<tr>
<td>$\beta_{\text{bAC}}$</td>
<td>0.471</td>
<td>0.333</td>
<td>0.187</td>
</tr>
<tr>
<td>$a_{\text{p,flow}}$</td>
<td>0.470</td>
<td>0.295</td>
<td>0.187</td>
</tr>
<tr>
<td>$k_{\text{miner,anox,sed,T0,gre}}$</td>
<td>0.462</td>
<td>0.280</td>
<td>0.187</td>
</tr>
<tr>
<td>$k_{\text{miner,alco,sed,T0,gre}}$</td>
<td>0.420</td>
<td>0.279</td>
<td>0.187</td>
</tr>
<tr>
<td>$K_{\text{O}_2,\text{miner}}$</td>
<td>0.307</td>
<td>0.215</td>
<td>0.187</td>
</tr>
<tr>
<td>$K_{\text{O}_2,\text{max}\text{,X}_1}$</td>
<td>0.275</td>
<td>0.174</td>
<td>0.187</td>
</tr>
<tr>
<td>$P_{\text{inflow}}$</td>
<td>0.272</td>
<td>0.167</td>
<td>0.187</td>
</tr>
<tr>
<td>$D_{\text{NO}_3}$</td>
<td>0.272</td>
<td>0.167</td>
<td>0.187</td>
</tr>
<tr>
<td>$w_{\text{ZOO}}$</td>
<td>0.269</td>
<td>0.167</td>
<td>0.187</td>
</tr>
<tr>
<td>$K_{\text{O}_2,\text{winter,gre}}$</td>
<td>0.244</td>
<td>0.167</td>
<td>0.187</td>
</tr>
<tr>
<td>$a_{\text{N}}$</td>
<td>0.232</td>
<td>0.127</td>
<td>0.187</td>
</tr>
<tr>
<td>$k_{\text{O}_2}$</td>
<td>0.231</td>
<td>0.097</td>
<td>0.187</td>
</tr>
<tr>
<td>$K_{\text{N,ALG}}$</td>
<td>0.183</td>
<td>0.090</td>
<td>0.187</td>
</tr>
<tr>
<td>$k_{\text{resp,ALG,T0}}$</td>
<td>0.163</td>
<td>0.090</td>
<td>0.187</td>
</tr>
<tr>
<td>$k_{\text{porgpart}}$</td>
<td>0.073</td>
<td>0.090</td>
<td>0.187</td>
</tr>
<tr>
<td>$k_{\text{upt,\text{X}_1}}$</td>
<td>0.072</td>
<td>0.090</td>
<td>0.187</td>
</tr>
<tr>
<td>$K_{\text{N}_3,\text{miner}}$</td>
<td>0.070</td>
<td>0.090</td>
<td>0.187</td>
</tr>
<tr>
<td>$a_{\text{C}}$</td>
<td>0.053</td>
<td>0.038</td>
<td>0.187</td>
</tr>
<tr>
<td>$k_{\text{miner,alco,wat,T0}}$</td>
<td>0.046</td>
<td>0.038</td>
<td>0.187</td>
</tr>
<tr>
<td>$K_{\text{O}_2,\text{ads}}$</td>
<td>0.044</td>
<td>0.038</td>
<td>0.187</td>
</tr>
<tr>
<td>$P_{\text{alco}}$</td>
<td>0.034</td>
<td>0.038</td>
<td>0.187</td>
</tr>
<tr>
<td>$P_{\text{alco,NH}_4}$</td>
<td>0.031</td>
<td>0.038</td>
<td>0.187</td>
</tr>
<tr>
<td>$w_{\text{sed,INORG}}$</td>
<td>0.013</td>
<td>0.038</td>
<td>0.187</td>
</tr>
<tr>
<td>$K_{\text{O}_2,\text{miner}}$</td>
<td>0.009</td>
<td>0.038</td>
<td>0.187</td>
</tr>
<tr>
<td>$K_{\text{O}_2,\text{miner}}$</td>
<td>0.008</td>
<td>0.038</td>
<td>0.187</td>
</tr>
<tr>
<td>$P_{\text{alco,NH}_4}$</td>
<td>0.008</td>
<td>0.038</td>
<td>0.187</td>
</tr>
</tbody>
</table>

Continued on next page
B.3 Results - Model calibration

In the following, the model results, concentrations of phytoplankton, zooplankton, oxygen, nitrate and phosphate in the epilimnion and hypolimnion of Greifensee (Fig. B.2), Lake Zurich (Fig. B.3) and Walensee (Fig. B.4) during the whole simulation period (19/30 years), are shown in comparison to data. The model results represent the best simulation results at the maximum of the non-approximated posterior distribution found by estimating the most influential model parameters for a calibration period of 9 and 20 years.

For Lake Zurich it seems that, in the years where zooplankton was measured as one integrated sample, the model overestimates the zooplankton concentration of the entire lake. It becomes obvious that the dynamic of calculated zooplankton concentrations in the hypolimnion is out-of-phase in many years. This is not true for the zooplankton concentrations of entire Greifensee neither for the epilimnion of Lake Zurich, where the overall annual dynamic of the calculations matches the measurements.

The results for Walensee can be found in Fig. B.4. Similar to Lake Zurich and Greifensee, the results for oxygen in the epilimnion fit the data best. The dynamics in the annual pattern of phosphate in the hypolimnion are not as pronounced as in the measurements. This was a problem also in Greifensee. Additionally, the concentrations of phytoplankton in the hypolimnion and zooplankton are, similar to Greifensee, too low. The averaged zooplankton concentrations in the beginning of the simulation time are rather overestimated as in Lake Zurich.

For all three lakes the oxygen concentrations (for Greifensee mainly in the epilimnion) and partly also phosphate and nitrate concentrations are the output variables best described by the model.
Figure B.2: Simulation results for Greifensee for the years 1987-2005. Data points (lines with markers) and model results (bold lines) for phytoplankton, zooplankton, phosphate, nitrate and oxygen in the whole lake (“lake”, black) and in epilimnion (“epi”, green) and hypolimnion (“hypo”, blue), respectively.
Figure B.3: Simulation results for Lake Zurich for the years 1976-2005. Data points (lines with markers) and model results (bold lines) for phytoplankton, zooplankton, phosphate, nitrate and oxygen in the whole lake (“lake”, black) and in epilimnion (“epi”, green) and hypolimnion (“hypo”, blue), respectively.
Figure B.4: Simulation results for Walensee for the years 1976-2005. Data points (lines with markers) and model results (bold lines) for phytoplankton, zooplankton, phosphate, nitrate and oxygen in the whole lake ("lake", black) and in epilimnion ("epi", green) and hypolimnion ("hypo", blue), respectively.
Chapter 4

Bayesian inference of a lake water quality model with a Gaussian process emulator

Anne Dietzel and Peter Reichert
Submitted to Water Resources Research
Abstract We use a Gaussian stochastic process emulator to interpolate the posterior probability density of a computationally demanding application of the lake model BE-LAMO to accelerate statistical inference of model parameters. This model consists of a mechanistic description of key processes influencing the mass balance of nutrients, dissolved oxygen, organic particles, and phyto- and zooplankton in the lake as well as another Gaussian stochastic process to describe the remaining model bias. A small subsample of the Markov chain representing the posterior of the model parameters is propagated through the model to get uncertainty estimates of model predictions. We expect this approximation to be more accurate at only slightly higher computational costs compared to using a normal approximation to the posterior and linear error propagation to the results as we did in an earlier paper. The performance of the two techniques is compared for a didactical example (for which we can also compare the two techniques to a full Bayesian inference without using an emulator) as well as for the lake model. As expected, for the didactical example, the use of the emulator led to posterior marginals of the model parameters that are closer to those calculated by Markov chain simulations using the full model than to those based on the normal approximation. For the lake model, the new technique proved applicable without an excessive increase in computational requirements (as would be the case for doing Markov chain simulations with the full model) and the results seem to be reasonable. We can thus expect that emulating the posterior probability density for getting a Markov chain sample of the parameters and propagating a subsample to the results is a computationally efficient alternative to using the full model. This technique outperforms a normal approximation to the posterior. In situations in which there is no excessive need for posterior simulations of model results, this technique can thus be an alternative to emulating the full, dynamic model, that is much less difficult to implement and use for arbitrary models.

Keywords: Gaussian stochastic process emulator; Posterior distribution; Lake water quality model; Structurally uncertain model; Model complexity.

4.1 Introduction

Environmental models are often complex, computationally demanding and affected by bias. A general methodology has been developed to deal with these difficulties by emulating the computer code of the model by a conditioned statistical model to reduce the computational demand and by describing bias by a stochastic process (Kennedy and O’Hagan, 2001; Higdon et al., 2004; Bayarri et al., 2007).
4.2. Methods

However, many environmental simulation models are dynamic, which makes these standard emulation techniques based on Gaussian stochastic processes difficult to apply. There are (at least) the following three options to deal with this problem:

1. Extend emulators to cope explicitly with dynamic models (Castelletti et al., 2012). Several approaches have been suggested to do so. Some rely on statistical dynamic models (Liu and West, 2009; Young and Ratto, 2011) or on dynamic models that approximate the mechanisms of the full model (Reichert et al., 2011; Albert, 2012) as a basis for the emulator. Others emulate the time-stepping mechanism of the full model to get an emulator of the time series (Bhattacharya, 2007; Conti and O’Hagan, 2010; Conti et al., 2009).

2. Only use statistical bias description and avoid emulation completely (Reichert and Schuwirth, 2012) or rely on approximations to the posterior if a full Markov chain Monte Carlo technique cannot be done due to computational limitations (Dietzel and Reichert, 2012).

3. Do not emulate the full model, but only the log posterior density of the model for statistical inference and rely on the full model (using a subsample) for prediction.

As dynamic emulation still needs further development and avoiding emulation requires either a large number of model runs or relatively crude approximations to the posterior, we explore the potential of the third of these options in this paper.

This paper is structured as follows. The methodology is described in section 4.2. In the following section 4.3 it is applied to the same didactical example as used by Reichert and Schuwirth, 2012 and Dietzel and Reichert, 2012. This makes a direct comparison possible between the full Bayesian approach, the use of the emulator of the posterior, and the use of a normal approximation to the posterior combined with linearized error propagation to the model results. Section 4.4 describes the application of the methodology to the same lake model case study as used in Dietzel and Reichert, 2012. Finally, we draw our conclusions in section 4.5.

4.2 Methods

The methodology applied in this paper consists of the following steps:

1. Formulating a posterior based on a likelihood function that considers bias in the form of a Gaussian stochastic process. Evaluating this posterior for a design
parameter set spread over a reasonable domain of parameter values. This domain can be chosen as an environment around the numerical approximation to the parameter set that leads to the maximum posterior value. This step is slow, as simulations of the model are required.

2. Conditioning a Gaussian stochastic process emulator to interpolate the log posterior between the points of the design parameter set and estimate interpolation accuracy. This step is fast, as there is no need for further model evaluations. It is also simple to implement as it only requires a very simple, scalar emulator that is independent of the dynamic structure of the model.

3. Applying a Metropolis Markov chain Monte Carlo technique to derive a Markov chain sample of the emulated posterior. This step is again relatively fast, as only the emulator, not the model has to be evaluated. This computational efficiency makes it possible to generate long Markov chains that sample the posterior in a reliable way.

4. Propagating of a subsample of the Markov chain through the model to estimate prediction uncertainty. This step is slow, as it requires the full model to be evaluated, potentially even to a longer time period than used for inference. However, only a small (approximately independent) subsample of the Markov chain can be used for this purpose.

These four steps are discussed in more detail in the following four subsections.

### 4.2.1 Derivation of posterior

As suggested by Craig et al., 1996, Craig et al., 2001, Kennedy and O’Hagan, 2001, Higdon et al., 2004, Bayarri et al., 2007, we describe the observations as the sum of the output of a deterministic model $y_M^L(x, \theta)$, the bias $B_M^L(x, \xi)$ and the observation error $E^L(\psi)$:

$$Y_M^L(x, \theta, \psi, \xi) = y_M^L(x, \theta) + B_M^L(x, \xi) + E^L(\psi).$$

(4.1)

The vector $Y_M^L$ of random variables represents the observations as described by the model $M$ at the observation layout $L$. The observation layout defines the output variables and the time points and locations at which they are observed or evaluated. $Y_M^L$ depends on the external influence factors $x$, unknown model parameters $\theta$ and additional parameters $\psi$ and $\xi$ of the error terms. It is composed of the deterministic function $y_M^L(x, \theta)$, representing our knowledge of the system mechanisms as a function
of model parameters and input, the random process \( B_L (x, \xi) \), expressing our knowledge of model bias depending on external inputs and additional parameters, and of \( E_L (\psi) \), a vector of random variables that represent observation errors and can depend on additional parameters. Equation (4.1) leads to a hierarchical model with the bias, \( B_L (x, \xi) \), as an intermediate variable.

We assume normally distributed observation errors with mean zero and covariance matrix \( \Sigma_{E_L} (\psi) \)

\[
f_{E_L | \psi} (\epsilon^L | \psi) = \frac{1}{\sqrt{2\pi^L}} \frac{1}{\sqrt{\det (\Sigma_{E_L} (\psi))}} \exp \left( -\frac{1}{2} (\epsilon^L)^T \Sigma_{E_L} (\psi)^{-1} \epsilon^L \right)
\]

(4.2)

and a Gaussian stochastic process for the bias with covariance matrix \( \Sigma_{B_L} (\xi, x) \)

\[
f_{B_L | \xi} (b^L | \xi, x) = \frac{1}{\sqrt{2\pi^L}} \frac{1}{\sqrt{\det (\Sigma_{B_L} (\xi, x))}} \exp \left( -\frac{1}{2} (b^L)^T \Sigma_{B_L} (\xi, x)^{-1} b^L \right)
\]

(4.3)

The diagonal of the covariance matrix of the observation error is assumed to consist of

\[
\Sigma_{E_{i,i}} (\psi) = \sigma_{E_{i,i}} (\psi)^2
\]

all other matrix elements are assumed to be zero. The covariance matrix of the bias is assumed to consist of the matrix elements

\[
\Sigma_{B_{i,i,j}} (\xi, x) = \sigma_{B_{i,i}} (\xi) \sigma_{B_{j,j}} (\xi) \exp \left( -\left( \frac{t_i - t_j}{\tau} \right)^2 \right)
\]

(4.5)

depending on points in time \( t_i \) and the correlation time \( \tau \). For output variables \( y^L_i \) and \( y^L_j \) of different type, the matrix elements are assumed to be zero.

Under these assumptions, and using the joint prior, \( f_{\Theta, \Psi, \Xi} \), of the model parameters, \( \Theta \), of the bias parameters, \( \Psi \), and of the observation error parameters, \( \Xi \), their posterior for given data for the observation layout \( L_1 \) is given by

\[
f_{\Theta, \Psi, \Xi | Y^L_{M}} (\theta, \psi, \xi | y^L_{1,1}, x) \propto f_{\Theta, \Psi, \Xi} (\theta, \psi, \xi) \frac{1}{\sqrt{\det (\Sigma_{E_{i,j}} + \Sigma_{B_{i,j}})}} \cdot \exp \left( -\frac{1}{2} [y^L_{1} - y^L_{M} (x, \theta)]^T (\Sigma_{E_{i,j}} + \Sigma_{B_{i,j}})^{-1} [y^L_{1} - y^L_{M} (x, \theta)] \right)
\]

(4.6)

(see Reichert and Schuwirth, 2012 and Dietzel and Reichert, 2012 for more details).

To account for heteroscedasticity in model output, as in Dietzel and Reichert, 2012, we applied the technique to Box-Cox transformed data and model results (Box and Cox, 1964; Box and Cox, 1982).
4.2.2 Gaussian stochastic process emulator

In a previous study (Dietzel and Reichert, 2012), a normal distribution was used to approximate the posterior (4.6) at its maximum to reduce the computational demand resulting from the evaluation of the deterministic model, $y^{L^i}_M(x, \theta)$, during the inference process. This approximating normal distribution was constructed based on evaluations of the posterior for a parameter sample from a normal or uniform ball distribution around the maximum of the posterior. This maximum was estimated before using a global optimization algorithm (Dietzel and Reichert, 2012).

In this paper, we are trying to improve this approximation by emulating the log of the posterior instead of approximating it by a parameterized (in our case normal) distribution. As the emulator interpolates between the evaluations of the log posterior for a given design data set, we can expect a better representation of its shape than what we got with the normal distribution. As a basis for conditioning the emulator, we draw a random sample, $\gamma^D = (\gamma^D, \gamma^D, \ldots, \gamma^D)$, of size $n_D$ from a multivariate normal distribution around the maximum of the posterior as a design parameter set. Here, $\gamma = (\theta, \psi, \xi)$, is used as an abbreviation for all parameters. Using the same notation for the corresponding random variables, $\Gamma = (\Theta, \Psi, \Xi)$, this design data set can be written in the form

$$D = (\gamma^D, f^D) = \left(\gamma^D, f_{\Gamma|Y^L_M}(\gamma^D)\right)$$

with the posterior densities $f^D = f_{\Gamma|Y^L_M}(\gamma^D)$ calculated according to equation (4.6).

The log of the posterior probability density given by equation (4.6) is to be interpolated between the values at the design parameter set by a Gaussian stochastic process emulator (O’Hagan, 2006) for fast statistical inference. This can be done as described in Reichert et al., 2011. For this special case of a univariate emulator, the expected value $\log f$ of the multivariate normal distribution describing the GAussian Stochastic Process (GASP) emulator conditioned on the design data set, is given by:

$$\log f(\gamma, D) = \log(f_M)(\gamma) + k^{(n_D)}(\gamma, \gamma^{D,1}, \ldots, \gamma^{D,n_D})^T$$

$$\cdot \left(K^{(n_D)}(\gamma^{D,1}, \ldots, \gamma^{D,n_D})^{-1} \cdot \begin{pmatrix} \log(f^{D,1}) - \log(f_M(\gamma^{D,1})) \\ \vdots \\ \log(f^{D,n_D}) - \log(f_M(\gamma^{D,n_D})) \end{pmatrix} \right). \quad (4.8)$$

In this equation, $\log(f_M)$ represents the simplified model of the log posterior underlying the emulator. For the application to a didactical example, we use the log values of the normal density approximating the posterior distribution resulting from the importance sampling technique described in Dietzel and Reichert, 2012 as the deterministic
4.2. Methods

simplified model (shifted by a constant to account for the lack of normalization of the posterior density). For the lake water quality model application, we used the mean of the log posterior density values of the design data set. Specifically, the values \( f_{M'}(\gamma^D) \) represent the results of the simplified model of the posterior at the design parameter vectors, while \( f^D \) give the exact posterior values (see above). The matrix \( K^{(n)} \) defines the covariance structure of the unconditioned Gaussian process for an arbitrary set of \( n \) parameter vectors:

\[
K^{(n)}(\gamma^1, \ldots, \gamma^n) = 
\begin{pmatrix}
K(\gamma^1, \gamma^1) & \cdots & K(\gamma^1, \gamma^n) \\
\vdots & \ddots & \vdots \\
K(\gamma^n, \gamma^1) & \cdots & K(\gamma^n, \gamma^n)
\end{pmatrix}
\]

(4.9)

with

\[
K(\gamma^i, \gamma^j) = \sigma^2 \cdot \exp \left( -\sum_{k=1}^{n_\gamma} \left( \frac{|\gamma^i_k - \gamma^j_k|}{\lambda_k} \right)^\alpha \right).
\]

(4.10)

Equation (4.10) defines the correlation structure of the covariance matrix and defines the standard deviation \( \sigma \) of the prior Gaussian process. For the applications in this paper, we consider only one output of the "true" model, the log of the posterior probability density. As we only consider the mean of the conditioned Gaussian process to interpolate the posterior probability density, the value of \( \sigma \) does not influence the results (it would influence, however, the variance of the predicted log posterior density). The parameters \( \lambda_k \) are the correlation lengths for each model parameter \( \gamma_k \), \( n_\gamma \) is the number of model parameters, in our case parameters of the deterministic model and of the error model. Generally, the larger the correlation lengths, the smoother the interpolation by the emulator. Additionally, the correlation matrix depends on the exponent \( \alpha \). Only for \( \alpha = 2 \) the mean of the emulator is (infinitely often) differentiable at the design data points. As this choice is susceptible to oscillations of the mean of the emulator, we use values of \( \alpha \) between 1 and 2 and compare the quality of emulation for the didactical example. In equation (4.8), the matrix \( K^{(n_D)} \) only depends on the design parameter set and hence can be calculated according to equation (4.9) and inverted already during the setup of the emulator. The same is true for the last factor in the same equation. The value of \( f_{M'}(\gamma) \) is given by the result of the last factor in the same equation. The vector \( k \) is the only other component of equation (4.8) which depends on the new parameter values. It is given by:

\[
k^{(n)}(\gamma, \gamma^1, \ldots, \gamma^n) = 
\begin{pmatrix}
K(\gamma, \gamma^1) \\
\vdots \\
K(\gamma, \gamma^n)
\end{pmatrix}
\]

(4.11)

with components as given by equation (4.10).
4.2.3 Markov chain Monte Carlo sampling based on emulator

The emulator conditioned to the design data as described in section 4.2.2 is used to calculate a Markov chain sample of the (approximate) posterior calculated by $\log \tilde{f}(\gamma, D)$ according to equation (4.8) using a Metropolis algorithm (Gelman et al., 1995; Gemanman, 1997). The jump distribution is optimized by 4 iterations with shorter Markov chains before the final Markov chain is calculated resulting in a sample of the posterior. For visualization of the results, a density estimator is applied to this sample to provide estimates of posterior marginal distributions of the parameters $\gamma$.

4.2.4 Model predictions

Model predictions for new observations (sum of the deterministic model output, bias and observation error) are distributed according to

$$f_{Y_{L_2}^M|Y_{L_1}^M} (y_{L_2}, x) = \frac{1}{\sqrt{2\pi}^n} \int \frac{f_{\Theta, \Psi, \Xi|Y_{M}^L}(\theta, \psi, \xi | Y_{L_1}^M, x)}{\sqrt{\det(\text{Var}[Y_{L_2}^M|Y_{L_1}^M, \Theta, \Psi, \Xi])}} \cdot \exp \left( -\frac{1}{2} \left( y_{L_2} - E[Y_{M}^L|Y_{L_1}^M, \Theta, \Psi, \Xi] \right)^T \cdot \text{Var}[Y_{L_2}^M|Y_{L_1}^M, \Theta, \Psi, \Xi]^{-1} \cdot \left( y_{L_2} - E[Y_{M}^L|Y_{L_1}^M, \Theta, \Psi, \Xi] \right) \right) \, d\theta \, d\psi \, d\xi \quad (4.12)$$

with

$$E[Y_{M}^L|Y_{L_1}^M, \Theta, \Psi, \Xi] = y_{M}^L(x, \theta) + (\Sigma_{E_{L_1}} + \Sigma_{B_{M}})_{L_2,L_1}^{-1} \cdot (y_{L_1}^M - y_M^L(x, \theta)) \quad (4.13)$$

and

$$\text{Var}[Y_{M}^L|Y_{L_1}^M, \Theta, \Psi, \Xi] = \Sigma_{E_{L_2}} + \Sigma_{B_{M}} - (\Sigma_{E_{L_1}} + \Sigma_{B_{M}})_{L_2,L_1} \cdot (\Sigma_{E_{L_1}} + \Sigma_{B_{M}})^{-1} \cdot (\Sigma_{E_{L_1}} + \Sigma_{B_{M}})_{L_2,L_1}^T \quad (4.14)$$

(Reichert and Schuwirth, 2012). In these equations, $L_1$ refers to the observation layout (variables and points in time and space) used for calibration, $L_2$ to the layout for which results are to be predicted (see Reichert and Schuwirth, 2012 for more details). From these results, approximated uncertainty bounds can be derived for the results of the
4.3 Didactical example

deterministic model considering the contribution of uncertainty due to model bias and observation error. Omitting the terms $\Sigma_{E^{L_2}}$ and $(\Sigma_{E^{L_1,2}})_{L_2,L_1}$ in equations (4.13) and (4.14), results in predictions of the “true” system output before observation (results of the deterministic model plus bias) in equation (4.12).

Numerically, a sample from this distribution is drawn by first choosing a subsample (for increasing the numerical efficiency) of the Markov chain derived as described in section 4.2.3. Then, for each parameter vector, $(\theta, \psi, \xi)$, of this subsample, it is drawn from the normal distribution defined by the equations (4.13) and (4.14) while either using or omitting the terms $\Sigma_{E^{L_2}}$ and $(\Sigma_{E^{L_1,2}})_{L_2,L_1}$. This results in sampling either from $Y_{M}^{L_2} \mid Y_{M}^{L_1}$ or from $y_{M}^{L_2} + B_{M}^{L_2} \mid Y_{M}^{L_1}$, respectively.

The emulation technique was implemented with the statistics software R (http://www.r-project.org).

4.3 Didactical example

We use the same microbial growth model as used in [Reichert and Schuwirth, 2012] and [Dietzel and Reichert, 2012] as a simple example for the application of the suggested emulation approach and for an informative comparison to the calibration and uncertainty analysis approaches described in the other two papers.

4.3.1 Model description, observation data and prior distributions

This model describes microbes’ concentration $C_M$ and substrate concentration $C_S$ by the following mass balance equations:

$$\frac{dC_M}{dt} = \mu \frac{C_S}{K + C_S} C_M \exp(c(T - T_0)) - b C_M - q C_M$$

$$\frac{dC_S}{dt} = -\frac{\mu}{Y} \frac{C_S}{K + C_S} C_M \exp(c(T - T_0)) + q (C_{S,in} - C_S)$$

with the initial conditions

$$C_M(0) = C_{M,ini}$$

$$C_S(0) = C_{S,ini}$$

and the temperature variations

$$T = T_0 + A \sin(2\pi (t - 0.25d))$$

(4.15)
In these equations, \( C_M \) is the concentration of microbes in the reactor (ML\(^{-3}\)), \( C_S \) is the concentration of substrate in the reactor (ML\(^{-3}\)), \( t \) is time (T), \( \mu \) is the maximum specific growth rate of the microbes (T\(^{-1}\)), \( K \) is the concentration of substrate at which the microbes grow with half of their maximum specific growth rate (ML\(^{-3}\)), \( b \) is the specific death rate of the microbes (T\(^{-1}\)), \( q \) is the volumetric flow rate through the reactor per unit of reactor volume (T\(^{-1}\)), \( Y \) is the yield (produced amount of biomass per consumed amount of substrate) (–), \( C_{S,\text{in}} \) is the concentration of substrate in the inflow (ML\(^{-3}\)) (the concentration of microbes in the inflow is assumed to be zero), \( T \) refers to temperature, \( T_0 \) is a reference temperature assumed to be 20°C, \( A \) is the amplitude of temperature variations, \( c \) is a temperature dependence coefficient assumed to be 0.046 per degree Celsius, \( C_{M,\text{ini}} \) is the initial concentration of microbes in the reactor (ML\(^{-3}\)), and \( C_{S,\text{ini}} \) is the initial concentration of substrate in the reactor (ML\(^{-3}\)).

With this model, we could produce synthetic observations for a simulation period of 4 days, an amplitude of temperature variation of 4°C and the following parameter values: \( \mu = 4 \), \( K = 10 \), \( b = 1 \), \( q = 1 \), \( Y = 0.6 \), \( C_{S,\text{in}} = 100 \), \( C_{M,\text{ini}} = 10 \), \( C_{S,\text{ini}} = 40 \), and \( \sigma_{E_{C_M}} = 0.5 \) and \( \sigma_{E_{C_S}} = 0.5 \) (in model output units) as standard deviations of normally distributed observation errors in both model output variables. As the generated data does not contain heteroscedasticity, no Box-Cox transformation was conducted.

The first two days out of these data were used to estimate the parameters of a model that describes the same system but contains structural error due to neglecting the temperature dependence of substrate and microbial growth as described in equations (4.15). The vector \( \theta = (\mu, K, b, q, Y, C_{S,\text{in}}, C_{M,\text{ini}}, C_{S,\text{ini}})^T \) gives the parameter vector of the deterministic model. The parameters \( \psi = (\sigma_{E_{C_M}}, \sigma_{E_{C_S}})^T \) consist of the standard deviations \( \sigma_{E_{C_M}} \) and \( \sigma_{E_{C_S}} \) of independent normal distributions characterizing the observation error of \( C_M \) and \( C_S \) at all points in time. Finally, the vector \( \xi = (\sigma_{B_{C_M}}, \sigma_{B_{C_S}}, \tau)^T \) combines the standard deviations \( \sigma_{B_{C_M}} \) and \( \sigma_{B_{C_S}} \) for the bias in each of the model variables \( C_M \) and \( C_S \) and the correlation time \( \tau \). The priors of \( \sigma_{B_{C_M}} \) and \( \sigma_{B_{C_S}} \) were chosen as normal distributions with mean zero and a standard deviation of 0.5. They were truncated at zero to avoid negative values. For the parameters \( \theta \) of the deterministic model, we used independent lognormal priors with the means at the correct values given above and standard deviations of 50%. For the parameters \( \psi \) and the correlation time \( \tau \), lognormal priors with the means at the correct values in case of the observation error parameters (given above) and at 0.3 for \( \tau \) and with standard deviations of 10% were used. Differently to the application in [Dietzel and Reichert, 2012], it was not needed to calibrate the logarithm of all parameters instead of the parameters themselves.
4.3. Didactical example

For more details about the didactical example we refer to [Reichert and Schuwirth, 2012](#) and [Dietzel and Reichert, 2012](#). For the emulation of the log of the posterior probability of the microbial growth model and its error model, a design data set with a set of 1000 design parameter combinations and the respective log posterior density calculated according to equation (4.6) was used. All parameters $\gamma = (\theta, \psi, \xi)$ were used. The design parameter set was randomly drawn from the normal distribution representing a simple approximation to the posterior distribution of the model that was derived from one importance sampling from a ball and 10 subsequent iterative importance samplings from a normal distribution as described in [Dietzel and Reichert, 2012](#). The emulator was conditioned on this design data set and the same normal distribution was used as a base model for the emulator. We ran a Markov chain Monte Carlo sampling with this emulator. Four iterations with chains of length 10000 were run for the optimization of the jump distribution, the final chain was calculated with a length of 100000. The chain was thinned by a factor of 5 and every 100th element of the resulting chain was propagated through the microbial growth model for the predictions of model results and derivation of uncertainty bounds as described in section 4.2.4 i.e. 200 simulations with the microbial growth model were needed.

We evaluated the performance of the emulator for different setups by applying the emulator to additional new input data also drawn from the basic normal distribution and again calculating the real log posterior value by applying equation (4.6) to the new parameter set and comparing the results to the emulator results.

4.3.2 Emulator setup and performance

In this section we discuss aspects of the most suitable setup of the emulator, mainly focused on the choice of the exponent $\alpha$, the correlation lengths $\lambda_k$ (see equation (4.10)), the design data set and the simplified model underlying the emulator. Some problems appeared during the setup of the emulator and the subsequent calculation of a Markov chain Monte Carlo sample with the presented emulator giving the result of the log of the posterior density of the microbial growth model and its error model. A first difficulty arose from the choice of a suitable design data set. One critical aspect is the size of the design data set. A size of 100 parameter samples and the resulting log posterior densities led to numerical problems already in the initialization of the emulator during the calculation of the covariance matrix. Although the didactical example is relatively simple, it is comprised of a parameter space with already 13 dimensions. This makes an interpolation in all dimensions difficult if only few points are available. A design data
set of size 1000 appeared to be adequate. Furthermore, tests with a parameter sample from a uniform ball in the neighborhood of the maximum of the posterior failed. We used such samples in earlier studies (Reichert and Schuwirth, 2012) for the start of an importance sampling technique aiming for an approximation of the posterior by a normal distribution (Dietzel and Reichert, 2012). With this sample, it was possible to initialize the emulator and train it with this design data set. However, the calculation of a Markov chain failed, typically because the rejection frequency was too large, while the chains for each parameter stayed within a small range which led to unrealistically small standard deviations for all parameters. This is an indication that, due to the "volume effect", only few points of the design data set are close to the posterior maximum for a uniform sample in a high-dimensional ball. Hence, the emulator provided a poor approximation to the log posterior with local maxima close to the design points but poor emulation in between. It became obvious that a design parameter set drawn from a normal distribution with high densities around the maximum of the posterior is much more appropriate and avoids this effect. Finally, we used a random parameter sample of size 1000 from the normal distribution achieved from iterative importance sampling for an approximation to the posterior in an earlier study (Dietzel and Reichert, 2012). The same normal distribution was used as the underlying model of the emulator. In principle, any other simple model would work as well if the correlation lengths are chosen adequately, but the normal distribution including a maximum should approximate the shape of the posterior much better than, for example, a linear model.

As a second step, we aimed for a good performance of the emulator by choosing the optimal combination of exponent $\alpha$ and correlation lengths $\lambda_k$. We did so by minimizing the root mean square error between the emulator interpolating between the design data points for a new set of parameters and the true posterior density of the same parameter set. As we are not limited by computational time for the didactical example, we were able to explore the results of the method for different emulator setups compared to the results of the real model. For that purpose, we drew 1000 parameter samples from the approximate normal distribution and calculated the log posterior densities for those parameter combinations, once with the emulator trained on a different design data set and once by the true model according to equation (4.6). The root mean square errors (RMSE) could be calculated between both results and be compared for different combinations of $\alpha$ and $\lambda$. For the exponent $\alpha$, we tested seven discrete values between 1 and 2. For the 13 correlation lengths $\lambda_k$ for each parameter, we only tested different correlation lengths factors ($\lambda_{fact}$) between 0.0001 and 75 which were multiplied with the
4.3. Didactical example

difference between the maximum and the minimum parameter values contained in the design parameter set resulting in correlation lengths for each parameter relative to the range of each parameter within the design set. Fig. 4.1 shows the RMSEs for all tested correlation length factors for $\alpha$ equal 1, 1.5, 1.9 and 2.

Figure 4.1: Performance of the GASP emulator for different exponents $\alpha$ or the correlation structure and depending on chosen correlation length factors.

It is important to point out that the results depend on the random choice of new input parameter sets for which the emulator and the true log posterior are calculated. That slightly changes the results, while the overall pattern stays the same, as we saw in several replicates of the same calculation. In general, Fig. 4.1 indicates that for small correlation lengths the root mean square error between emulation and true value decreases with increasing correlation lengths. This is meaningful, as a larger correlation should smooth the interpolation between design data points and decrease the influence of the underlying model. However, there seems to be an optimum for the correlation lengths, after which the RMSEs increases again for even larger correlation lengths.
Additionally, we can see that for increasing values of $\alpha$ the RMSE is decreasing more strongly. We have to emphasize that an emulation with $\alpha = 2$ in combination with large correlation lengths leads to the problem that the inversion of the covariance matrix $K^{(nD)}$ (see equation (4.8)) leads to numerical difficulties. All in all, the smallest RMSE can be achieved with $\alpha = 2$ and $\lambda_{\text{fact}} = 5$. In cases where a larger correlation length factor did not lead to numerical problems, there could not be accomplished a smaller RMSE. The first tries with the emulator were therefore done with $\alpha = 2$ and $\lambda_{\text{fact}} = 5$. However, for this combination the Markov chain Monte Carlo sampling resulted in diverging chains causing bimodal and sometimes unrealistic wide posterior parameter distributions. This also caused unrealistic best estimates for the model parameters leading to poor model predictions. A possible explanation for these phenomena is that for $\alpha = 2$ the domains of good emulation and of poor emulation due to oscillations are very close to each other regarding the $\lambda_k$ and not even clearly distinguishable in the space at emulated parameters. Although $\alpha = 1.9$ and $\lambda_{\text{fact}} = 1$ shows a slightly higher RMSE, it seemed to be a better choice for the described technique due to its higher robustness against oscillations. Similarly good results could even be achieved with $\alpha = 1$ and $\lambda_{\text{fact}} = 1$.

### 4.3.3 Results and discussion

In the following, we show the results for the calibration and uncertainty analysis of the microbial growth model with the help of the GASP emulator as introduced in section 4.2. We focus on the results for $\alpha = 1.9$ and $\lambda_{\text{fact}} = 1$. Fig. 4.2 shows the marginal prior and posterior distributions of all model parameters. The posterior distributions are derived from the Markov chain with 100000 elements based on the calculation of the log of the posterior by the GASP emulator. They are compared with the approximation by a normal distribution (in that case their logarithm was approximated by a normal distribution) as shown in Dietzel and Reichert, 2012 and by a Markov chain Monte Carlo technique based on the original model as presented in Reichert and Schuwirth, 2012.

In general, all marginal distributions derived from emulation show a better accordance with the full MCMC method than the ones of the normal approximation. Moreover, the best estimates of the parameters represented by the mode of the posterior distribution are closer to the true parameter values (vertical lines). Again, this indicates an improved posterior approximation by the GASP emulator than by the normal distribution. The results for the prediction as described in section 4.2.4 are presented in Fig. 4.3 again as a comparison of the three different prediction approaches. For the GASP emulator
4.3. Didactical example

Figure 4.2: Prior (dotted, supported by light grey shading of the areas below the lines) and posterior (supported by dark grey shading of the areas below the lines) marginals of the model parameters by a Markov chain Monte Carlo method based on the GASP emulator (long-dashed), posterior approximation by a normal distribution from importance sampling (short-dashed) (Dietzel and Reichert, 2012) and a Markov chain Monte Carlo method with the original model (solid) (Reichert and Schwirth, 2012). Vertical lines indicate the true values of the model parameters used for producing the synthetic data (there are no true values for the parameters of the bias).
approach, every 500th element of the full Markov chain (every 100th element of the thinned chain), likely to represent a random sample, was propagated through the model (compared to the full MCMC method for which the whole Markov chain was used).

Overall, the results for the three different approaches are very similar. As the best parameter estimates for the Markov chain on basis of the emulator are closer to the full Markov chain Monte Carlo results and to the true values, the deterministic model results and the bias-corrected output are also closer to the full Markov chain results than for the posterior approximation by a normal distribution. The 95% credibility intervals for model predictions resulting from parameter uncertainty, bias and observation error as well as for the bias and of the observation error of the two output variables $C_M$ and $C_S$ are very similar between the three approaches. Due to the smaller subsample of the Markov chain propagated through the model in case of the emulator application, the boundaries of the credibility intervals are slightly less smooth than for the other two approaches.

### 4.4 Application to biogeochemical and ecological lake model

We applied the emulation technique described in section 4.2 also to a real case study, the calibration of the lake model BELAMO to long-term simulations of Lake Zurich. Due to its high computational demand, a full Markov chain Monte Carlo sampling is not applicable for this model. The possibility to calibrate this model and estimate uncertainties by a normal approximation of the posterior distribution to reduce the computational demand was explored in an earlier study (Dietzel and Reichert, 2012). The posterior approximation by a normal distribution seemed to be a crude simplification and the need to gain more information about the shape of the posterior distribution appeared. This is done by the application of a GASP emulator in the next section.

#### 4.4.1 BELAMO: model description

The Biogeochemical and Ecological LAke MOdel (BELAMO), implemented in AQUASIM (version 2.1f, [http://www.aquasim.eawag.ch](http://www.aquasim.eawag.ch)), a computer program for the identification and simulation of aquatic systems (Reichert, 1994; Reichert, 1998), describes concentrations of nutrients, oxygen, organic particles, phyto- and zooplankton in a lake. It is based on differential equations formulating mass balances. Originally, it was im-
4.4. Application to biogeochemical and ecological lake model

Figure 4.3: Results of the median and 95% credibility intervals for both variables $C_M$ and $C_S$ (top row; observations used for calibration are marked by solid symbols, those not used for calibration by open symbols; output of the deterministic model (long-dashed line), bias-corrected output (deterministic model plus bias; median as solid line, 95% credibility bounds as dark grey area bounded by dashed lines), and prediction for new observations (including observation error; median as solid line (same as for bias-corrected output), 95% credibility bounds shown as dark and light grey areas bounded by dotted lines)), results of the median and 95% credibility intervals for the bias (solid line and dark grey areas bounded by dashed lines) and of the observation error (dots and vertical line segments) for the variables $C_M$ (middle row) and $C_S$ (bottom row). The left column shows the results for a Markov chain Monte Carlo method based on the GASP emulator, the middle column by posterior approximation by a normal distribution from importance sampling (Dietzel and Reichert, 2012) and the right column the results of the full MCMC method (Reichert and Schuwirth, 2012).
implemented as a 1d model for Lake Zurich (Omlin et al., 2001b), then analyzed for its transferability to the lakes Greifensee and Walensee (Mieleitner and Reichert, 2006) and subsequently spatially simplified to a 4-box version for an analysis of functional phytoplankton groups (Mieleitner and Reichert, 2008). Dietzel and Reichert, 2012 applied a calibration technique to BELAMO that makes it possible to estimate the prediction uncertainty of structurally uncertain and computationally demanding models, while considering and quantifying the contribution of model bias. This study was only done for Lake Zurich, but was followed by an uncertainty estimation and joint calibration of the lakes Greifensee, Lake Zurich and Walensee (Dietzel et al., 2012). Dietzel et al., 2012 also changed aspects of the model to reduce the bias to use the model for long-term simulations of the three lakes during a time with decreasing input loads.

The current version of BELAMO calculates concentrations of ammonium, nitrate, phosphate, oxygen, degradable and inert dead organic particles and the total biomass of phytoplankton and zooplankton in the epilimnion, hypolimnion and two sediment compartments of the lakes. The four model compartments are considered as well-mixed boxes, the model accounts for inflow to and outflow from the water column, gas exchange between the water column and the atmosphere, mixing between the two water compartments, sedimentation of particles, advection and diffusion of dissolved substances in the water column and in the pore water of the sediment compartments, sediment accumulation and permanent burial. Within the model compartments, the model includes the description of growth, respiration and death of phyto- and zooplankton, aerobic, anoxic and anaerobic mineralization, nitrification, methane oxidation and phosphate adsorption to sinking degradable and inert organic particles.

To learn about lake biogeochemistry and ecology from a model like BELAMO or to even use it for environmental decision support in the area of water quality and plankton dynamics, it is needed to assess the uncertainty of model predictions. As stated by Dietzel and Reichert, 2012 the introduced lake model is complex, though a highly simplified description of the real system which causes a considerable amount of bias, and simulations are time-consuming. During the development of the model and while conducting different application studies, it became obvious that the calibration of the model is a non-trivial problem. This is especially true for the application to a long-term period of up to 30 years of simulation time and for a joint calibration of three lakes differing in their trophic regime, while trying to use only few lake-specific parameters (Dietzel et al., 2012). The large contribution of bias to the model results, which is even increased by considering the bias to be the same for each output variable across lakes, increases this calibration problem. Time-consuming and difficult calibrations in the high-dimensional parameter and output spaces brought up the idea to use an emulator for
a better approximation of the posterior distribution. This would serve for exploring the shape of the posterior distribution, which was only coarsely approximated by a normal distribution before \cite{DietzelReichert2012, DietzelReichertFreyforDietzel2012}. In the following, we apply the emulator described in section \ref{sec:emulator} to the simulation of long-term observations of Lake Zurich with BELAMO to estimate the model output uncertainty and calibrate the model for a better description of the observations.

\subsection*{4.4.2 Study area}

Lake Zurich is a deep mesotrophic lake in the north-eastern part of the Swiss plateau. It is separated into two parts by a natural dam, lower Lake Zurich, the larger and deeper part of the two, served as a study lake. In lower lake Zurich, measurements were taken at the location of largest depth (136 m). The watershed of the lake is mainly used for settlements and to a smaller amount for agriculture. More detailed descriptions of the properties of the lake can be found in \cite{MieleitnerReichert2006, DietzelReichertFreyforDietzel2012} both show a map of the study area as well.

\subsection*{4.4.3 Observation data}

Observations from Lake Zurich of the years 1976 to 2005 (zooplankton only since 1985) serve as validation data and for the calculation of the posterior distribution estimated by our emulator. Ammonium, nitrate, phosphate, oxygen and temperature were measured as profiles, consisting of samples in 19 different depths. Phytoplankton data consists of samples in 14 different depths, while zooplankton was only measured as two integrated samples over the whole lake depth. These data were collected monthly by the Water Supply Authority of Zurich (Wasserversorgung Zürich, WVZ). Information about the nutrient input loads from inflowing rivers, waste water treatment plants and wet deposition, as well as data about meteorological forcing could be obtained from federal and cantonal agencies and from technical reports. For a detailed description of data compilation and processing we refer to \cite{DietzelReichertFreyforDietzel2012} specifically to the supporting information to this paper.

We assume that observation and structural errors are smaller for smaller concentrations of chemical and biological variables. To apply the emulation of the posterior to our chosen error model and to account for this heteroscedasticity of the data, a Box-Cox transformation of data and model results was performed \cite[section 4.2.1]{BoxCox1964}. For the shown application, the transformation parameters were chosen as $\lambda_1 =$
0.5 and $\lambda_2 = 0$.

### 4.4.4 Prior distributions

The prior distributions and the descriptions of the parameters of the deterministic model as well as of the error model are shown in Tables 4.1–4.3.

We assumed lognormal priors for the standard deviations of the observation errors representing the parameter vector $\psi$. These values are listed in Table 4.2 in Box-Cox transformed units.

Table 4.3 contains the priors for the standard deviations of the bias in Box-Cox transformed units. We chose exponential distributions for which the probability density increases with decreasing value of the standard deviation. This reflects our desire to avoid bias if possible. The listed priors were used for the calculation of the posterior densities according to equation (4.6) at the design data points with which the emulator was trained.

### 4.4.5 Emulator setup

A similar setup of the emulator was chosen for the lake model as for the didactical example. The emulator parameters were chosen as $\alpha = 1.9$ and $\lambda_{fact} = 1$, as this appeared a reasonable choice in the application to the didactical example. As a design data set, we draw a random sample of 2000 parameter combinations from an uncorrelated normal distribution with the mean at the maximum of the posterior that was derived from an earlier calibration (Dietzel and Reichert, 2012). A simple constant model, the mean of the log posterior densities within the design data set, was used as the underlying model for the emulator. With this, it was possible to calculate a Markov chain of length 100000, after four iterations of adjusting the jump distribution with Markov chains of length 10000. To derive prediction uncertainty bands, every 20th element from the final Markov chain was propagated through the model, which resulted in 1000 simulations with the full lake model.

### 4.4.6 Results and discussion

For the emulator setup described in the previous section, Fig. 4.4 shows the marginal prior and posterior distributions of all lake model and error model parameters. The
Table 4.1: Prior marginals for the calibrated parameters $\theta$ of the deterministic model.

<table>
<thead>
<tr>
<th>Name</th>
<th>Unit</th>
<th>Distribution</th>
<th>Mean</th>
<th>StDev</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_{\text{gro,ALG},T_0}$</td>
<td>$d^{-1}$</td>
<td>Lognormal</td>
<td>1.79</td>
<td>0.5</td>
<td>Growth rate of phytoplankton at reference temp./sat. light intensity</td>
</tr>
<tr>
<td>$k_{\text{death,ALG},T_0}$</td>
<td>$d^{-1}$</td>
<td>Lognormal</td>
<td>0.0721</td>
<td>0.05</td>
<td>Death rate of phytoplankton at reference temperature</td>
</tr>
<tr>
<td>$k_{\text{gro,ZOO},T_0}$</td>
<td>$d^{-1}$</td>
<td>Lognormal</td>
<td>0.397</td>
<td>0.25</td>
<td>Growth rate of zooplankton at reference temperature</td>
</tr>
<tr>
<td>$k_{\text{death,ZOO},T_0,\text{zh}}$</td>
<td>$d^{-1}$</td>
<td>Lognormal</td>
<td>0.0675</td>
<td>0.05</td>
<td>Death rate of zooplankton at reference temperature</td>
</tr>
<tr>
<td>$K_{\text{Feed}}$</td>
<td>gDM $m^{-3}$</td>
<td>Lognormal</td>
<td>0.295</td>
<td>0.3</td>
<td>Threshold phytopl. conc. when zoopl. feeding switches to Monod limit.</td>
</tr>
<tr>
<td>$f_{X_1,\text{rivers (transformed)}}$</td>
<td>–</td>
<td>Normal</td>
<td>2</td>
<td>1</td>
<td>Inert fraction of allochthonous organic particles</td>
</tr>
<tr>
<td>$k_{\text{miner,aero,\text{sed},T_0,\text{zh}}}$</td>
<td>$d^{-1}$</td>
<td>Lognormal</td>
<td>0.295</td>
<td>0.3</td>
<td>Aerobic mineralization rate at reference temperature</td>
</tr>
<tr>
<td>$k_{\text{miner,anae,\text{sed},T_0,\text{zh}}}$</td>
<td>$d^{-1}$</td>
<td>Lognormal</td>
<td>0.295</td>
<td>0.3</td>
<td>Anaerobic mineralization rate at reference temperature</td>
</tr>
<tr>
<td>$k_{\text{miner,anox,\text{sed},T_0,\text{zh}}}$</td>
<td>$d^{-1}$</td>
<td>Lognormal</td>
<td>0.295</td>
<td>0.3</td>
<td>Anoxic mineralization rate at reference temperature</td>
</tr>
</tbody>
</table>

Table 4.2: Prior marginals for the parameters $\psi$ of the error model (on a Box-Cox transformed scale).

<table>
<thead>
<tr>
<th>Name</th>
<th>Unit</th>
<th>Distribution</th>
<th>Mean</th>
<th>StDev</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_{E,\text{ALG}}$</td>
<td>gWM$^{1/2}$ m$^{-3/2}$</td>
<td>Lognormal</td>
<td>0.07</td>
<td>0.02</td>
<td>Standard deviation of measurement error in phytoplankton</td>
</tr>
<tr>
<td>$\sigma_{E,ZOO}$</td>
<td>gWM$^{1/2}$ m$^{-3/2}$</td>
<td>Lognormal</td>
<td>0.08</td>
<td>0.02</td>
<td>Standard deviation of measurement error in zooplankton</td>
</tr>
<tr>
<td>$\sigma_{E,\text{NO}_3}$</td>
<td>gN$^{1/2}$ m$^{-3/2}$</td>
<td>Lognormal</td>
<td>0.022</td>
<td>0.002</td>
<td>Standard deviation of measurement error in nitrate</td>
</tr>
<tr>
<td>$\sigma_{E,O_2}$</td>
<td>gO$^{1/2}$ m$^{-3/2}$</td>
<td>Lognormal</td>
<td>0.11</td>
<td>0.01</td>
<td>Standard deviation of measurement error in oxygen</td>
</tr>
<tr>
<td>$\sigma_{E,HPO_4}$</td>
<td>gP$^{1/2}$ m$^{-3/2}$</td>
<td>Lognormal</td>
<td>0.007</td>
<td>0.0002</td>
<td>Standard deviation of measurement error in phosphate</td>
</tr>
</tbody>
</table>
Chapter 4. Bayesian inference with a Gaussian process emulator

Table 4.3: Prior marginals for the calibrated parameters of $\xi$ of the bias (on a Box-Cox transformed scale).

<table>
<thead>
<tr>
<th>Name</th>
<th>Unit</th>
<th>Distribution</th>
<th>Mean</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_{B,ALG}$</td>
<td>gWM$^{1/2}$ m$^{-3/2}$</td>
<td>Exponential</td>
<td>0.3</td>
<td>Standard deviation of bias in phytoplankton</td>
</tr>
<tr>
<td>$\sigma_{B,ZOO}$</td>
<td>gWM$^{1/2}$ m$^{-3/2}$</td>
<td>Exponential</td>
<td>0.4</td>
<td>Standard deviation of bias in zooplankton</td>
</tr>
<tr>
<td>$\sigma_{B,NO3}$</td>
<td>gN$^{1/2}$ m$^{-3/2}$</td>
<td>Exponential</td>
<td>0.11</td>
<td>Standard deviation of bias in nitrate</td>
</tr>
<tr>
<td>$\sigma_{B,O2}$</td>
<td>gO$^{1/2}$ m$^{-3/2}$</td>
<td>Exponential</td>
<td>0.5</td>
<td>Standard deviation of bias in oxygen</td>
</tr>
<tr>
<td>$\sigma_{B,HPO4}$</td>
<td>gP$^{1/2}$ m$^{-3/2}$</td>
<td>Exponential</td>
<td>0.035</td>
<td>Standard deviation of bias in phosphate</td>
</tr>
</tbody>
</table>

results of the emulator are compared to the posterior marginals derived from a normal approximation by an importance sampling technique shown in [Dietzel and Reichert, 2012]. In many cases both techniques show similar results, for many but not for all parameters the emulation leads to wider posterior marginals.

The model results and their prediction uncertainties for the considered model output variables phytoplankton, zooplankton, nitrate, phosphate and oxygen derived from the emulation of the posterior distribution and propagation of a Markov chain subsample are presented in Figs. 4.5–4.7. The results show large similarities with the results calculated by a normal approximation to the posterior shown in [Dietzel and Reichert, 2012].

Several problems emerged during the setup of the emulator and the subsequent Markov chain Monte Carlo sampling. Some problems, the suitable choice of the exponent $\alpha$ and correlation lengths factors $\lambda_{fact}$ as well as the choice of the design data set, had already been explored by the didactical example. Additionally, the design data set had to be increased for the lake model application due to the higher dimension of the parameter space. For the lake model application, it was not possible to derive the design data set from the normal distribution found as an approximation of the posterior distribution in an earlier study ([Dietzel and Reichert, 2012]), neither to use this approximation as the underlying model for the emulator. These problems resulted from the large correlations and narrow form of the normal distribution derived from the importance sampling. They could partly be solved by a more regional sample of the posterior from an uncorrelated normal distribution around the maximum of the posterior as a design data set. Using a different, even simpler model underlying the emulator, improved the emulator performance as well. Furthermore, a larger subsample of the resulting Markov chain propagated through the lake model enhanced the estimation of prediction uncertainty bands.
Figure 4.4: Prior (dotted, supported by light grey shading of the areas below the lines) and posterior (supported by dark grey shading of the areas below the lines) marginals of the model parameters from the posterior approximation by the GASP emulator (solid) and a normal approximation from a ball/normal importance sampling (dashed).
Figure 4.5: Phytoplankton ((a),(b)) and zooplankton ((c),(d)) concentrations in the epilimnion ((a),(c)) and hypolimnion ((b),(d)) of Lake Zurich. Data points (markers), output of the deterministic model (long-dashed), median (solid) and 95% credibility bounds (dark grey area with dashed boundaries) of bias-corrected output and median (solid; same as for bias-corrected output) and 95% credibility bounds (dark and light grey areas with dotted boundaries) of predictions of new observations (including observation error) for the whole simulation time.
4.5 Conclusions

In the current study, we suggest to replace an approximate parametric description of the posterior by the use of a Gaussian stochastic process (GASP) emulator to be able to do Markov chain Monte Carlo sampling for Bayesian inference with computationally demanding models. A subsample of the Markov chain can then be propagated through the full model to obtain model predictions. The suggested procedure is based on a likelihood function which includes the description of model bias by a stochastic process to ensure a profound uncertainty estimation also for structurally uncertain models. The suggested posterior emulation is an alternative to emulating the full model that is very easy to implement in a universal way as it is independent of the model structure. On the other hand, it has the disadvantage of still requiring the full model for prediction. This disadvantage is reduced by only using a small subsample of the Markov chain for prediction. Both aspects, reducing the computational burden of Bayesian inference...
Figure 4.7: Nitrate ((a),(b)) and phosphate ((c),(d)) concentrations in the epilimnion ((a),(c)) and hypolimnion ((b),(d)) of Lake Zurich. Data points (markers), output of the deterministic model (long-dashed), median (solid) and 95% credibility bounds (dark grey area with dashed boundaries) of bias-corrected output and median (solid; same as for bias-corrected output) and 95% credibility bounds (dark and light grey areas with dotted boundaries) of predictions of new observations (including observation error) for the whole simulation time.
4.5. Conclusions

Figure 4.8: Median and 95 % credibility interval of posterior estimate of bias (solid line and shaded area bounded by dashed lines) and median posterior estimate and uncertainty of observation error (markers and vertical lines) of observed output variables in Lake Zurich. The units refer to the units of respective output variables on a Box-Cox transformed scale.

and prediction and considering systematic model errors, are of particular interest in environmental modeling where often computationally demanding, highly complex and hence structurally uncertain models are used.

We could show that the presented method results in a better posterior approximation than an approximation of the posterior by a normal distribution. At least for the didactical example, the results are more similar to a Markov chain Monte Carlo sampling of the full model. This could be achieved at only slightly higher computational costs than for the normal approximation and linear error propagation. A first maximization of the posterior
distribution is needed for both techniques. For the emulation procedure, Markov chain Monte Carlo sampling of the emulator is needed, which is computationally very cheap because the emulator needs fractions of seconds to evaluate whereas the full model can take minutes to hours, as in case of the lake model BELAMO. For the following propagation of a subsample of the Markov chain around 1000 simulations with the full model are needed, instead of around 100 times more as for a MCMC sampling with the full model. Additionally, these 1000 simulations can be run in parallel, which is not possible for the MCMC sampling.

Besides these positive results of the presented emulation technique, we also discovered some problems with the optimal setup of the GASP emulator. It became obvious that the evaluation of the emulator performance by its root mean square error (RMSE) for a small test sample can be misleading. Due to the small size of the test sample, it is possible to miss oscillations of the emulator leading to poor behaviour of the Markov chain even for emulator parameters for which the emulator seems to approximate the posterior well. In particular, the choice of the design data set, the underlying simplified model, and the emulator parameters $\alpha$ and $\lambda_k$, are crucial factors for the quality of the method. These choices are difficult, as in situations in which emulation is interesting, the computational demand of the full model does not allow extensive testing of the emulator performance. The application to the lake water quality model BELAMO showed that these difficulties increase with increasing complexity of the simulation model. This indicates that the suggested emulation technique may not be applicable to models of arbitrary complexity.

The suggested emulation technique may replace parametric approximations of the posterior of computationally demanding models, which is still the dominant approximation technique in textbooks about Bayesian inference. The main advantage over the parametric approximation is the better performance at only slightly increased computational effort. On the other hand, the difficulties in finding the optimal emulator parameters also show the limitations of this approach and indicate that more research in this and alternative techniques is still required.

4.6 Acknowledgments

This research was funded by the Swiss National Science Foundation (SNSF). We are grateful to Richard Forster and Oliver Köster from WVZ, Barbara Känel, Pius Niederhauser, Irene Purtschert and Peter Spohn from AWEL and Daniel Streit and Christian
Leu from BAFU for the kind provision of data on chemistry, biology and physics of Lake Zurich, discharging waste water treatment plants and inflowing rivers.
Chapter 5

Conclusions and Outlook

This work explores the possibilities of modeling the biogeochemistry and the main aspects of the plankton dynamics of three Swiss lakes by the lake model BELAMO. This is important, as modeling environmental systems can help to increase the understanding about these systems, test hypotheses about the mechanisms driving their dynamics and predict their future development under changes in driving forces. This is a substantial element of environmental decision support and can contribute to the preservation of biodiversity and ecosystem services like drinking water supply, as potentially the case for lakes.

However, systematic deviations of model results from observations remain and hamper this important contribution of environmental modeling to ecological research and management. This bias is primarily caused by the simplified description of the reality by the model and lack of precise knowledge of driving forces. In the case of BELAMO, systematic deviations mainly occur due to the high spatial and structural aggregation, the difficult task of jointly calibrating the model for three lakes with only few lake-specific parameters, the attempt to close the mass balances by mechanistically describing the mineralization processes in the sediment, and uncertainty about the input to the lakes. The high spatial aggregation that results from the description of the lake water column by only two mixed boxes leads to a coarse representation of vertical mixing processes and makes it more difficult to find the right aggregation of measurements to be compared with model results. The large structural aggregation results from modeling the total sum of phytoplankton and sum of zooplankton instead of a more detailed separation of the plankton community. The structural aggregation does not account for any variability in the composition, temporal succession or adaptation. The task of jointly calibrating the model to three lakes of different trophic state over a long-term period increases the difficulties in finding the most suitable universal model structure and parameterization.
Chapter 5. Conclusions and Outlook

The consideration of mineralization in the sediment increases the problem of finding adequate mineralization rates over the whole simulation period. And finally, inputs are only sparsely measured and the measurements do not in all cases directly correspond to model state variables (in particular, biodegradability of organic particles is relevant for the model and is usually not known).

This thesis followed three approaches to improve the lake model BELAMO and to estimate its prediction uncertainty.

First, an attempt was made to reduce the bias by improving the model. Several modifications to the model were needed to increase its universality. In particular, active movement of zooplankton to the epilimnion and sorption of phosphate to sedimenting not degradable organic particles was introduced into the model, the formulations of zooplankton growth, of grazing of zooplankton by diptera and of anaerobic mineralization were improved and the biodegradability of external organic particles was reduced. These modifications considerably reduced the bias and led to model results that represented the key features of the data reasonably well.

Despite this adequate description of the data, significant discrepancies between model results and data remain, which are larger and more systematic than expected for observational errors. To get adequate uncertainty estimates of model parameters and predictions, these remaining systematic deviations have to be considered. This can be done by including a separate bias term into the error model, which distinguishes autocorrelated structural error from the approximately independent and normally distributed observation error. This term accounts for larger uncertainties due to bias in model results and for larger uncertainties during the prediction period compared to the calibration period when knowledge about the bias declined. The use of this technique is essential for structurally uncertain environmental models like the studied lake model, but when included in a traditional approach for Bayesian inference based on a Markov chain Monte Carlo technique, it requires many model runs to be performed. As the consideration of bias for uncertainty estimation is needed, but for computationally demanding models, the computation times can be prohibitive, we developed computationally less demanding, approximate techniques for parameter estimation and prediction.

As a second step of this study, we used a normal approximation to the posterior close to its maximum and linear error propagation to model results to reduce the computational burden. Although this method is a simplification, it can be used to gain information about the parameter, observation and model structure uncertainty and the resulting overall model prediction uncertainty. We introduced different approaches to approximate the posterior by a normal distribution, suitable to be applied to models of different
complexities and shapes of posterior distributions. These approximations do not rely on the specific formulation of the likelihood function used herein, but could be used in other contexts where an approximation of the posterior distribution by a normal distribution seems to be meaningful and where a full Markov chain sampling is computationally too expensive.

The replacement of the posterior by a normal distribution appeared to be a very coarse approximation. To find an intermediate method between a normal approximation of the posterior and a full Markov chain Monte Carlo sampling of the posterior, a Gaussian stochastic process emulator was used as a third contribution of this thesis. This emulator interpolates the posterior probability density to accelerate statistical inference of model parameters. A small sub-sample of the Markov chain sampled from the model and error parameters, for which the emulator is evaluated instead of the model, is propagated through the full model to get uncertainty estimates of model predictions. This method causes a more realistic posterior approximation than a normal distribution while only slightly increasing the computational burden.

The modifications made to the structure and parameterization of the lake model BE-LAMO, made a representation of the key pattern in the observations possible. The findings of this first step improved our understanding of the relevant processes and mass fluxes in the lake. The methods developed in the second and third step were required for conducting the study, but they are relevant in a much more general context of environmental modeling. In general, environmental models typically suffer from large computation times, uncertain model structure and large simplifications to depict complex real systems, as it was the case for the long-term simulations with BELAMO. These complexities and uncertainties hamper model calibration. The resulting uncertainties of model predictions need to be addressed, especially if the model is used in the context of decision support, environmental management or communication to society. The current work makes a significant contribution to this area of research by introducing calibration and uncertainty estimation techniques that are suitable for computationally demanding models that suffer from large structural deficits.

Despite the progress made in this thesis, research needs remain in both fields, mechanistic ecosystem modeling of lakes and methodologies to improve parameter inference and prediction.

As discussed above, major deficiencies of the model and its application are the coarse spatial, structural and temporal resolution of the model, the coarse description of mineralization processes in the lake sediment, and input uncertainty. Further improvements could thus be:
• use of a higher spatial resolution of the lake (1d or even 3d) and coupling to a hydro-dynamic model to quantify the mixing processes and their dependence on atmospheric driving forces;

• use of a higher time resolution of the simulations to account for daily variations, in particular in primary production and surface layer mixing;

• introduction of functional groups of phytoplankton and zooplankton or even taxonomic groups, potentially supported by applying the metabolic theory of ecology (Brown et al., 2004) to limit the increase in the number of model parameters;

• coupling the model of the processes in the water column to a more sophisticated sediment model that has a higher spatial resolution and a more complete representation of mineralization processes;

• improved probabilistic description of inputs to the lake and other driving forces.

Investigations of the adequacy of such model extensions could be supported by the application to more lakes, also in different climate zones. Insight into the structural uncertainty could be improved by comparing results of different lake models or model variants to the same lakes.

The bias description technique applied in this study may not need much further development and could become a standard approach for dealing with remaining bias in environmental modeling. However, application of such a technique should always be complemented by improvements of the model structure and the numerical efficiency of the inference and prediction techniques. Besides the specific suggestions for the lake model given above, relevant methodological contributions could be:

• introduction of stochasticity into the deterministic model could account for remaining aggregation errors and driving forces that could not be considered and could support the identification of structural model errors;

• instead of focusing on the mean of the result of the presented emulator, the consideration of the standard deviation could give information about the uncertainty induced by the emulator;

• further development of dynamic emulators could make MCMC techniques applicable to an approximation of the full model instead of only emulating the posterior probability density;
• use of adaptive MCMC techniques (Gilks et al., 1998; Liu et al., 2000; Haario et al., 2001; Haario et al., 2006; Andrieu and Thoms, 2008) could further improve the efficiency of the numerics of the inference procedure;

• use of stochastic models will require further development of numerical algorithms for statistical inference, such as Approximate Bayesian Computation (ABC) techniques (Marjoram et al., 2003; Beaumont, 2009).

In general, research into all these directions is needed for being able to test different ecological hypotheses and future management scenarios with environmental models.
Bibliography


List of Figures

2.1 Calibration didactical example: prior and posterior marginals ............. 33
2.2 Calibration didactical example: uncertainty results .......................... 34
2.3 Calibration Lake Zurich: prior and posterior marginals ...................... 44
2.4 Calibration Lake Zurich: uncertainty plankton ............................... 45
2.5 Calibration Lake Zurich: uncertainty oxygen ............................... 46
2.6 Calibration Lake Zurich: uncertainty nutrients ............................... 47
2.7 Calibration Lake Zurich: bias and observation error ......................... 48

A.1 Calibration didactical example, wrong priors: prior and posterior marginals 55
A.2 Calibration didactical example, wrong priors: uncertainty results ............ 56

3.1 Map study area . .............................................................................. 64
3.2 BELAMO: structure and physical processes ................................. 65
3.3 BELAMO: biological and chemical processes ............................. 66
3.4 Nutrient input loads ................................................................. 76
3.5 Simulation & uncertainty results: Greifensee ................................. 87
3.6 Simulation & uncertainty results: Lake Zurich ............................. 88
3.7 Simulation & uncertainty results: Walensee ................................ 90
3.8 Mass fluxes nitrogen ................................................................. 92
3.9 Mass fluxes phosphorus ............................................................. 93

B.1 Functional response zooplankton grazing ....................................... 99
B.2 Simulation results: Greifensee ..................................................... 108
B.3 Simulation results: Lake Zurich .................................................. 109
B.4 Simulation results: Walensee ..................................................... 110

4.1 Performance emulator ............................................................... 125
4.2 Emulation didactical example: prior and posterior marginals ........... 127
4.3 Emulation didactical example: uncertainty results ............................ 129
4.4 Emulation Lake Zurich: prior and posterior marginals ...................... 135
4.5 Emulation Lake Zurich: uncertainty plankton ................................. 136
4.6 Emulation Lake Zurich: uncertainty oxygen .................................... 137
4.7 Emulation Lake Zurich: uncertainty nutrients .................................. 138
4.8 Emulation Lake Zurich: bias and observation error .......................... 139
List of Tables

2.1 Didactical example: Comparison number of simulation runs . . . . . . . . 32
2.2 Calibration BELAMO, Lake Zurich: parameter prior definitions . . . . . . 40
2.3 Calibration BELAMO, Lake Zurich: prior definitions observation error . . . 40
2.4 Calibration BELAMO, Lake Zurich: prior definitions bias . . . . . . . . . . 41
2.5 BELAMO: Comparison number of simulation runs . . . . . . . . . . . . . 42
3.1 BELAMO: Model parameters . . . . . . . . . . . . . . . . . . . . . . . . . 75
3.2 Calibration BELAMO, all lakes: parameter prior definitions . . . . . . . . 81
3.3 Joint sensitivity analysis all lakes . . . . . . . . . . . . . . . . . . . . . . . 83
3.4 Calibration BELAMO: estimated parameter values . . . . . . . . . . . . . 85
B.1 Lake-specific sensitivity analysis . . . . . . . . . . . . . . . . . . . . . . . 105
4.1 Emulation BELAMO, Lake Zurich: parameter prior definitions . . . . . . . 133
4.2 Emulation BELAMO, Lake Zurich: prior definitions observation error . . . 133
4.3 Emulation BELAMO, Lake Zurich: prior definitions bias . . . . . . . . . . . 134
Acknowledgments

First of all I want to thank my supervisor Peter Reichert for his everlasting support, understanding, practical help, patient explanations and motivation. I am thankful to have learned a lot from his broad knowledge and for him always creating a comfortable working atmosphere.

I would also like to thank my co-examiners Jukka Jokela and Thomas Petzoldt for their willingness to be part of my doctoral committee and for improving my thesis.

I am thankful for the nice collaboration with the Polyfaktor Designbüro (http://www.polyfaktor.com/) and their professional layouting of some of the figures in chapter 3, the cover and the separating pages of my thesis.

I was lucky to work in a nice atmosphere throughout my PhD. For that, I especially would like to thank my office mates Fauzia Ahmed, Carlo Albert, Judith Ellens, Renata Hari, Salvador Peña Haro and Zhao Xu, who were all nice roomies and expanded my horizon from Adelboden to the Wallis, from Ghana to México and Beijing, from Lindor chocolate to climate-friendly food and from "Science" to "Sendung mit der Maus". I am happy about the plenty funny and interesting lunch, coffee or cake breaks, for which I thank everyone from the Siam department, in particular Jafet Andersson, Saeid Ashraf Vaghefi, Hans-Peter Bader, Thomas Buchsbaum, Emmanuelle Cohen-Shacham, Nadja Contzen, Monireh Famarzi, Karin Ghilardi, Alexandra Huber, Jenny Inauen, Lukas Indermaur, Nasreen Khan, Christine Kündig, Judit Lienert, Ivana Logar, David Machac, Hansi Mosler, Silvie Palacios, Elham Rouholahnejad, Ruth Scheidegger, Jacqueline Schlosser, Florian Schnetzer, Lisa Scholten, Elisabeth Seimetz, Rosi Siber, Colombe Siegenthaler-Le Drian, Andrea Tamas, Robert Tobias, Sören Vogel and Eike von Lindern. I thank Johanna Mieleitner for the great help in the first months, Simon-Lukas Rinderknecht for being my PhD and exercise buddy, Raoul Schaffner especially for all the Mac and server help and Simone Langhans and Nele Schuwirth for "golden confetti" memories, professional aid and a lot of fun.
As I did my diploma thesis in the Uchem department, I also felt connected to that department and associated people. I especially would like to thank Aurea Chiaia, Tobias Doppler, Anna Gulkowka, Irene Hanke, Tobias Heck, Martin Frey, Susanne Kern, Luba Kovalova, Merle Richter, Andreas Skretschmann, Holger Tülp and Jürgen van der Voet for funny karaoke evenings and other nice occasions. I thank Sebastian "Baschi" Huntscha for enduring music supply, Stephan Koepke for a bit of Rheinland feeling and Aduccia Sciacovelli for being my "little dark twin".

I am glad that I got to know Irene Wittmer and Saskia Zimmermann, great girls, great company for all kinds of activities.

A thank goes to the Eawag restaurant aQa for constant cappuccino, muffin and mood food supply, to the Eawag reception team for a lot of help and stamps and to Ilse Hildbrand for four years of relaxing Monday yoga.

I am grateful to my Wehntaler flat mates Nicole Schüpfer, Liliana Belchior, Yvonne Kunz and especially Jacqueline Reiser during that time and thank my Español fellows Katharina Leucht and Lea Fünfschilling for inspiring lunch breaks.

Furthermore, I was lucky to also have some friends back in the Rhineland and in Osnabrück. Although they were too far away most of the time, it was a great encouragement to know they are somewhere, anyway. I especially would like to thank Christiane Zarfl, Christof Söger, Jochen Sprickerhof and the rest of the (Ex-)Mathe-Info-Sywi-Fachschaft of the University of Osnabrück and my school girls Sandra Hamacher, Nadia Kadereit and Eva Reiff. Special thanks go to Rekha and Alexander Witzke, Diana "procrastination" Tessmer, Petra "wir telefonieren dann nochmal" Lange and Simone "eschte Fründe" Leyendecker.

Unfortunately, my father Rolf Dietzel missed my defense for 20 years. All the more, I deeply thank my mother Theodora Dietzel for her loving support in all parts of my life, my aunt Cäcilia for her encouragement, my sister Ruth, Torsten and my brother Mathias for always being there for me in academic as well as in personal affairs, Timo and Sabina for showing what else matters. Salvador, for being the best that happened to me...
Curriculum vitae

Anne Dietzel
Born on September 24, 1983 in Langenfeld (Rheinland), Germany

Academic Education

12/2007 – 06/2012  PhD student at the Department of Systems Analysis, Integrated Assessment and Modelling, Eawag Dübendorf and the Institute of Biogeochemistry and Pollutant Dynamics, ETH Zurich
06/2002
08/1993 – 06/2002  Abitur (general qualification for university entrance)
08/1990 – 06/1993  High school: Erzbischöfliche Marienschule, Leverkusen-Opladen, Germany
08/1990 – 06/1993  Primary school: Martinusschule, Langenfeld-Richrath, Germany
Professional Experience

2009 – 2011  Teaching assistance in Environmental Systems Analysis, Summer School, Eawag, Dübendorf
2008 – 2011  Teaching assistance in Modelling of Aquatic Ecosystems, Spring semester, D-UWIS, ETH Zurich
2006  Teaching assistance in Environmental Systems Analysis, Winter semester, University of Osnabrück, Germany
07/2006 – 10/2006  Research assistance at the Institute of Environmental Systems Research (USF), University of Osnabrück, Germany
08/2005 – 10/2005  Student research project at the Department of Environmental Chemistry, Eawag and the Institute of Environmental Systems Research, University of Osnabrück
2005  Tutor in Theoretical Informatics, Summer semester, University of Osnabrück, Germany
2004 – 2006  Tutor in Systems Science I, Winter semester, University of Osnabrück, Germany
"Der Rhing der is zick neustem
d'r Kochpott der EG.
Die Schweiz die deit jet Öl dabei,
von Frankreich wed met Salz jewürz.
Un bei uns do wed dat Janze chemisch konservet.
In Holland wed dat Süppche dann
als Trinkwasser probet.

Einmol em Johr kült d'r Rhing us em Bett,
nämlich dann wenn hä Huhwasser hätt.
Un dann freuen sich de Fesch,
dat es doch klor,
denn dann han se widder Sauerstoff
et eetste mol em Johr.
Un dann freut sich och d'r Minsch,
dat es doch klor,
denn dann stink et nur noch hallef esu vill
et eetste mol em Johr."

("Einmol em Johr", De Bläck Fööss, 1975)