Lake modelling

an interdisciplinary context

Koji Tominaga

Dissertation for the degree of *Philosophiae Doctor*

Department of Biosciences
Faculty of Mathematics and Natural Sciences
University of Oslo, Norway
2013

© Koji Tominaga, 2013

Series of dissertations submitted to the
Faculty of Mathematics and Natural Sciences, University of Oslo
No. 1368

ISSN 1501-7710

All rights reserved. No part of this publication may be
reproduced or transmitted, in any form or by any means, without permission.

Cover: Inger Sandved Anfinsen.
Printed in Norway: AIT Oslo AS.

Produced in co-operation with Akademika Publishing.
The thesis is produced by Akademika publishing merely in connection with the
thesis defence. Kindly direct all inquiries regarding the thesis to the copyright
holder or the unit which grants the doctorate.

Acknowledgements

I would like to thank my main supervisor Prof. Tom Andersen for making everything about my PhD work possible. Tom has always been caring, polite, understanding, and patient. For such a stubborn person like me, the trust that Tom is always willing to listen to my ideas and he is happy to discuss research opportunities and other matters with me was essential. Thank you very much for supporting me with your vast knowledge and insights in an unpretentious way. If I ever become a supervisor, I would try my best to show respect to my students as much as Tom has been showing us. I truly enjoyed my research at the University of Oslo, and it is my honour to be one of his students. Tom, I hope you did not feel that I take your kindness for granted or that I took advantage of it, and I sincerely hope you enjoyed supervising me. Thanks, again.

My co-supervisor and EUTROPIA project leader Prof. Rolf D. Vogt is a very enthusiastic researcher who helped me even with many non-academic logistics. Rolf also kept relevance of my research to the concerned matters in real world. The other co-supervisor Dr. Tuomo Saloranta was also of great help, especially for his knowledge on intricate modelling of lake physical dynamics. Come to think of it, I was fortunate because all three supervisors were not just being always available, but they are all attending to and interested in my work progress and results. I also got all the scientific and knowledge provision I needed.

Another aspect I was fortunate about was that I could work together with many helpful and responsible people. The studies included in this thesis were only possible with their help and expertise. Here, I decided to make a single long paragraph, and listed them alphabetically. Dr. Aleksandra Trnic Romarheim collected many of the data used in Paper I, and wrote most part of it. We worked together for months on the presented modelling work. Although the work was both academically challenging and interesting, the collaborating effort was pleasant and memorable. Alexander Melvold Engebretsen managed a huge geographical dataset and the catchment model SWAT that I depended on, in addition to assisting me with lake sampling and laboratory analyses, while he worked on his own fieldwork, experiments, laboratory analyses and teaching. Being the other PhD student in the EUTROPIA project, Alex's contribution to the this thesis was tremendous. Dr. Anders Gravbrøt Finstad gave me through his network an opportunity to participate in his project, which effectively worked as a parallel study to EUTROPIA, presented in Paper II. Although the task was challenging, I got much freedom to design the study, and the process and the results were exciting. Besides him being friendly and agreeable, I relied also on his skill to predictably complete a vast amount of work in a short amount of time. Dr. Jostein Starrfelt worked on the model calibration method MCMC-DREAM and the model application of INCA-P, which I depended on. Jostein was the main co-worker I spent time with the most at the Norwegian Institute for Water Research (NIVA), and his professionalism and reliability were important for me.

Furthermore, he is very insightful about modelling and science, and many concepts that I present in this thesis is a fruit of frank discussions with him. Dr. Øyvind Kaste has been an attentive and considerate boss of mine at NIVA, and arranged a generous offer for me to work part-time at NIVA on modelling and various projects. These opportunities themselves widened my academic perspectives and international research arena, on top of providing me with extra time for working with the thesis. Øyvind helped me with many of the manuscripts included in this thesis.

EUTROPIA is an interdisciplinary project. And, because of the delivery time and scope of the thesis, research efforts provided by these researchers are not presented here, but they make my study meaningful and relevant to the abatement strategy study at the Morsa-Vansjø waters: Anne Bjørkenes Christiansen, Dr. Ann Lyche Solheim, Dr. Brian Lutz, Dr. Csilla Farkas, Dr. David N. Barton, Dr. Richard R. Wright, Dr. Eirik Romstad, Dr. Geir Orderud, Helga Gunnarsdottir, Dr. Isabel Seifert, Dr. Jannicke Moe, Dr. Marianne Beckmann, Dr. Oddvar Røyseth, Dr. Raoul-Marie Couture, Silje Holen. I am looking forward to further collaborations with some of these researchers towards the completion of the project and more. Furthermore, there are many more people who have been collecting and analysing samples in the lake, in the catchment and at the weather stations. Some of these people made logistics for my own field samples and measurements. These include Berit Kaasa, Christian Mohr, Prof. Gunnhild Riise, Marte Rosnes, Dr. Per Johan Færøvik, Dr. Sigrid Haande, Prof. Thomas Rohrlack. They also contribute to knowledge around the catchments and the lake through their own researches. In addition, I would like to send my appreciation to people who were involved in earlier projects in the Morsa catchment and those who work on monitoring by NIVA, the Norwegian Water Resources and Energy Directorate (NVE), the Norwegian Meteorological Institute, the Norwegian Institute for Agricultural and Environmental Research (Bioforsk), and the Norwegian University of Life Sciences (UMB). I regret that I may not even recognise the names of the individuals but their contribution is enormous and this needs to be noted. This is particularly important for me as an environmental modeller who relies on long continuous data. I also want to thank two students that I 'partially coached' during my position at the University of Oslo: Lluís Gómez Gener and Vincent Hess. Lluís and Vincent, you were both independent and very capable in your respective research fields, and showed great potentials to be prominent researchers someday. It was my pleasure to having been able to work with you – thank you. Many MSc students and visiting researchers at the Department of Chemistry at the University of Oslo are also involved in EUTROPIA and their research efforts are also acknowledged. In particular I want to thank Frøydis Meen Wærsted for picking up some of the studies at Centre for Accelerator-based Research and Energy Physics (SAFE) that I unfortunately skipped due to time constraints. I also would like to thank people at SAFE especially Frøydis Schulz and Inger-Marie Hansen, although I did not complete my studies there.

I also thank those at the university or at NIVA who made my stay enjoyable. You may not have had anything to do directly with my research contents for the thesis, but your friendliness was important during my stay. Andrea L, Andrea G, Anna, Bente, Birger, Eirin, Elianne, Gro, Heleen, Heidi, Jan Erik, Jonas P, Jonas T, Josefin, Kathrin, Kari, Ketil, Line, Maja, Marcia, Marcin, Markus, Maryia, Mazy, Oda, Rita, Sagnik, Sandra, Serena, Simon, Sissel, Therese, Thijs, Thorjørn, Toffe, Wenche and Yan: thank you! Special thanks go to Kathrin who were not only a wonderful office mate but also a person who would arrange with others social gatherings that were very important come to think of it. Thank you for the others who also made them happen, especially those PhD students and postdocs at the Department of Biology (now the Department of Biosciences). Special thanks also go to Bente who was a very pleasant and helpful neighbour and made my stay in the last year very comfortable. My friends in Oslo: Hiro, Maki, Erik, Michiko, Patrick, Kubota-san, Hino-san, Gabriel, and Lena, thank you for the wonderful times in Oslo. I would like to thank especially Hiro-san for all the generous helps and wonderful times together.

I thank Sumera and Helen for being a reliable gateway to the administration works.

I also would like to sincerely appreciate the committee members Dr. Rita Adrian, Dr. Martin Schmid, and Prof. Asbjørn Vøllestad for taking their precious time to read a draft of the thesis.

I also would like to take this opportunity to thank Dick and Thorjørn. Because you knew both my previous supervisors and my current supervisors, your advise four years ago was valuable for me when deciding to come to Oslo. I hope my performance over the last four years came close to what you had imagined originally then.

My appointment at the University of Oslo was provided by EUTROPIA, which is a project through the Research Council of Norway. My part-time appointment at NIVA in addition made my research period longer, which was essential.

Finally, and most important, I thank my family, especially Nelly and my mother for supporting me.

In Osaka and Singapore, and in transit

Koji Tominaga

Table of contents

There are 6 figures and 4 tables in this document.

List of papers

This thesis is based on the following manuscripts, and they will be referred to hereinafter using the Roman numerals.

Paper I Romarheim, A. T., Tominaga K., Riise G., and Andersen T. Modelling the responses of a cold temperate lake to changes in external forcing. *Manuscript*.

Paper II Tominaga, K., Finstad, A. G., Kaste, Ø., and Andersen T. Melting and warming of northern lakes. *Manuscript*.

Paper III Tominaga, K., Starrfelt, J., Kaste, Ø., and Andersen T. Simultaneous calibration of two sequentially connected models. *Manuscript*.

Paper IV Tominaga, K., Engebretsen, A. M., Starrfelt, J., Kaste, Ø., Vogt, R. D., and Andersen T. Calibration uncertainty in sequential model application. *Manuscript*.

I contributed also to the following papers during the doctoral programme. These papers are not included in the thesis.

- Mooij, W. M., *et al.* Challenges and opportunities for integrating lake ecosystem modelling approaches. 2010. *Aquatic Ecology 44*: 633–667.
- Tominaga, K., *et al.* Predicting acidification recovery at the Hubbard Brook Experimental Forest, New Hampshire: evaluation of four models. 2010. *Environmental Science & Technology 44*: 9003–9009.
- Trolle, D. *et al.* A community-based framework for aquatic ecosystem models. 2012. *Hydrobiologia 683*: 25–34.

Summary

As a deliverable of the applied environmental science project Eutropia, this thesis elaborates on an interdisciplinary theme between two academic disciplines: limnology and modelling. The main goal of my tasks in the project was to simulate the impact of nutrient loading on the lake Vansjø using a lake model MyLake. In addition to documenting the model configuration, this thesis also presents novel parallel studies that I conducted with my colleagues. Findings and conclusions from these studies include predictions of dramatic changes in lake ice phenology and temperature during this century, a classification for model parameter uncertainty typology, simultaneous calibration of multiple models, and importance of year-to-year stochasticity in model inputs on lake simulation. These parallel studies assisted in improving the relevance of the main modelling task at Vansjø.

1 The thesis

As *Stock and Burton* (2011) summarise, integration of what used to be separate traditional academic disciplines is now regarded as a fundamental instrument for addressing environmental issues. This is also reflected in an increasing fraction of integrated disciplines in the literature over the past decades (Figure 1). The Norwegian research community is not an exception. The Research Council of Norway (RCN/NFR) for example runs the Miljø2015 programme (English: Norwegian Environmental Research Towards 2015), in which a focal point is integration of various existing knowledge (physics, chemistry, biology, social sciences and cultural studies). The University of Oslo has recently established seven new interfaculty research areas with the recognition that knowledge demands in society are complex and require interdisciplinary collaboration in some fields of application, notably in environmental and sustainability research. The building at the new research park in Oslo (Norwegian: Forskningsparken) is even named as such: CIENS (pronounced 'science,' Oslo Centre for Interdisciplinary Environmental and Social Research).

Many terms have been coined to describe integration of disciplines, e.g., multidisciplinary, interdisciplinary, transdisciplinary and crossdisciplinary. Whereas they seem to vary in the extent of disciplinary integration and the ultimate goal, precise definition is not always straightforward. Typological issues aside, the pertinent observation is that the interaction is now called for on multitude of levels, from interaction among scientists to interaction between a research team and policy makers.

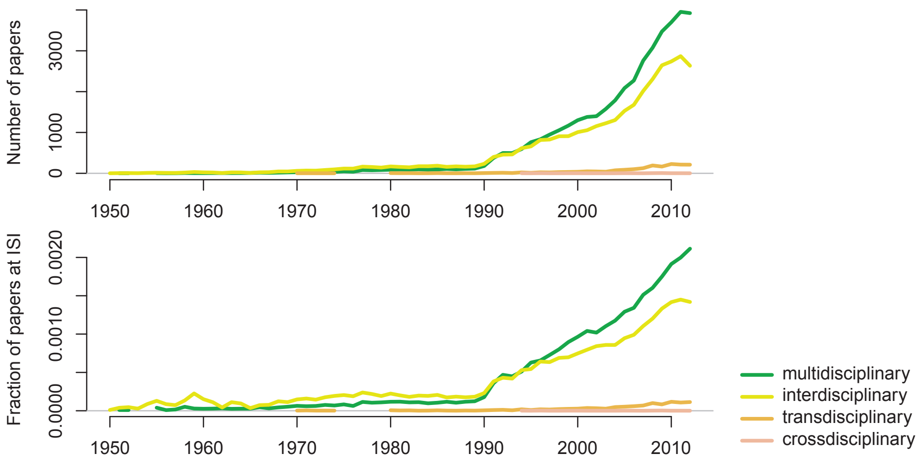


Figure 1: Frequencies of four discipline integration keywords in the literature index ISI Web of Science as of 8 February 2013. Entries to the recent years are underestimated as they may not have been registered yet in the database. Fraction of occurrence shows weighted representability of these words in the literature, as the total number of entries has increased over the decades.

This thesis is about integration between limnology and modelling, thus exemplifying the interaction between the two natural science subjects, in an effort to provide remedial knowledge for eutrophication at Vansjø (59.42° N, 10.86° E) in southeastern Norway (Figure 2). Such integration is considered important, as new knowledge is expected to emerge by mixing the two subjects. It is also challenging due to subject specialisation (own jargon, advanced concepts and unfamiliarity), as discussed in the next sections. Nonetheless, successful integration is a key element in the funding project Eutropia. The RCN-funded Eutropia project (full project name: Watershed Eutrophication Management through System Oriented Process Modelling of Pressures, Impacts and Abatement Actions) has several goals revolving around eutrophication issues at Vansjø. The research approach taken in this project is process-based modelling, in particular modelling of effectiveness of abatement actions from both socioeconomic and natural science perspectives. Eutropia is a Miljø2015 project designated under the Crosscut theme (Norwegian: Tvers), with the emphasis of knowledge integration from all relevant sciences. In light of that, my contribution, namely the *limnology-modelling* connection, is a pillar of the project, for the project is aiming to orchestrate interactions beyond natural sciences and even beyond the research communities to the public.

In the next sections, I will introduce some key concepts in each of the two subject fields that I have tried to integrate. These concepts may perhaps be so fundamental and so familiar to subject experts that my following illustration may appear superfluous. On the opposite extreme, my illustrations may appear over-simplistic on contentious research matters of the subject today. This impression may of course reflect my shallow expertise in these subjects, but in some cases I am doing this deliberately in order to effectively communicate the relevant knowledge to non-experts. The need for easing communication is indeed the challenge that all researchers involved in interdisciplinary studies face. Possibly, the type of knowledge researchers in a particular discipline unanimously perceive interesting, novel and valuable may not be pertinent when the discipline participates in an interdisciplinary study. The summary given below is only my interpretation as to what is the most applicable and useful in modelling eutrophication in light of interdisciplinarity.

2 Limnology and eutrophication

Limnology is defined in the Oxford English Dictionary as *'the study of the biological, chemical and physical features of lakes and other bodies of fresh water.'* Perhaps a more precise definition that relates more strongly to the research community and working tendency of the subject is given for example by *Wetzel* (2001): *Limnology is the study of the structural and functional relationships and productivity of organisms of inland aquatic ecosystems as they are regulated by the dynamics of their physical, chemical, and biotic environments*. These

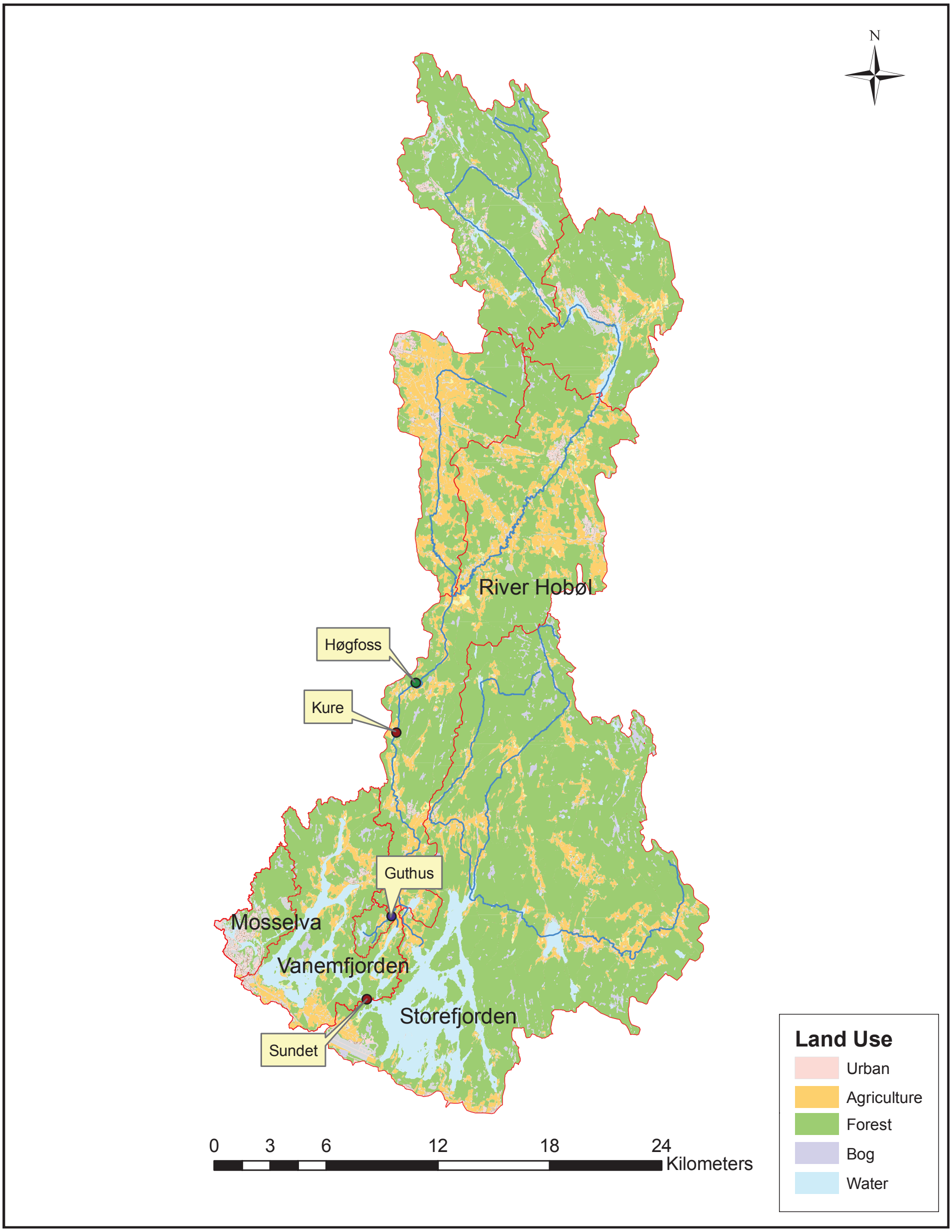


Figure 2: Map of the Vansjø catchment. The lake Vansjø consists of Storefjorden, Vanemfjorden, and Mosselva in this map. Map production by A. M. Engebretsen.

definitions reveal themselves that the subject of limnology is in its own right interdisciplinary (*Duarte and Piro*, 2001). Open a limnology textbook, and the reader will find that there is a discussion of earth sciences, physics, chemistry, and of course biology throughout the book.

Here, I would like to describe important advancements in limnological knowledge that contributed to better understanding of the eutrophication phenomenon.

Eutrophication is a relative term for elevated nutrient status in water systems. Anthropogenic eutrophication of freshwater ecosystem is often considered a form of pollution due to reduced transparency, unpleasant appearance and odour, occurrence of potentially toxic organisms, and their consequences for commercial and recreational use of the lake. Limnological understandings helped identify the main mechanisms through which anthropogenic eutrophication happens.

Very early times, lake researchers thought of intra-connections of players within the lake community to be important (see, for example, *Forbes*, 1887). This microcosm view of the lake ecosystem later became the basis of early development of ecology. Eventually the limnological focus shifted from the lake as an isolated entity to its role as a landscape element, demarcated by ecologically relevant geographical units defined by catchment, or drainage area. It became evident by then that the externally sourced materials such as nutrients affect the lake's ecological conditions and that the flow of water is the major medium for transporting nutrients. This was particularly relevant for eutrophication problems; in most anthropogenically eutrophic lakes much of the phosphorus found in lake water originated in the catchment. Most of phosphorus compounds that are biologically useful such as orthophosphate can bind to the fine mineral particles, such that much of the phosphorus transport occurs when these mineral particles are transported. Erosion and particle transport are linked to rainfall and runoff through complex and highly non-linear relationships. When mineral particles enter the lake, their adsorbed phosphorus may be scavenged by algae or sink with bound phosphorus. The phosphorus attached to the sinking particles may be permanently buried or reintroduced to the biologically active layers of lake water when these particles are resuspended. Phosphorus may also be released from sediment particles to water by dissolution under chemically reduced conditions.

Because phosphorus is one of the major essential elements for primary production, *i.e.*, the basal level biological production, change in phosphorus supply is known to possibly create drastic changes in the ecology of fresh waters. The paradigm of 'phosphorus as the limiting factor' was strikingly demonstrated by a study at the Experimental Lake Area in northern Ontario, Canada (*Schindler*, 1977). In this study a lake was separated into two parts using an impermeable curtain and phosphorus fertiliser was added only to one part. Later, algal bloom and eutrophication were observed only in the part of the lake with phosphorus treatment, clearly supporting the paradigm. Generally, total phosphorus concentration is positively correlated to chlorophyll concentration, which is a generally accepted proxy of algal content in the water.

Increase of algae content has cascading impacts in the lake. First, algae are not transparent and therefore attenuate solar irradiance. This results both in selective heating of near-surface water and in dimming of the deeper layers. The reduced light penetration will give an advantage to planktonic algae suspended, which can trigger dramatic extinctions of bottom vegetation. The excessive amount of algae will eventually be preyed upon by a member of the upper trophic level, or die by themselves, sink to the bottom, and decompose. Oxygen consumption by this decomposition process may eventually cause sufficient oxygen depletion to make the lake bottom uninhabitable by important components of the lake's fauna. Algal bloom is often considered a nuisance to humans, and in case of dominance by toxic cyanobacteria (blue green algae), animals including humans could be harmed.

Other limnological knowledge assists in understanding eutrophication by providing physical bases of lake systems. A key physical perspective is the thermal energy budget of the lake. In lakes most thermal energy originates in shortwave radiation (visible and near infrared) from the sun and will be eventually lost back to the atmosphere as long-wave radiation (thermal infrared). Thermal energy has a direct impact on many biological processes, but it also has indirect impact of the whole-lake process by affecting its transport and storage medium, water. From the physical chemistry point of view, water is a peculiar liquid; water has an unusually high heat capacity and is able to store a large amount of heat. This is a crucial consideration in understanding many lake processes. Perhaps more relevant, water has a unique density characteristic; in contrast to most other substances, liquid water does not attain maximal density at the melting point. From the molecular level point of view, balance between two opposing processes explains this density characteristic of water. First, as temperature of water increases, hydrogen bonds among water molecules that set them apart break, and this makes the molecules come closer to each other and more compacted. Second, increase in temperature causes molecules to move faster, thereby taking up larger space and lowering the density of water. The end result of these two mechanisms working against each other is the characteristic maximal density at a temperature (3.98 °C) that lies between the melting (0 °C) and boiling (100 °C) points (under atmospheric pressure). In the ordinary solid phase of water (*i.e.*, ice), hydrogen bonds are arranged hexagonally in a sparse and structured manner, which makes ice lighter than water.

This density characteristic of water has a significant consequence in seasonal temperature distribution in lakes. Let us take a lake that freezes in the winter as an example. Ice being lighter than liquid water means that lake freezes at the surface and that the formed ice insulates the rest of waterbody from the cold atmosphere. And, because the water at 0 °C is lighter than the water at 3.98 °C, bottom of the lake remains warmer than the top of water column interfacing ice (0 °C). This density gradient and sheltering by ice cover prevents mixing of layers until the ice is melted and the top part of water warms up to 3.98 °C. In the summer, solar heating declines exponentially downward from the lake surface, which makes the surface layers lighter than the bottom layers, again separating the top water from the bottom water. The thickness of the isothermal top layer is determined by the balance of wind-driven mixing and the reinforcement of the density gradient by solar heating near the surface. Layers in a stratified water column do not mix very well, but there is heat exchange due to turbulent diffusion, slowly warming the bottom part of the lake as summer progresses. In the autumn, heat loss on the surface due to the thermal radiation exceeds the input of solar irradiance on a diurnal basis. The corresponding density changes cause convective mixing, which lowers the mixing depth. As the mixing depth increases, the whole water column will eventually be mixed and become isothermal. This well-mixed state will continue until the water temperature becomes less then 3.98 °C; at this point lighter and colder water near surface reinforces its position by losing heat due to thermal radiation and making itself even lighter. These physical phenomena, namely, freezing and thawing of water, depth distribution of solar energy absorption, convection due to density differences, wind-driven surface-water mixing, and diffusion of heat due to temperature gradients, are precisely explained by physical laws and principles. In limnology, vertical separation of lake water owing to density difference at depths is termed *thermal stratification*, and the transient periods between winter cold stratification and summer warm stratification is referred to as *spring* and *autumn turnover*. For a sufficiently deep lake this cycle is evident, and stratification cuts deep sediment from the most of the biological happenings in the top part of the lake. This is relevant, because sediment has been known as a source of nutrients and major consumer of oxygen, with great significance to the cause and impact of eutrophication.

Limnology of course goes beyond the knowledge that I have just introduced. For example, food web and biodiversity is a very highly researched area in freshwater systems. Cycles of other elements than phosphorus, for example, nitrogen, silica, and sulphur are also important for understanding of major chemical pathways and related processes. Many complex combinations of chemical reactions and species interactions happen in the benthic environment, and accumulated sediment could be sampled to study the past conditions of the lake and the surrounding (paleolimnology). The phosphorus limitation paradigm that I introduced also is recently more and more disputed with suggested alternative limiting factors among others being nitrogen and light (*Walz and Adrian*, 2008).

3 History of lake models

Numerical models have been created in order to represent some subset of the processes within a lake, although the intention of creating a model varied. Motivation for developing lake models classified as either heuristic or utility purposes. Some models are developed in order to test the conceptual connections among known processes that are felt important, and are used to learn about the system (*conceptual model*). Other models are created and used for real world purposes, in particular, to provide *projection* and *prediction* to help formulate management plans.

Review papers such as the one by *Mooij et al.* (2010) summarise the development of lake models, and their variety in approaches and degree of detail in description of lake (Table 1). Briefly speaking, the earliest lake models are of static chemical model type (average, and non-changing by time) that describes the balance of fluxes into or out of lake to determine the steady-state pools of chemical constituents of the lake. From this type, we saw advancement in modelling more biological and physical aspects of the lake, and most models are now designed to be able to predict temporal changes by including the time dimension. These are called *dynamic* models. Incorporating known processes into models was a logical way forward, and models have become more complex. Increase in complexity is a rather common phenomenon also seen in other domains of environmental modelling. Inflating complexity unfortunately led also to operational hurdles; complex models in general take longer computing time, require more details about lake and inputs, and may become ambiguous and be able to explain everything, but possibly for wrong reasons. Today, a variety of challenges are identified for lake modelling, from the need for geochemical and ecological knowledge advancements to numerical and technical issues. Evidently, the question of what challenge is the most urgent is often subjective. In my opinion, finding an optimal balance between complexity and simplicity is the most critical. I find the question of optimal complexity of lake models interesting, because it is something that is very difficult answer in the traditional science framework. I also think that the question of optimal model complexity is critical in providing useful information for the model user.

Table 1: Comparison of models as found in *Mooij et al.* (2010). MyLake is one of the simplest in the dynamic model category, yet it handles surface ice formation. This increases the relevance of the model to the present application at Vansjø. Further details of the table notations are given in the following. +: fully covered. +/- partially covered. -: not covered. * VOL: Vollenweider; CAE: DYRESM-CAEDYM (1-DV) and ELCOM-CAEDYM (3D); CEQ: CE-QUAL-W2; D3D: DELFT3D-ECO; MYL: MyLake; PCL: PCLake; SHR: Lake Shira Model; IPH: IPH-TRIM3D-PCLAKE; PRO: PROTECH; SAL: SALMO; CHAR: CHARISMA; PIS: PISCATOR. ** These columns do not refer to a specific model but to a group of models instead. *** Category abbreviations: STA: Static model; CDN: Complex Dynamic; MDN: Minimal Dynamic; SIB: Super-Individual-Based; S/PM: Stage-Structured and Physiologically Structured; TBM Trait Based. **** Spatial dimension abbreviations: 0-D: 0-dimensional; 1-DV: 1-dimensional vertical; 1-DH: 1-dimensional horizontal; 2-DV: 2-dimensional vertical; 2-DV*: 2-dimensional vertical using compartments (see figure 2); 3-D: 3-dimensional. All individual models except for VOL were checked by the model developer.

Model name*	VOL	CAE	CEQ	D3D	MYL	PCL	SHR	IPH	PRO	SAL	**	CHA	PIS	**	**
Category***	STA	CDN	CDN	CDN	CDN	CDN	CDN	CDN	CDN	CDN	MDN	SIB	SIB	S/PM	TBM
Spatial dimension****	0-D	1-DV/3-D	2-DV	3-D	1-DV	2-DV*	1-DV	3-D	1-D	1-DV	0-D	2-DH	0-D	0-D	0-D
Stratification	-	+	+	+	+	-	+	+	-	+	-	-	-	-	-
Sediment	-	+	+	+	+	+	-	+/-	-	+	-	+	-	-	-
Littoral zone	-	+	-	-	-	+	-	-	-	-	-	-	+	-	-
# Phytoplankton groups	0	7	3+	3–6	1	3	2	3	10	2–10	1–3	0	0	0	1
# Zooplankton groups	0	5	3+	1–3	0	1	1	1	1	1		0	1	1	1
# Benthic groups	0	6	0	1	0	1	1	1	0	0		0	1		
# Fish groups	0	3	0	0	0	3	0	3	0	0		1	ca. 8	1-2	1
# Macrophyte groups	0	1	3+	0	0	1	0	1	0	0		ca. 5	0	0	0
# Bird groups	0	0	0	0	0	0–1	0	0–1	0	0		1	3	0	0
Hydrodynamics	-	+	+	+	+	+/-	+	+/-	+/-	+	+/-	-	-	-	-
Temperature dynamics	-	+	+	+	+	+	+	+	+/-	+	+/-	+/-	+/-	+/-	+/-
Oxygen dynamics	-	+	+	+	-	+	+	+	-	+	-	-	-	-	-
CO_2/DIC dynamics	-	+	+	+	-	-	-	-	-	-	-	-	-	-	-
DOC/POC dynamics	-	+	+	+	-	+	+	+	-	+	-	-	-	-	-
Microbial dynamics	-	+	+	+	-	+/-	+/-	+/-	-	-	-	-	-	-	-
P loading	+	+	+	+	+	+	+	+	+/-	+	+/-	+/-	+/-	+/-	+/-
N loading	+	+	+	+	-	+	+	+	+/-	+		+/-	+/-		
Internal P dynamics	-	+	+	+	+	+	+	+	+	+	-	-	-	-	-
Internal N dynamics	-	+	+	+	-	+	+	+	+	+	-	-	-	-	-
Internal Si dynamics	-	+	+	+	-	+/-	-	+/-	-	-	-	-	-	-	-
Sedimentation/resuspension	+	+	+	+	+	+	+	+	+	+	+/-	-	-	-	-
Diagenesis	-	+	+/-	+	-	+/-	-	-	-	-	-	-	-	-	-
Fisheries	-	+/-	-	-	-	+	-	-	-	+/-	+/-	-	+	+	+
Dredging	-	-	-	+	-	+	-	-	-	-	-	-	-	-	-
Mowing	-	-	-	-	-	+	-	-	-	-	-	+	-	-	-

4 The model

The model used in this thesis is called MyLake. The MyLake model is a *'one-dimensional process-based model code for simulation of daily vertical distribution of lake water temperature and thus density stratification, evolution of seasonal lake ice and snow cover, sediment water interactions, and phosphorus-phytoplankton dynamics'* (*Saloranta and Andersen*, 2007). Compared with other lake models, MyLake has a strong flavour of physical representation of a lake, and a very basic and aggregated representation of biological processes. The temporal and physical resolution is minimalistic: time and depth dimensions only, on a rather coarse grid compared to many other physical models. The design principle of the model is to keep the processes accountable, attempted by incorporating primarily established physical laws and simplifying or avoiding processes that are less established. The model is therefore not able to simulate certain processes such as trophic interactions among species, resource competition among species at the same trophic level, or detailed chemical speciation in the sediment under varying chemical reduction states.

The design and decision to simplify or exclude processes come from various grounds. One primary reason for being simple is accountability; simple model construction reduces the risk of unintentionally faking the fit during the calibration process. It has been known that the more processes and details are added, the easier it becomes to appear consistent with independently taken observation for model testing purpose (see next section). The simple construction also makes it possible to explain deviations of model simulations, if they appear to be governed by parts of the system that are excluded or simplified. One would expect the model to operate poorly on these missing processes.[1] Another primary reason for the simple construction is the availability of independently taken observation to challenge the model simulation. Modelling long-term processes such as the nutrient sourcing from the sediment requires long-term observation data. In our case and in many other applications, observation record for decades of change in sediment-water interaction is unavailable, and therefore makes the application arduous and uncertain. Another advantage is MyLake's implementation of ice-heat dynamics (*Fang and Stefan*, 1996; *Leppäranta*, 1993), which is not present in many lake models (an exception is the FLake model (*Mironov et al.*, 2010)), despite being an essential seasonal phenomenon in northern lakes. Because almost all the lakes that were considered in the present thesis or the relevant projects freeze during the winter, this was deemed a necessary constituent.

Also important, no model is perfect and it is impossible to fully justify the choice of the model, or the model construction and design. Thus, with everything taken into consideration, the choice of model remains subjective. In our case it is true that the model was chosen

[1] One example quote describing such a situation is: '*Models are often more interesting when they fail than when they succeed*' (*Aber and Driscoll*, 1997). This perspective was used to provide depth in their discussion.

also on the ground of familiarity. It is also true that the model does not explain variations among different algae species, or their interactions with their predators. In addition, it is true that model does not simulate horizontal gradients of the lake, nor does it manage water level fluctuations; these were assumed to be negligible. Therefore, sincerely put, the model choice must be considered a provisional decision. One way of reducing the subjectivity about the model choice is to incorporate more than one model and use them in a similar manner (for example, *Solomon et al.*, 2007; *Tominaga et al.*, 2010). But the logistic constraints in the project regrettably hindered parallel applications of many other models. Nonetheless, a limited model comparison study was conducted on the catchment models (see the relevant section).[2]

The MyLake model is rather new compared with many other lake models, and applications had been so far limited only in Norway, Finland, and Canada (*Dibike et al.*, 2012, 2011; *Kankaala et al.*, 1996; *Saloranta et al.*, 2009). The main pedigree of MyLake is the Minlake model (*Riley and Stefan*, 1988), which was developed and been applied in Minnesota, USA many times.[3] Overall, the model applications have shown that the model is reliable especially in the temperature and ice phenology dynamics. Some aspects of the model that we feel needed challenges are addressed now in the present thesis.

5 Numerical modelling as a subject discipline

In our context, the word 'model' is used to represent the complex real world in a simplified manner. The present thesis revolves around the concept of process-oriented dynamic models. Contrasting process-oriented dynamic models with empirical stiatistical models, which are perhaps more commonly used in ecology, is useful for clarification. Examples of empirical models include traditional statistics, which are for example in biology more commonly used to derive empirical relationships between variables by taking samples or conducting experiments. Empirically based statistical models are useful in their own right; however, they do

[2] There is also a community-based approach in model creation that are currently being developed so that many different model choices could be more easily compared (*Trolle et al.*, 2012).

[3] MyLake was named as such in order to reflect an attitude among not all but some limnologists who recognise that there are a vast variety of lakes with different processes and dominant phenomena, and say to another limnologist, 'Yes, your theory might be right in your lake but in *my lake*, the situation is different...' The name also reflects its ability to model ice and snow on surface, as this enables simulation over winter, hence the *multi-year* (*my*) lake modelling. Furthermore, the model name is a multi-lingual word play. MyLake is based on a model Minlake from Minnesota. MyLake was developed in Norway, and in Norwegian one of the first-person posessive pronoun forms (equivalent to the English word *my*) is *min*, so MyLake is indeed Minlake. Although originally not intended by the model authors, there is one more inter-linguistic connection with these two models. The state name Minnesota is named after a river of the same name, which derives the first part of its name (*minne*) from the word in the Dakota language that means water (transcribed as *mni*, *mini*, or *minne*) and is also used for various water bodies. In Norwegian, a typical word for water is *vann* or *vatn*, which is often used for naming lakes. So the first part of the name MyLake (*my*) goes around a long way to come back to the second part of the name (*lake*).

not explain directly the mechanisms in the natural world. Process-oriented models instead *consist of* mechanisms. Some of the processes in process-oriented models may also include empirically derived relationships to represent an aggregated process, but the aim remains the same: to describe the system by orchestrating processes.

Many types of process-based models exist, but they have common features; they have temporal and spatial boundaries and within the boundaries there is a system. To a modeller, a system means a set of state variables that are connected with processes. Processes are mathematically described, *i.e.*, by formulae, and these formulae come in many flavours. Some formulae are arithmetic and some are differential equations describing rates of change. For a model author it is important that these formulae are trustworthy (the best likely knowledge) and not overly complex (Occam's razor). Decisions about model design, from the spatiotemporal dimensions to the degree of details, are author specific and subjective. So, reductionistically put, any process-based model is a collection of deliberately chosen formulae.

As such, the question of credibility arises when a model is used for a real-world application. From the philosophical point of view, process-based models cannot be supported in the same manner as a scientific hypothesis is supported through Popperian falsificationism (but see *Tarantola* (2006) for a treatment of how a model application could be seen falsifiable). One way to gain confidence in model usage is instead through comparison between the model output and independently taken observation counterpart. The common practice is to accept the model if its output resembles separately taken observation. What is considered similar enough and therefore good enough for credibility is arbitrary decided. In this paradigm, the observations are considered the absolute target although they must at least include measurement errors. Perhaps even more frequently conducted is the parameter adjustment procedure (‘*calibration*’) where parameters governing the formula composite (*i.e.*, model constituents) are varied until the model output better resembles the target observation. With this, the acceptance level still remains arbitrary, and furthermore, the model user has to decide when to stop adjusting. This approach may seem counterintuitive and unaccustomed to a natural scientist, who often relies on the hypothetico-deductive scheme or Popperian falsificationism. But this is currently the most common approach in applied environmental modelling and most modellers characterise the problem as an *inverse problem*, in which a model is chosen and its parameters are calibrated to reproduce the observation as close as possible. The usual rationale is that a model should have predictive power for the future, because the model itself is based on previously established mechanisms and is able to produce the observation counterpart after adjusting parameterisation of the mechanisms. It is usually assumed that adjustment of parameters for process formulae does not undermine the mechanistic representation.[4]

Today, the art of model calibration goes beyond subjective manual adjustment of model parameters in an effort to 'aim' and 'shoot' the target observation. Here, I would like to describe popular techniques that are relevant to this thesis. For simplicity most of the description assumes that there is only one parameter that we wish to calibrate. But these methods are all suitable even if we have multiple parameters to calibrate simultaneously for a particular model, as conducted in the included papers.

In a *Monte Carlo method* (see, for example, *Mosegaard and Tarantola*, 1995), parameter values are repeatedly taken randomly and a range of outputs are produced. This removes the subjectivity about parameter value choices by allowing ambiguous (*i.e.*, a range of) outputs. The range of outputs represent uncertainty due to what we do not know precisely about the parameters. Today, honest description of this uncertainty is increasingly appreciated. Relating to this uncertainty, there is a concept of *equifinality* that describes the situation where multiple different combinations of parameters in a highly non-linear and complex model produce equal or similar outputs (hence the word *equi-final-ity*) (*Beven and Freer*, 2001). The Monte Carlo method therefore has a close affinity with the equifinality phenomenon and is in principle capable of containing the equifinality of the model. The Monte Carlo method was once a popular calibration method (and is still no less invalid today) because the range of outputs could then be compared against the observation and a modeller could decide which parameter value is most suitable for corresponding to observation, or choose a subset of parameter values that produce outputs that are within an acceptable level of correspondence. But for the calibration purpose, the Monte Carlo method is now less preferred because of its computational inefficiency. As the number of model parameters increases, the number of all possible combinations of parameter values increases exponentially, even with a Latin hypercube sampling.

The *Markov chain Monte Carlo (MCMC) method* (see, for example, *Andrieu et al.*, 2003) tries to circumvent the computational inefficiency of the 'classical' Monte Carlo method, described above. MCMC is an automatic machine learning method for finding optimal parameter values. MCMC works in sequence; at each step the performance of a new parameter

[4] Example demonstration of MyLake processes upon calibration is given here. Let us say we increase the effectiveness of wind to mix the water column. It means that during the summer more heat will be distributed to the deeper part of the lake, changing the temperature distribution at depths. The change of the temperature affects algae growth. The model outputs (temperature and phosphorus concentrations) can then be compared against the observation counterpart, such as temperature and algae concentration at depths, to see whether the increase in effectiveness of wind mixing was appropriate. A relevant complication is that the temperature-specific algae uptake rate also influences the phosphorus fractions in water and an increase in algae would in turn increase the light attenuation, which again affects the heat distribution of solar radiation. So we have a complex relation between the combination of parameter values to the set of model outputs that can be compared to corresponding observations. The relation is dynamic (temporally changing) and non-linear (not just a matter of scaling and summation). A typical model application of MyLake involves many more parameters for many more processes resulting in many more output variables to be compared to multiple series of observations.

value is tested against that of its preceding parameter and decision is made regarding whether or not to accept and move to the new value. Sometimes the new parameter value that immediately does not perform as well as the preceding value can be accepted and chain can move to the inferior value (Metropolis-Hastings algorithm), in order to avoid getting trapped in a local good value when the best value lies somewhere else (*Metropolis and Ulam*, 1949). After a while the sequence becomes long enough and the sequence itself is a collection of values that are optimised.[5]

One of the most recent developments in this direction is the *DREAM (Differential Evolution Adaptive Metropolis) method* (*Vrugt et al.*, 2009). DREAM is a variation of MCMC and addresses a shortcoming of an ordinary MCMC, namely about determining chain convergence (see footnote 5). A typical MCMC before DREAM had only one chain of parameter, or in case where multiple parameters were varied at the same time, one set of parameter chains. DREAM uses instead multiple sets of parameter chains, and these sets of parameter chains at each step 'communicate' among each other to decide on acceptance, and propose the new candidate parameters. Simply put, DREAM is operating multiple MCMCs at the same time. Because DREAM runs multiple sets of chains at the same time, the parameter range is more thoroughly searched and convergence can be more easily and confidently concluded. This can be favourably contrasted to the ordinary MCMC, in which convergence is hard to distinguish from interim stagnation of parameter chains. When model calibration is in mind, multiple chains are gathered and parameter values are collected.

[5] A more technical overview of a typical random walk Metropolis-Hastings algorithm MCMC variation in environmental modelling, under the Bayesian inference paradigm, with the assumption that the observation error is Gaussian distributed, is given in the following. A candidate parameter is drawn from the possible value range (the proposal probability distribution) and the model output using the parameter value is produced. By collecting the errors of these outputs against the target observation, one gains likelihood of that particular parameter value, using the error distribution that was assumed beforehand. By keeping the parameter-likelihood pair and using a proposal probability distribution, a new parameter is sampled. Using this new candidate parameter we calculate the new corresponding model output likelihood in the same manner. We then make a decision whether or not accept the new parameter, taking into account both i) how more probable the new parameter is compared with the that by the old parameter according to a priori knowledge, and ii) how more likely the model outcome using the new parameter compared with that using the old parameter. One way to make the acceptance decision is to do it probabilistically using the prior probability and the likelihood that have just been evaluated, but in such a way that the less probable and less likely results are accepted sometimes (Metropolis-Hastings algorithm). Once the decision is made, we either adopt the new parameter or keep the old parameter, and around the chosen parameter another new candidate parameter is sampled, and the procedure is repeated until the chain of parameter appears converged. It is not always possible to know the variance of the assumed Gaussian distribution of observation error that is required for likelihood calculation, and therefore this can be sometimes included in a list of parameters to estimate (*Gelman et al.*, 2003). If calibration of a model is concerned, we can take the resulting chain and collect the values to use for the model parameter in subsequent simulation analyses. Typically, if the parameter is sensitive and therefore changing its values changes the model outcome substantially, the range of collected values from the converged chain (often referred to as the posterior distribution) becomes narrower. Disadvantages of this type of MCMC are that the number of chain steps it takes to truly converge depends on which parameter value to start with, and that one would never be completely confident that the chain has converged, because the algorithm might not yet have searched everywhere possible.

With all these advancements in calibration methods, one may get a false impression that the models are becoming more accurate and more reliable. This is by no means implied and not necessarily true. Let us remind ourselves that the term calibration is in fact a jargon used by environmental modellers; the word choice is rather unfortunate because the word calibration is in ordinary context used to describe adjustment of a *high accuracy instrument* against some *absolute standard*. I write unfortunate because neither of these terms are very prominent in the model calibration procedure. First, an environmental model is simply an aggregation of formulae and mechanisms and as a whole it does not have the accuracy status of instrument such as a thermometer; an environmental model may be partially consisting of formulae that are not universally accepted among scientists in the particular field of study, and they carry the equifinality issue as described above. Second, the model calibration is done against real world observation, but it is impossible to avoid error in observation due to inaccuracy in measurement and error due to inherent temporal fluctuation (e.g., weekly samples may miss an episode that happens within the course of a few hours). Therefore, the observed data to be matched against never have the status of an absolute standard. Hidden jargons such as the terminology 'calibration' require extra attention in the context of interdisciplinarity because it may cause misunderstanding. The same precaution should be given to even stronger words that implies confirmation, such as '*verification*' and '*validation*,' both of which are frequently found in the literature in environmental models. Precise meanings of these terms might be agreed among modellers, but these words are somewhat illegitimately and unfortunately chosen, and they are misleading.[6] See *Oreskes et al.* (1994) for a rigorous examination of these words in the environmental models.

Thus, model calibration in any way does not verify the model itself and its application. It is simply a popular practice whose sole purpose is to adjust the model to observations, to address the parameter uncertainty of environmental models, and to explore the variation in outputs due to the choice of parameter values. Other important sources of uncertainty in environmental models are:

- Structural uncertainty: Uncertainty due to the way the system is represented (use ensemble of models to minimise this, but still not possible to eliminate).

- Forcing data uncertainty (*Vrugt et al.*, 2008): Uncertainty in the accuracy of the data that are fed into the system (e.g., weather or nutrient loading in the case of MyLake). If the models are used in a sequence, all the uncertainty in the upstream model will cascade over to the downstream model.

[6] Although he may or may not agree with every detail of these points here, I would like to acknowledge my colleague J. Starrfelt, for my observations I have written here are a fruit of a casual but enlightening conversation with him.

- Projection uncertainty: Our inability to know the future conditions. If we use models to predict future statuses of the environment, the conditions that we pose on the model may be different in the real time future.

- Epistemological (Epistemic) uncertainty: Uncertainty due to the fact that we do not know beforehand what we do not know, but we know that there exist something we do not know.

- Technical error: There is always a possibility that the researchers and model users may have made mistakes in creating the model code and in administrating the model.

We have seen the present modelling-specific issues that many limnologists and scientists who are not modellers themselves may feel unfamiliar with. The art of modelling has very different operating principles than the Popperian definition of science. In my opinion, the most curious and striking about modelling in contrast to science is its arbitrariness. The design and formulation is arbitrarily chosen due to authors' opinion or their familiarity (own experiences or colleagues expertise) rather than scrutinising all candidate choices. The decision for not using other equally applicable models is often not well justified either. The 'calibration procedure' is inherently arbitrary. However, despite these properties, environmental models have been considered useful and have gained a strong share in the academic literature. The core source of confidence in the model application remains the same (*Knutti*, 2008): the models are mechanistic themselves, and with calibration they are able to reproduce the observation counterparts. This thesis focuses much on the physical and chemical processes of lakes, which are mostly based on first principles, or widely accepted 'laws' of nature. At the same time, deliberately excluded are probably important knowledge and concepts of community ecology, nitrogen limitation, and sediment interactions. We are disregarding these processes either because they cannot be adequately formulated or because we may be lacking adequate data to confront the model with. In the Eutropia project, the biological response to change in lake's physical and chemical conditions will be statistically predicted using empirical models rather than from dynamical models.

6 Present challenges, approaches and findings

Within the framework of the Eutropia project, my role was to set up the MyLake model in the two basins of Vansjø (Storefjorden and Vanemfjorden, Figure 2), connect the lake model with two catchment models (INCA-P (*Wade et al.*, 2002) and SWAT (*Gassman et al.*, 2007)), and improve the simulation. We identified challenges in achieving them, and our attempts in addressing these were summarised in Table 2, and the details can be referred to in the included papers (Papers I to IV).

Table 2: Summary of identified challenges, approaches and findings. Catchment models were arranged by my colleagues (primarily, A. M. Engebretsen for SWAT and J. Starrfelt for INCA-P, with many more people who were involved in collecting and collating required data). I adopted the DREAM method implemented by J. Starrfelt, for calibrating MyLake. Technical assistance with internals of the MyLake model was given by T. Saloranta and T. Andersen. The Bayesian Belief Network application was coordinated by D. N. Barton.

Challenges	Approaches and relevant knowledge	Findings and suggestion
MyLake's ability of simulating winter chemistry has never been evaluated.	We used Årungen, a lake located approximately 30 km north of Vansjø, where a more detailed loading data and in-lake winter chemistry data exist, to test MyLake (Paper I).	MyLake was able to perform well for winter chemistry (Figure 3).
Long-term climate change and short-term interannual variation in weather may be a confounding factor.	We conducted a large scale regional study in Nordic countries (Paper II) and effect decomposition study of year-to-year variation at Årungen (Paper I).	Vansjø may experience a large change in mean summer water temperature, and longer ice-free period upon climate change (Table 3). Short-term year-to-year stochasticity in weather can also affect the lake water quality (Figure 4).
The catchment is large and the model setup is resource demanding.	We used the simpler model INCA-P for loading to the larger Storefjorden basin while SWAT with detailed settings was set up for the smaller western Vansjø catchment. SWAT is more detailed and is able to simulate effectiveness of various abatement measures. This SWAT utility was concentrated on the western catchment into the Vanemfjorden basin, where the eutrophication impacts are considered less desirable (Paper III).	The model connections among the four waterbodies, namely, the eastern catchment (INCA-P), the Storefjorden basin (MyLake), the western catchment (SWAT and INCA-P) and the Vanemfjorden basin (MyLake) (see Figure 2 for the map), were properly set up (see Table 4 and Figure 5 for summary results).

Table 2: Summary of identified challenges, approaches and findings. (continued)

Challenges	Approaches and relevant knowledge	Findings and suggestion
Some of the in-lake processes were not very well known.	Identification and quantification of these processes by modelling studies are still difficult at this point. Continued efforts in chemical studies around the catchment are ongoing as of writing. A modelling experiment was however conducted in this thesis, in which the water input from the Storefjorden basin was separated in order to identify the role of this water in determining the water quality of the Vanemfjorden basin (Paper III).	The modelling experiment supported that the water from the Storefjorden basin is always a diluting agent for local runoff loading at the western catchment. This means that the management of the local catchment is crucial in improving the water quality of the Vanemfjorden basin. The findings also meant that the Vanemfjorden basin always has a water quality condition that is worse than the Storefjorden basin. This is primarily because Storefjorden is able to sink phosphorus by sedimentation because it is deep. On the other hand, the Vanemfjorden is shallow and has a limited ability in removing phosphorus in the lake water (Table 4 and Figure 5).
Probabilistically based calibration methods such as the ones we employ in Eutropia are effective in describing parameter uncertainty of a model, but this was often not treated properly when models were used in sequence. The magnitude of uncertainty due to such negligence was not evaluated before.	We deviced a simultaneous calibration in order to assess the extent of such uncertainty (Paper IV), and we also studied how parameter uncertainty of our models can cascade downstream (Paper III).	Parameter uncertainty of both the catchment models and the lake model are significant, and therefore they we need to take them into consideration (Figures 5 and 6).
Further cross-disciplinary assessment (*i.e.*, beyond the limnology-modelling theme of this thesis, such as further cooperation with social and economical disciplines) is needed.	A Bayesian Belief Network (BBN) model that connects all the models and expert knowledge based on probabilstic understanding (see, for example, *Barton et al.*, 2012) is being set up as of writing.	The BBN was setup for the natural science models and this operates as expected as of writing. Further connections are to come this year.

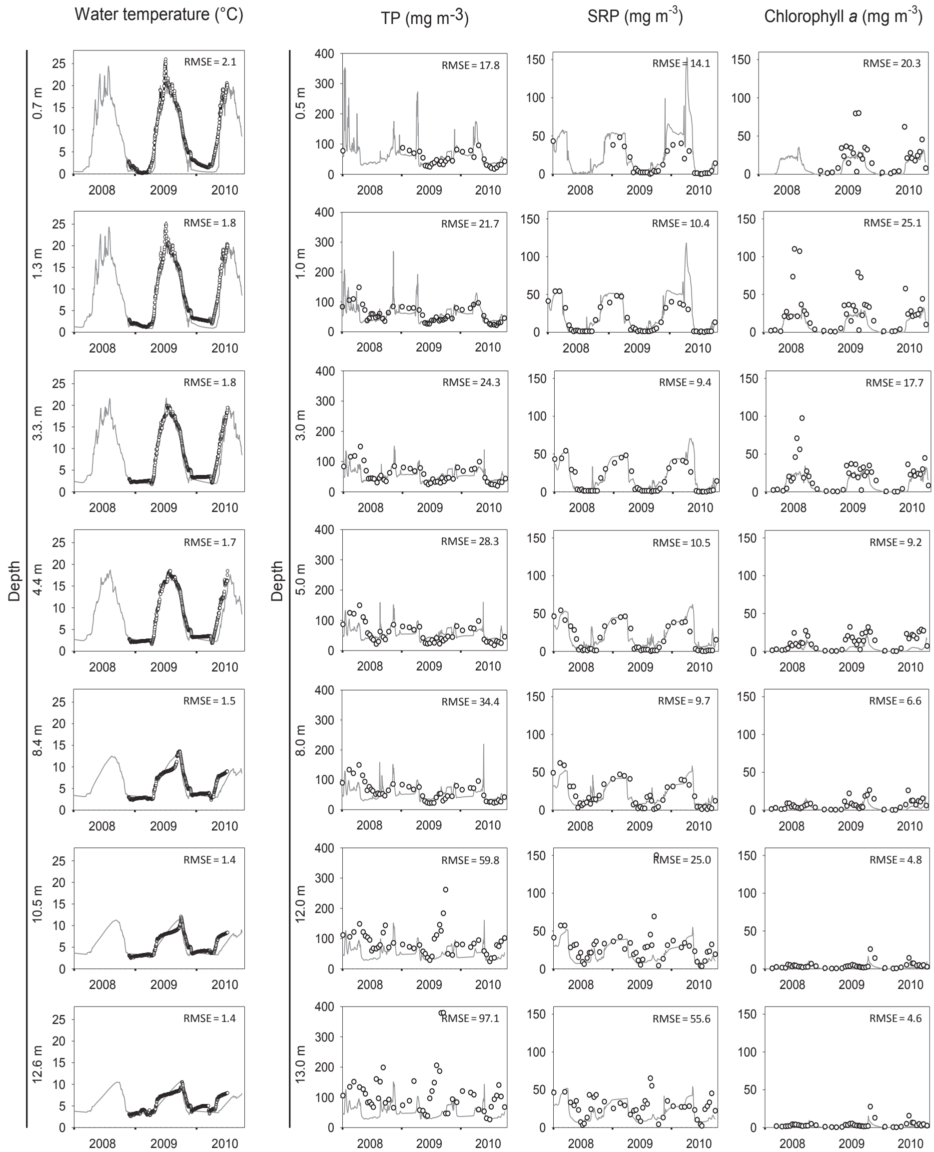


Figure 3: MyLake's performance at Årungen for all seasons. The circles represent lake observations whereas the lines are MyLake's simulation after calibration. Winter in-lake data at Vansjø are scarse, so a separate model test such as this increases confidence in the model through multiple years. Root mean square error (RMSE) is a convenient statistic that shows on quadratic average how far a model simulation is from the lake observations. RMSE has the same unit as the measurement unit. Lake observation data and figure production mainly by A. T. Romarheim. I performed the calibration. This Figure is taken from Paper I.

Table 3: Forcasted climate change impacts on lake physical condition (median of 2070-2100 compared to median of 1970-2000). These forecasts are based on generalised additive models from Paper II and are produced using a generic lake type that resembles Vansjø Storefjorden at a grid location in the lake. See the manuscript for details.

Variable	Change
First ice-on	27 days later
Last ice-off	24 days earlier
Mean ice thickness between the first ice-on and the last ice-off	16 cm thinner
Duration of spring turnover	9 days longer
Duration of warm stratification	11 days longer
Duration of autumn turnover	28 days longer
Mean temperature for the top two metres in January	0.2 °C warmer
Mean temperature for the top two metres in February	0.1 °C warmer
Mean temperature for the top two metres in March	0.2 °C warmer
Mean temperature for the top two metres in April	0.8 °C warmer
Mean temperature for the top two metres in May	2.1 °C warmer
Mean temperature for the top two metres in June	3.6 °C warmer
Mean temperature for the top two metres in July	3.0 °C warmer
Mean temperature for the top two metres in August	2.5 °C warmer
Mean temperature for the top two metres in September	2.4 °C warmer
Mean temperature for the top two metres in October	2.8 °C warmer
Mean temperature for the top two metres in November	1.8 °C warmer
Mean temperature for the top two metres in December	0.8 °C warmer

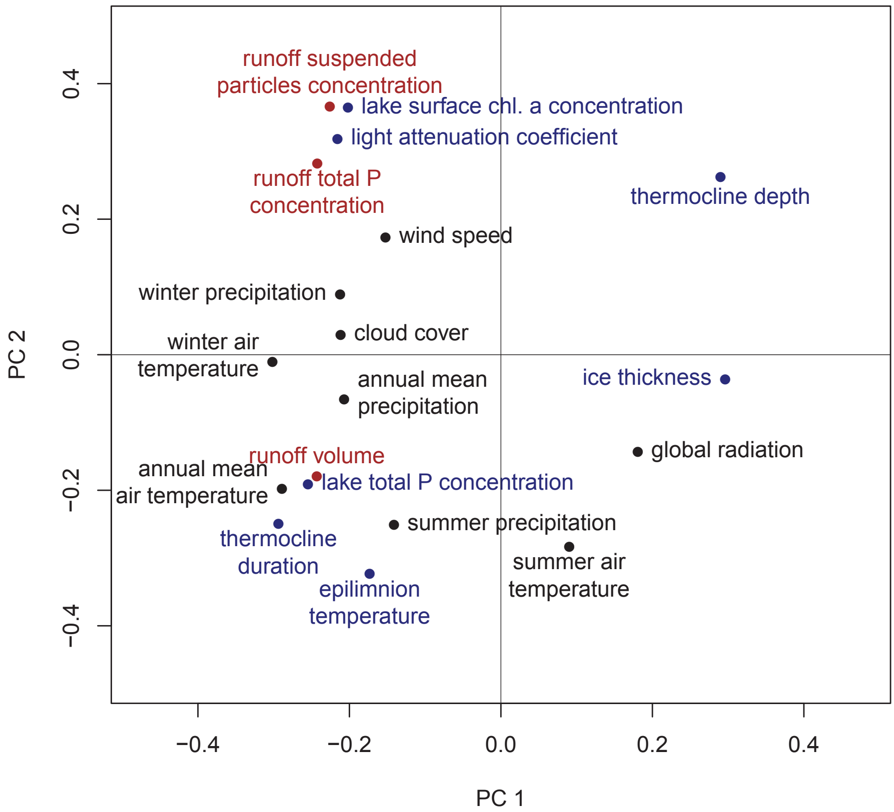


Figure 4: Ordination plot among the weather inputs (black), loading inputs (brown), and the model outputs (blue) using annually based aggregations of a MyLake application at Årungen for 16 years. Only the loadings from the first two principal components are shown. The variables that occur in the same direction from the origin or in the opposite direction are closely related. This Figure is taken from Paper I.

Table 4: Summary of model connection configurations for MyLake at the Vanemfjorden calibration and contrasts among the configurations. Corresponding results are shown in Figure 5. Calibration A refers to calibration using the best SWAT input together with median Sundet whereas Calibration B refers to calibration using INCA-P's median together with median Sundet as the input. Sundet is the unidirectional channel between the upstream Storefjorden basin and the downstream Vanemfjorden basin. This table is taken from Paper III; see the manuscript for information regarding these calibrations.

				Uncertainty assessment				Experiment	
Model configuration	Sundet alternative	Vanemfjorden local catchment alternative	Vanemfjorden MyLake basis calibration inputs	I	II	III	IV	V	VI
1	Median	SWAT's 95-percentile	Calibration A	x				x	
2	Median	SWAT's best	Calibration A	x		x	x	x	x
3	Median	SWAT's 5-percentile	Calibration A	x				x	
4	Median	INCA-P's 95-percentile	Calibration B		x				
5	Median	INCA-P's median	Calibration B		x	x			
6	Median	INCA-P's 5-percentile	Calibration B		x				
7	Median	INCA-P's median	Calibration A			x			
8	5-percentile	SWAT's best	Calibration A				x		
9	95-percentile	SWAT's best	Calibration A				x		
1v	Median	SWAT's 95-percentile with VFS	Calibration A					x	
2v	Median	SWAT's best with VFS	Calibration A					x	
3v	Median	SWAT's 5-percentile with VFS	Calibration A					x	
10	Not used	SWAT's best	Calibration A						x
11	Median	Not used	Calibration A						x

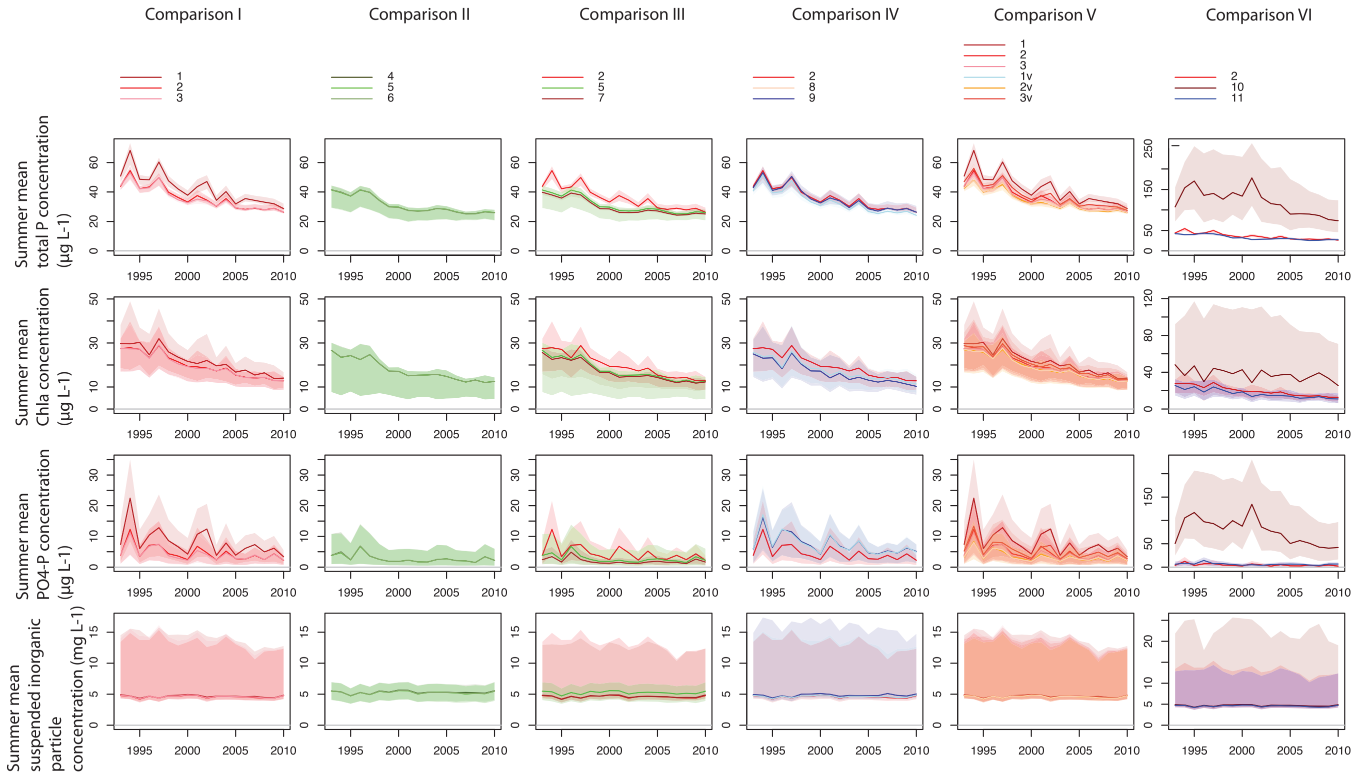


Figure 5: Main results of model connections after Table 4. Comparison of output of MyLake at the Vanemfjorden basin using total of 14 model configurations are shown here. The solid lines are MyLake's median around the OPU ('own parameter uncertainty'). The shades are the 5- and 95-percentiles of the same OPU are shown by shaded area of the same colour. Summer mean concentration here is defined as the mean of daily values over the three months (JJA). The figure is taken from Paper III; see the manuscript for description of OPU and other details.

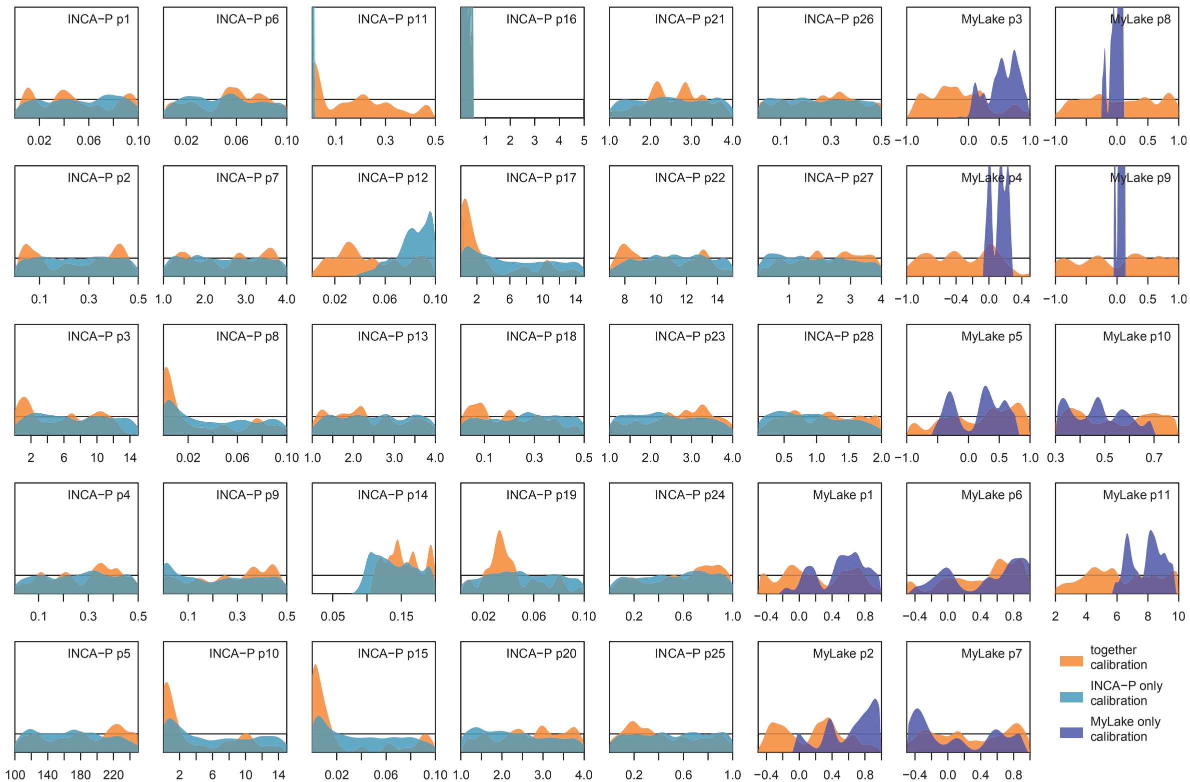


Figure 6: Parameter uncertainty of INCA-P and MyLake as applied on the Storefjorden watershed. Posterior probability density functions for calibrated parameters (*i.e.*, after-calibration parameter values) are shown. The 'interaction' between INCA-P parameters and MyLake parameters are illustrated; the optimum parameter values become different when these models are calibrated together. Calibration schemes are overlaid to emphasise the difference in the calibrated parameter density. Gaussian normal kernel density was assumed to produce the density functions. The prior uniform distributions are indicated here by the horizontal line. This figure is taken from Paper IV, and see the manscript for details of these parameters and other design aspects.

6.1 Other notable findings and relevance to other studies in the project

The Årungen application showed that MyLake was capable of simulating lake physics and chemistry at all depths for all seasons (Paper I). The observation reproducibility was in fact better than MyLake applied at the Vansjø basins. There can be many possible reasons for this. First, Årungen has a much simpler bathymetry. Second, Årungen is much more eutrophic. But in my opinion the most crucial reason is most likely because the loading estimate to the lake was more relevant and accurate. In Paper I we used scaled up the volume-weighted total loading data at the Skuterud observation site located upstream of Årungen. This provided loading data that are accurate in the amount of phosphorus and sedimenting particles. In Vansjø, such data are not available as river samples at Hobøl or Guthusbekken only represent instantaneous concentration separated by periods of weeks (see Paper III). The catchment models are therefore calibrated to data that may not be ideal for lake water quality, although that was the best data available for us. Despite this, the monitoring data from the watershed and the lake were indispensable for the modelling studies.

Because of the very design of the model, the simulation outcome honestly depicts the limitation of included processes. In this sense, an apparent failure of the model can be considered a virtue. In case of an apparent success of the model, one will always need to keep in mind the possibility of the model appearing to be correct for wrong reasons. From my experiences, there were two notable deviations of MyLake simulations from lake observations: the 2006 algae bloom in Vanemfjorden and the 2008 spike in inorganic particulate and total phosphorus conentrations in Storefjorden (see Paper IV). In both cases MyLake honestly showed where omission of unknown processes was significant. Many hypotheses have been created to explain the 2006 algae bloom in Vanemfjorden. They may for example have the perspectives of agricultural studies, biogeochemistry, lake ecology, algae physiology, meteology or combinations of these. These explanations may focus on catchment processes or in-lake processes. In the Eutropia project, many participants from the Department of Chemistry at the University of Oslo have been studying biogeochemical hypotheses (relating to recent reduction in pH in surface waters and iron-phosphorus reduction-oxydation reactions, for example). In principle, once these processes are scientifically identified, they can be incorporated in catchment models or lake models. However, it seems to me that at the moment of writing this thesis there may not be relevant data to test a model construction using such a newly identified processes in Vansjø and its catchments even if these hypotheses are well supported. The 2008 spike in inorganic particulate phosphorus concentration is probably caused by a major landslide happened along the river Hobøl. A record high total phosphorus observation gained much attention from the management bodies. However, this was not coupled with a large algae bloom. These observations seem to point to a possibility that much of phosphosrus introduced after the landslide may be detected in laboratory analyses and therefore reported, but may not be important in other in-lake phosphorus cycles, such as algae growth. If this is the case, both catchment and lake models may improve simulation by for example introducing a separate phosphorus variable that does not participate in many phosphorus transformation processes but is still detected in chemical analyses. But the difficulties remain in predicting such a sudden, accute, and abrupt incident by an ordinary deterministic models such as those used in the project. In addition, we may lack relevant data to test such a model construction. From the management point of view, accountability of identified processes are important, in order to gain cooperation among the stakeholders. Accountability of certain processes can be gained for example by quantification, mass balance, or knowledge conforming to studies from other similar systems. From this perspective, the current modelling approaches in Eutropia such as the one presented in Paper IV or the one to be presented using a Bayesian belief network should provide useful knowledge for management purposes.

We also reckon that the project's lifespan (three years) was short compared with the actual changes we expect in the future, both owing to management measures and to the short-term and long-term changes in the nature. Monitoring (*Lovett et al.*, 2007), evaluation of management measures, and improvement in modelling all require future commitments.

7 Concluding remark

I would like to end this thesis by stating that interdisciplinarity of the Eutropia project across natural and social sciences was crucial in conducting these studies. It enabled my attention on both the modelling and limnology disciplines, but it also gave me a rewarding feeling that my findings have wider relevance to real world concerns.

8 References

Aber, J. D., and C. T. Driscoll (1997), Effects of land use, climate variation, and N deposition on N cycling and C storage in northern hardwood forests, *Glob. Biogeochem. Cy.*, *11*(4), 639–648.

Andrieu, C., N. de Freitas, A. Doucet, and M. I. Jordan (2003), An introduction to MCMC for machine learning, *Mach. Learn.*, *50*(1), 5–43.

Barton, D. N., S. Kuikka, O. Varis, L. Uusitalo, H. J. Henriksen, M. Borsuk, A. de la Hera, R. Farmani, S. Johnson, and J. D. Linnell (2012), Bayesian networks in environmental and resource management, *Integrated Environ. Assess. Manag.*, *8*(3), 418–429.

Beven, K., and J. Freer (2001), Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology, *J. Hydrol.*, *249*(1–4), 11–29.

Dibike, Y., T. Prowse, T. Saloranta, and R. Ahmed (2011), Response of northern hemisphere lake-ice cover and lake-water thermal structure patterns to a changing climate, *Hydrol. Process.*, *25*(19), 2942–2953.

Dibike, Y., T. Prowse, B. Bonsal, L. d. Rham, and T. Saloranta (2012), Simulation of north american lake-ice cover characteristics under contemporary and future climate conditions, *Int. J. Climatol.*, *32*(5), 695–709.

Duarte, C. M., and O. Piro (2001), Interdisciplinary challenges and bottlenecks in the aquatic sciences, *Limnol. Oceanogr. Bull.*, *10*(4), 57–61.

Fang, X., and H. G. Stefan (1996), Long-term lake water temperature and ice cover simulations/measurements, *Cold Regions Sci. Technol*, *24*(3), 289–304.

Forbes, S. A. (1887), The lake as a microcosm, *Bull. Ill. Nat. Hist. Surv.*, *15*, 537–550.

Gassman, P. W., M. R. Reyes, C. H. Green, and J. G. Arnold (2007), *The Soil And Water Assessment Tool: Historical Development, Applications, And Future Research Directions*, Center for Agricultural and Rural Development, Iowa State University.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin (2003), *Bayesian Data Analysis, Second Edition*, 2 ed., Chapman and Hall/CRC.

Kankaala, P., L. Arvola, T. Tulonen, and A. Ojala (1996), Carbon budget for the pelagic food web of the euphotic zone in a boreal lake (lake pääjärvi), *Can. J. Fish. Aquat. Sci.*, *53*(7), 1663–1674.

Knutti, R. (2008), Should we believe model predictions of future climate change?, *Phil. Trans. R. Soc. A*, *366*(1885), 4647–4664.

Leppäranta, M. (1993), A review of analytical models of sea-ice growth, *Atmos. Ocean*, *31*(1), 123–138.

Lovett, G. M., D. A. Burns, C. T. Driscoll, J. C. Jenkins, M. J. Mitchell, L. Rustad, J. B. Shanley, G. E. Likens, and R. Haeuber (2007), Who needs environmental monitoring?, *Front. Ecol. Environ.*, *5*(5), 253–260.

Metropolis, N., and S. Ulam (1949), The monte carlo method, *J. Am. Stat. Assoc.*, *44*(247), 335–341.

Mironov, D., E. Heist, E. Kourzeneva, B. Ritter, N. Schneider, and A. Terzhevik (2010), Implementation of the lake parameterisation scheme FLake into the numerical weather prediction model COSMO, *Boreal Environ. Res.*, *15*(2), 218–230.

Mooij, W. M., D. Trolle, E. Jeppesen, G. Arhonditsis, P. V. Belolipetsky, D. B. R. Chitamwebwa, A. G. Degermendzhy, D. L. DeAngelis, L. N. D. S. Domis, A. S. Downing, J. A. Elliott, C. R. Fragoso, U. Gaedke, S. N. Genova, R. D. Gulati, L. Hakanson, D. P. Hamilton, M. R. Hipsey, J. 't Hoen, S. Huelsmann, F. H. Los, V. Makler-Pick, T. Petzoldt, I. G. Prokopkin, K. Rinke, S. A. Schep, K. Tominaga, A. A. Van Dam, E. H. Van Nes, S. A. Wells, and J. H. Janse (2010), Challenges and opportunities for integrating lake ecosystem modelling approaches, *Aquat. Ecol.*, *44*(3), 633–667.

Mosegaard, K., and A. Tarantola (1995), Monte carlo sampling of solutions to inverse problems, *J. Geophys. Res.*, *100*(B7), 12,431–12,447.

Oreskes, N., K. Shrader-Frechette, and K. Belitz (1994), Verification, validation, and confirmation of numerical models in the earth sciences, *Science*, *263*, 641–646.

Riley, M. J., and H. G. Stefan (1988), Minlake: A dynamic lake water quality simulation model, *Ecol. Model.*, *43*(3–4), 155–182.

Saloranta, T. M., and T. Andersen (2007), MyLake - a multi-year lake simulation model code suitable for uncertainty and sensitivity analysis simulations, *Ecol. Model.*, *207*, 45–60, 1.

Saloranta, T. M., M. Forsius, M. Jarvinen, and L. Arvola (2009), Impacts of projected climate change on the thermodynamics of a shallow and a deep lake in finland: model simulations and bayesian uncertainty analysis, *Hydrol. Res.*, *40*, 234–248, 2-3.

Schindler, D. W. (1977), Evolution of phosphorus limitation in lakes, *Science*, *195*(4275), 260–262.

Solomon, S., D. Qin, M. Manning, M. Marquis, K. Averyt, M. M. B. Tignor, H. L. Miller, and C. Zhenlin (2007), *Climate Change 2007: The Physical Science Basis*, Intergovernmental Panel on Climate Change.

Stock, P., and R. J. Burton (2011), Defining terms for integrated (multi-inter-trans-disciplinary) sustainability research, *Sustainability*, *3*(8), 1090–1113.

Tarantola, A. (2006), Popper, bayes and the inverse problem, *Nat. Phys.*, *2*(8), 492–494.

Tominaga, K., J. Aherne, S. A. Watmough, M. Alveteg, B. J. Cosby, C. T. Driscoll, M. Posch, and A. Pourmokhtarian (2010), Predicting acidification recovery at the hubbard brook experimental forest, new hampshire: Evaluation of four models, *Environ. Sci. Technol.*, *44*(23), 9003–9009.

Trolle, D., D. P. Hamilton, M. R. Hipsey, K. Bolding, J. Bruggeman, W. M. Mooij, J. H. Janse, A. Nielsen, E. Jeppesen, J. A. Elliott, V. Makler-Pick, T. Petzoldt, K. Rinke, M. R. Flindt, G. B. Arhonditsis, G. Gal, R. Bjerring, K. Tominaga, J. Hoen, A. S. Downing, D. M. Marques, C. R. Fragoso, M. Sondergaard, and P. C. Hanson (2012), A community-based framework for aquatic ecosystem models, *Hydrobiologia*, *683*(1), 25–34.

Vrugt, J. A., C. J. F. t. Braak, M. P. Clark, J. M. Hyman, and B. A. Robinson (2008), Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with markov chain monte carlo simulation, *Water Resour. Res.*, *44*(12), W00B09.

Vrugt, J. A., C. J. F. ter Braak, C. G. H. Diks, B. A. Robinson, J. M. Hyman, and D. Higdon (2009), Accelerating markov chain monte carlo simulation by differential evolution with self-adaptive randomized subspace sampling, *J. Nonlin. Sci. Num.*, *10*(3), 273–290.

Wade, A. J., P. G. Whitehead, and D. Butterfield (2002), The integrated catchments model of phosphorus dynamics (INCA-P), a new approach for multiple source assessment in heterogeneous river systems: model structure and equations, *Hydrol. Earth Syst. Sc.*, *6*, 583–606, 3.

Walz, N., and R. Adrian (2008), Is there a paradigm shift in limnology and marine biology?, *Int. Rev. Hydrobiol.*, *93*, 633–638, 4-5.

Wetzel, R. G. (2001), *Limnology: Lake and River Ecosystems*, 3 ed., Academic Press.

I

Modelling the responses of a cold temperate lake to changes in external forcing

Aleksandra T. Romarheim · Koji Tominaga · Gunnhild Riise · Tom Andersen

A. T. Romarheim · G. Riise

Department of Plant and Environmental Sciences, Norwegian University of Life Sciences, PO Box 5003, 1432 Ås, Norway

e-mail: aleksandra.romarheim@umb.no

K. Tominaga · T. Andersen

Department of Biology, University of Oslo, PO Box 1066, Blindern, 0316 Oslo, Norway

K. Tominaga

Norwegian Institute for Water Research, Gaustadalléen 21, 0349 Oslo, Norway

Keywords MyLake model · Cold temperate lake · Eutrophication · Weather · Loading · Lake Årungen

Abstract

The simultaneous action of multiple environmental pressures limits our understanding of the impact of individual factors on the lake processes. In this model experiment, the MyLake model was used to disentangle the importance of year-to-year variations of meteorological forcing and nutrient loading on the eutrophic and particle loaded Lake Årungen, Norway. Daily meteorological and runoff input data were used to reconstruct the lake response for the period 1994 to 2010, and to simulate the lake responses under different weather and loading scenarios for the same period. Both weather and loading exhibited year-to-year variation, and therefore affected physical, chemical and biological lake responses. Air temperature and precipitation were most variable during the winter. Variation in runoff volume was most prevalent during autumn and winter, while variation in phosphorus inflow was most extensive from late winter to early spring. Thermal related properties in the lake were highly determined by weather conditions, whereas loading was the most important factor for phytoplankton biomass and water transparency. Mild winters and increased input of suspended matter and phosphorus were followed by increased phytoplankton biomass and light attenuation. Thus, in lowland lakes surrounded by erosive soil, mild winters may promote summer blooms with low water transparency. Our study shows that future changes in the global climate may have important implications for local water management decision-making.

Introduction

Nutrient enrichment of lakes may lead to high phytoplankton mass development, low water transparency, and fish mortality due to oxygen depletion (Smith et al., 1999). Mainly two factors affect the nutrient loading to lakes: 1) the soil and land use in the lake catchment, and 2) the hydrology of the watershed. Phosphorus is generally regarded as the determining nutrient for phytoplankton production in freshwater lakes (Schindler, 1977). Much effort has therefore been given to reduce phosphorus input to aquatic ecosystems, which has demonstrably lead to reduced phytoplankton production and increased water transparency in many lakes in Europe and North America (Jeppesen et al., 2005). On the other hand, many lakes have showed delayed or negligible improvements in water quality despite reduced nutrients loading (Jeppesen et al., 2007a).

Year-to-year weather variation has also been recognised to affect physical, chemical and biological processes in lakes (Bailey-Watts & Kirika, 1999; Blenckner et al., 2007; Jeppesen et al., 2007b, 2009; Whitehead et al., 2009). Increased air temperature has been shown to increase the water temperature (George et al., 2007) and the stability of thermal stratification (Straile et al., 2003a), change the phytoplankton community towards dominance of species adapted to warmer water (Weyhenmeyer et al., 2002), and may lead to earlier and higher phytoplankton production (Weyhenmeyer et al., 2002; Huber et al., 2008). The changes in thermal conditions and mixing regime can in turn influence the light, oxygen and nutrient dynamics in the lake, and thereby impact the phytoplankton primary production and community structure (Tirok & Gaedke, 2007; Wilhelm & Adrian, 2008). Precipitation is also deemed as an important factor determining the water transparency, runoff intensity, and suspended matter discharge (Arheimer et al., 2005; Nõges et al., 2007; Ulén et al., 2007).

Norway has, generally, a low fraction of arable land (3%) and low population density (12 persons km^{-2}), so eutrophication is mainly recognizable in intensive agricultural districts at low altitudes. Lake Årungen is situated in a developed agricultural area south-east in Norway, and is one of the most nutrient rich lakes in the country. Geological studies suggest that the natural phosphorus concentration of lakes in this area is 7-8 mg m^{-3} (Borch et al., 2007). Eutrophication became a problem in the lake during the 1960s, with phosphorus concentration exceeding 400 mg m^{-3} in the 1980s (Løvstad & Krogstad, 1993). Algal blooms, low water transparency, malodorous water, reduced fish stocks, and occasional mass mortality of fish were observed in the lake in this period (Ensby et al., 1984). Heavy algal blooms still occur, despite investments in sewage treatment and extensive changes in agricultural practices since the 1970s to reduce nutrient leaching and erosion from the catchment.

Predicting eutrophication responses to nutrient loading is a complex task due to the temporal dynamics of lake's response to weather and runoff. This makes the traditional empirical approach to the problems unreliable. Thus, lake models based on system of processes have been identified as a primary tool for improving our understanding of recovery and progression of eutrophication (Mooij et al., 2010).

The main aim of the current study is to evaluate the relative importance of year-to-year variation of two major factors, namely meteorological forcing and nutrient loading, contributing to lake's physical, chemical and biological conditions. To this aim, 1) the MyLake model (Saloranta & Andersen, 2007) was calibrated against the lake data, 2) various meteorological and nutrient loading scenarios were then applied, and finally 3) the predicted lake responses were compared among different scenarios.

Material and methods

Study site

Lake Årungen is a dimictic lake with maximum and average depths of 13 and 8 m, respectively. The lake is located in south-east Norway, 25 km south of Oslo (59°41'18"N, 10°44'38"E; Fig. 1), and has a surface area of 1.2 km^2. The catchment area covers 51 km^2, where 53% is agricultural land, 34% forestry, 10% densely populated and 3% open water surfaces. The lake is highly exposed to agricultural runoff that causes high nutrient and particle loading. Runoff is mainly through 6 streams of 1.5 to 5 km length. The outlet connects the lake to the marine environment as Lake Årungen enters the Oslofjord through a 3 km long stream.

MyLake model inputs and outputs

MyLake is a one-dimensional lake model, adapted from MINLAKE (Riley & Stefan, 1988), which simulates daily changes in physical and chemical dynamics over the depth gradient (Saloranta & Andersen, 2007). The model simulates ice and snow dynamics in a mechanistic manner and it has been applied to winter-freezing lakes in Norway and Finland (Lydersen et al., 2003; Saloranta, 2006; Kankaala et al., 2006; Saloranta et al., 2009). It was therefore considered as a suitable model for Lake Årungen.

MyLake requires input of meteorological forcing, runoff volume and temperature, and fluxes of suspended inorganic particles and total phosphorus (TP) to model phosphorus and phytoplankton dynamics in the lake (Table 1). Meteorological data for daily air temperature, global radiation, cloud cover, precipitation, relative humidity and wind speed were obtained from the nearby meteorological station located at the Norwegian University of Life Sciences (59°39'37"N, 10°46'54"E). Time series of daily runoff volume, runoff water temperature, and fluxes of suspended inorganic particles and total phosphorus are available for the period 1994 to 2010 from the Skuterud monitoring station (Fig. 1) with a hydrovolumetric weir. The monitoring station is located at an inlet stream to Østensjøvann (59°41′18″N, 10°49′45″E), a small lake of 0.4 km^2 which drains into the Lake Årungen (Deelstra et al., 2007). Runoff from the other sub-catchments was estimated by scaling up the Skuterud data. The up-scaling was based on previously determined flow and nutrient scaling factors that take into account differences in area and land use between sub-catchments (Askilsrud, 2010). A separate MyLake model was set up for Lake Østensjøvann to account for the buffering effects of this lake in the largest sub-catchment of Lake Årungen. The simulated water properties of Lake Østensjøvann were combined with runoff from the other sub-catchments as an estimate of the total runoff to Lake Årungen.

Six variables including whole-lake average TP pool, mean surface chlorophyll concentration, light attenuation coefficient, thermocline depth, epilimnion temperature, and ice thickness were calculated from unprocessed model outputs (Table 1) in order to ease interpretation of the statistical analyses for scenario experiment described below.

Model calibration

Water temperature, TP, soluble reactive phosphorus (SRP), and chlorophyll *a* concentration from the deepest location in the lake were used to calibrate the model (Table 1). Vertical water temperature profile was continuously logged every hour at eight depths between 0.7 and 12.6 m by Hobo pendant temperature loggers (model 64K - UA-002-64; Onset Computer Corporation, Bourne, MA, USA) in the period from November 2008 to August 2010. Water samples for chemical and biological analyses were collected with a modified Rüttner water sampler at seven depths twice a month or monthly (n = 49) from January 2008 to September 2010. TP, SRP and chlorophyll *a* were determined spectrophotometrically (UV-VIS Spectrophotometer UV-1201, Shimadzu, Kyoto, Japan).

We deployed the Markov chain Monte Carlo (MCMC) method (Andrieu et al., 2003) to calibrate the model. The calibration consisted of two stages. The first MCMC calibration stage involved three physical parameters (see Table 2) that only affect heat dynamics, in particular thermocline depth. This first calibration was run against daily temperature measurements, using 2,000 MCMC steps with the first 1,000 for burn-in. The second MCMC calibration stage involving eight parameters (see Table 2) that affect phosphorus and chlorophyll dynamics, but not temperature, was run against measurements of TP, SRP, and chlorophyll *a* in 30,000 MCMC steps with the first 10,000 for burn-in. For these MCMC applications, convergence was monitored visually. Linear interpolation was used to match model outputs on a 0.5 m vertical grid to the actual measurement depths. Although it was not used directly during the MCMC calibration, model goodness of fit was assessed by root mean squared prediction errors (RMSE). The medians of the posterior parameter distributions generated from both stages of the MCMC calibrations were used for the scenarios experiments described in the following.

Model scenarios

The model was run for six nutrient loading and weather scenarios to quantify the impact of weather variation and loading conditions on phosphorus and phytoplankton dynamics (Table 3). These scenarios were based on input combination of observed data (original data, 1994-2010) and synthetic data, where the synthetic data were created by taking the year-to-year mean ($n = 16$) of each of the day of the year. Synthetic data repeats the calculated mean year with 365 days sixteen times. The 29th of February is removed in year-to-year mean calculation, and 28th of February was repeated to account for the 29th in leap years.

Statistical methods

A two-way analysis of variance (ANOVA) was run on the 16 years of water year based simulation statistics (water year mean, see Table 1), among scenarios A, B, C and D (two weather factors by two loading factors, see Table 3). All annual averages are computed over the period from 1st October to 30th September, commonly used in Europe to refer on a hydrological year, or a water year (Otnes & Ræstad, 1978). Since treatment contrasts are nested within water years we factored out the between-year variances to gain a greater power in the statistical tests. Principal component analysis (PCA) was used to explore the relationships between meteorological and land-related forcing and their relevance for the simulated lake response. Four water years with extreme PCA scores were selected for studying contrasting lake responses in closer details.

Results

Calibration

The simulated water temperature and thermal dynamics of the lake were in agreement with the lake observed data. Observed water temperature measurements were well predicted by simulation and the RMSE was less than 2°C at all lake depths (Fig. 2). After the water temperature calibration, parameters controlling TP, SRP, and chlorophyll *a* were calibrated against observed data for the period from January 2008 to September 2010. The epilimnion TP, SRP, and chlorophyll *a* were well predicted by the model, although their prediction was less successful than the prediction of the water temperature. The TP and SRP were better predicted by the model in shallower water than in deeper water whereas the chlorophyll *a* showed the opposite pattern. In general, the model simulated well TP and SRP, although both phosphorus forms were overestimated in early spring and autumn at shallow depths, while underestimated in bottom water. Simulated SRP concentrations were also somewhat higher than observed in winters. However, the simulation succeeded in showing a decreasing trend of lake phosphorus in spring and midsummer, and in mimicking its increase during the autumn mixing of water. Although the simulated chlorophyll concentrations were lower than measured, the model was able to predict seasonal variation in phytoplankton primary production and to simulate high phytoplankton biomasses in the lake epilimnion during midsummer.

Input variability

Inter-annual variation was expressed as the standard deviation in inputs and outputs between the years. All weather inputs varied between years (Fig. 3), with air temperature and global radiation having the strongest seasonal pattern in inter-annual variation (i.e. greater 16-year variation as compared to year-to-year variation on a day-of-year basis). The inter-annual variation in air temperature was strongest in the winter period, whereas global radiation varied most during the summer months. The variation in cloud cover, precipitation and relative humidity was generally similar across seasons. The year-to-year variation for precipitation was particularly high in December and in the period from July to September, the latter reflecting extreme precipitation events. Wind speed varied most in winters.

Runoff input data on water flow and concentrations of TP and suspended matter all varied seasonally and between years. The variation in runoff volume was greatest in the period from October to May. No clear seasonal pattern in the degree of variability could be found for suspended matter and TP fluxes, although the variation of TP influx seemed to peak in February-March.

Output variability

Differences in year-to-year variation among the scenarios (Fig. 4) and the annual statistics (Table 4) illustrate the seasonal influence of the external forcing on the thermal regime and the phosphorus and phytoplankton dynamics in the lake. The lake responded differently between years; all simulated outputs, except ice thickness, showed large variation in the beginning and at the end of the phytoplankton growing season (Fig. 4). All simulated output variables were influenced by external forcing to some extent as they varied inter-annually for all model scenarios. Ice thickness was significantly affected by weather ($P < 0.001$) as both air temperature and winter precipitation highly contributed to its variation between years (Table 4; Fig. 4). The variation in thermocline depth in May and October was well revealed by the model, and seemed to be equally dependent on weather and loading. The epilimnion temperature during the whole growing season was largely controlled by weather. The TP content in the lake was most variable in the period from November to January, and in April and July. Loading could mostly explain the inter-annual TP variation in the lake, whereas precipitation contributed to TP variation only in the spring, and air temperature only during the winter period. Loading was the overall most important factor in controlling the light attenuation coefficient ($P < 0.001$) and surface chlorophyll concentration ($P < 0.001$). Weather seemed to be important in controlling chlorophyll and light attenuation in early spring whereas loading was the most important factor controlling the both variables from June to September.

The years 1996, 2000, 2006 and 2007 were the four most extreme years determined on the basis of the PCA analysis (Fig. 5). The year 1996 was characterized by relatively low average annual air temperature, a thin cloud cover and low precipitation, which resulted in low epilimnion water temperature, short lasting thermocline, low runoff volume, and TP in the lake. The year 2007 represents an opposite to 1996, regarding weather characteristics and resulting model simulation with relatively high average annual air temperature and precipitation. Increased wind speed and low winter air temperatures and precipitation coincided with increased ice thickness and global radiation such as in 2003-2006, 2009 and 2010. These weather conditions resulted in lower suspended inorganic particles and TP in runoff which coincided with lower surface chlorophyll concentration and light attenuation. The year 2006 was identified as the extreme during this period, with a cold winter followed by a warm summer. In contrast, the year 2000 was characterized as a year with less global radiation, lower summer air temperature, and higher wind speed, but with higher winter temperature and precipitation. Such weather conditions pronounced higher TP and suspended particle in runoff compared with an average year, resulting in a high surface chlorophyll concentration and lower water transparency.

Discussion

Year-to-year variability in lake responses

Year-to-year weather variations, as well as the influence of catchment land-use and hydrology hinder our understanding of how individual stressors may affect the lake response (Blenckner, 2005). Our model experiment, which involved the input of weather and loading data for 16 years, was able to outline the importance of year-to-year variation in external forcing on physical, chemical and biological response in Lake Årungen. The combination of high forcing variability and high lake response sensitivity made the inter-annual variation most dramatically expressed in spring and autumn.

Air temperature, precipitation, and wind speed are the principal factors influencing freshwater ecosystems in a changing climate (Nickus et al., 2010). The lake thermal regime was to a large extent affected by weather conditions, particularly by air temperature. Variable winter air temperatures were an important factor influencing the heating and mixing processes during spring. A dynamic physical environment at the beginning of the growing season has considerable influence on the phytoplankton community structure and its dynamics (Weyhenmeyer et al., 2002). Increased surface water temperatures in the English Lake District (George et al., 2007) and incomplete water mixing in Lake Constance (Straile et al., 2003a) have earlier been associated with mild winters. High inter-annual variation in winter air temperatures in Lake Årungen was reflected in the simulated ice thickness and phenology of ice formation, with ice forming in December-January and disappearing in March-April. Also in other lake studies, decreased ice thickness and reduced duration of ice cover have been related to mild winters (Nickus et al., 2010). In this study, the time of thermocline formation varied among years by more than one month, from mid-April to mid-May. The large year-to-year variation in thermocline depth and duration could lead to changes in temperature, light, and nutrient condition in the lake, which further shape the phytoplankton community and determine its total biomass (Padisak et al., 2010; Zohary et al., 2010). For instance, early disappearance of diatoms and high development of cyanobacteria in European lakes has been related to winter warming and increased water temperature (Weyhenmeyer et al., 2002). Furthermore, increased water stability also favors the buoyant phytoplankton species such as bloom forming cyanobacteria (Reynolds et al., 1983; Winder & Hunter, 2008).

Year-to-year variation in phosphorus content in the lake was highly influenced by loading which indicated that the external nutrient supply remains an important source of phosphorus in the lake. Although the changes in nutrient loading are primarily linked to anthropogenic activities in the catchment, particularly to practices in agriculture, the short-term variations in weather and runoff will also influence the nutrient supply from external sources. Lake Årungen is surrounded by agricultural land, and is especially sensitive to variable weather conditions that promote nutrient loading from the catchment. Air temperature and rainfall frequency and intensity, all affect the runoff and the soil erosion pattern, particularly during the winter period. Increased winter temperatures with frequent freezing and melting events increase the risk of erosion, which again will increase the nutrient loading to the lake (Bechmann et al., 2005; Nõges et al., 2007; Jeppesen et al., 2009). Although not statistically demonstrated in the present study, the indirect impacts of weather conditions on discharge may still be important in regulating the nutrient dynamics. Variable winter weather conditions, and the time of ice out were most important cause of year-to-year variable phosphorus content in the period from November to January and in April in Lake Årungen. Enhanced phosphorus concentrations in streams during winter, and high phosphorus loading in early spring, both contribute to the total lake phosphorus concentration during the following summer in two Norwegian lakes with agricultural catchments (Bechmann et al., 2005). In addition, the variation in summer TP content could also be the result of between-year variation in rainfall, especially due to extreme precipitation events as observed for particular days in July. High inter-annual variation in TP content can consequently result in variable phytoplankton biomass between years.

Phytoplankton biomass and light were significantly affected by loading, although atmospheric forcing also contributed to their inter-annual variation in the lake. The effect of loading was pronounced during the whole algal growing season, whereas weather had the strongest effect in early spring and from mid-August to the end of the growing season. Thermal stratification is highly dependent on weather and may further influence water mixing as well as light and nutrient regimes, which are important in controlling the phytoplankton dynamics (Padisak et al., 2010; Zohary et al., 2010). Similar pattern of year-to-year variation in water temperature, chlorophyll concentration, and light attenuation indicates a close relationship between these variables. Increased air temperature promoted higher water temperatures and higher stability of the thermal stratification which enhanced phytoplankton production, particularly of bloom forming cyanobacteria (Reynolds et al., 1983; Weyhenmeyer et al., 2002). Increased runoff and soil erosion, caused by intense precipitation and frequent melting of snow and ice during mild winters, affect eutrophication and water turbidity (Bechmann et al., 2005; Jeppesen et al., 2009). Light, therefore, may limit phytoplankton growth more than nutrients in highly turbid lakes such as Lake Årungen (Dokulili, 1994). Reduced light availability may be crucial for the competitive success of cyanobacteria which are functionally adapted to low light conditions (Litchman, 1998). Particularly high dominance of cyanobacteria has been observed in the Lake Årungen after mild winters followed by low light conditions in spring (Romarheim et al., unpublished). Therefore, the additional measures to control soil erosion should be considered in water management, not only to reduce the supply of nutrients, but also to avoid low water transparency which may favour development of potentially toxic cyanobacteria.

Implications for lake management

According to our PCA analysis, most of the 1990s were categorized by winters with higher temperatures and more rainfall. The mild winters were related to a positive North Atlantic Oscillation (NAO) phase which has been shown to strongly influence physicochemical and biological responses in western European lakes (Weyhenmeyer et al., 2002; Straile et al., 2003b; George et al., 2007). The effect of climate condition on water ecosystems, however, should be considered individually as the lake response is also determined by the lake's geographical position, landscape topography, and the lake's morphometry and mixing regime (Nickus et al., 2010). Our model experiment indicated that increased inflow of suspended matter and phosphorus to the lake Årungen may be expected after mild winters with high precipitation. Consequently, high chlorophyll concentrations and increased light attenuation were predicted after mild and wet winters such in the year of 2000. Mild winters may thus counteract measures aimed to reduce external nutrient supply and to control phytoplankton production in temperate lakes. On the contrary, cold winters were related to thicker ice layer, less inflow of suspended matter and phosphorus, and low chlorophyll and light attenuation. This was consistent with the observed increase in water transparency and reduction of phytoplankton biomass, particularly of cyanobacteria, in Lake Årungen after the cold winter in 2010 (Romarheim et al., unpublished). Special attention must therefore be given to management practices, which should minimize the use of fertilizers and reduce the risk of nutrient runoff and soil erosion, especially in areas that drains directly into the lake. Increased annual air temperature coincided with warmer epilimnion, shallower thermocline and extended summer stratification such as for the year of 2007. In addition, high annual precipitation and runoff volume, particularly in summer, coincided with warmer years. The latest Intergovernmental Panel on Climate Change Assessment reported that all years in the period from 1995 to 2006, with exception of 1996, were globally among the warmest since 1850 (Trenberth et al., 2007). Similarly, the year 1996 was characterised with low average annual air temperature in our model experiment. Lower annual air temperature and low rainfall in 1996 lead to low epilimnion temperature, and a deep and short lasting thermocline. According to best available future climate predictions for Scandinavia, warmer winters and increased winter precipitation are expected in south-eastern Norway (Hanssen-Bauer et al., 2005). If so, we should also expect more intensive soil erosion, higher phosphorus loading, reduced water transparency, and increased phytoplankton biomasses, primarily of cyanobacteira in the lakes. Global climate changes and inter-annual variations in the local weather both directly, and indirectly through an impact on the catchment, influence the physicochemical and biological processes in lakes. Therefore, the effect of climate should be considered in future decision-making concerning water management.

Acknowledgements

We thank the Norwegian Institute for Agricultural and Environmental Research (Bioforsk) for providing runoff data from Skuterud monitoring station. We also thank Johnny Kristiansen at the Department of Plant and Environmental Sciences, Norwegian University of Life Sciences for excellent assistance during the field and laboratory work.

References

Andrieu, C., N. de Freitas, A. Doucet & M. I. Jordan, 2003. An introduction to MCMC for machine learning. Machine Learning 50: 5−43.

Arheimer, B., J. Andréasson, S. Fogelberg, H. Johnsson, C. B. Pers & K. Persson, 2005. Climate change impact on water quality: Model results from southern Sweden. Ambio 34: 559−566.

Askilsrud, H., 2010. Nutrient input to Lake Årungen - assessment of data availability to run the MyLake modell. M. Sc. thesis, Norwegian University of Life Sciences, Ås, Norway.

Bailey-Watts, A. E. & A. Kirika, 1999. Poor water quality in Loch Leven (Scotland) in 1995 in spite of reduced phosphorus loadings since 1985: the influences of catchment management and inter-annual weather variation. Hydrobiologia 403:135−151.

Bechmann, M. E., D. Berge, H. O. Eggestad & S. M. Vandsemb, 2005. Phosphorus transfer from agricultural areas and its impact on the eutrophication of lakes - two long-term integrated studies from Norway. Journal of Hydrology 304: 238−250.

Blenckner, T., 2005. A conceptual model of climate-related effects on lake ecosystems. Hydrobiologia 533: 1−14.

Blenckner, T., R. Adrian, D. Livingstone, E. Jennings, G. A. Weyhenmeyer, D. G. George, T. Jankowski, M. Jarviken, C. N. Aonghusa, T. Nõges, D. Straile & K. Teubner, 2007. Large-scale climatic signatures in lakes across Europe: a meta-analysis. Global Change Biology 13: 1314−1326.

Borch, H., A. Yri, Ø, Løvstad & S. Turtumøygard, 2007. Tiltaksplan for Årungen. Bioforsk rapport 52 (in Norwegian): 54 pp.

Deelstra, J., G. H. Ludvigsen, A. Pengerud, H. O. Eggestad, G. Tveiti & L. Øygarden, 2007. Jord- og vannovervåking i landbruket (JOVA). Skuterudbekken 2006. Bioforsk rapport 118 (in Norwegian): 25 pp.

Dokulil, M. T., 1994. Environmental control of phytoplankton production in turbulent turbid systems. Hydrobiologia 289: 65–72.

Ensby, S., R. Borgstrøm, G. Langeland, F. Rosland & S. Sanni, 1984. Årungen: tilstand, aktuelle sanerings- og restaureringstiltak: rapport utarbeidet på grunnlag av tverrfaglig forskningsaktivitet i perioden 1980-83 (in Norwegian). Institutt for Georessurs- og Forurensningsforskning, Ås, Norway: 30 pp

George, G., M. Hurley & D. Hewitt, 2007. The impact of climate change on the physical characteristics of the larger lakes in the English Lake District. Freshwater Biology 52: 1647−1666.

Hanssen-Bauer, I., C. Achberger, R. E. Benestad, D. Chen & E. J. Førland, 2005. Statistical downscaling of climate scenarios over Scandinavia. Climate Research 29: 255−268.

Huber, V., R. Adrian & D. Gerten, 2008. Phytoplankton response to climate warming modified by trophic state. Limnology and Oceanography 53: 1−13.

Jeppesen, E., M. Sondergaard, J. P. Jensen, K. E. Havens, O. Anneville, L. Carvalho, M. F. Coveney, R. Deneke, M. T. Dokulil, B. Foy, D. Gerdeaux, S. E. Hampton, S. Hilt, K. Kangur, J. Köhler, E. H. H. R. Lammens, T. L. Lauridsen, M. Manca, M. R. Miracle, B. Moss, P. Nõges, G. Persson, G. Phillips, R. Portielje, S. Romo, C. L. Schelske, D. Straile, I. Tatrai, E. Willén & M. Winder, 2005. Lake responses to reduced nutrient loading - an analysis of contemporary long-term data from 35 case studies. Freshwater Biology 50: 1747-1771.

Jeppesen, E., M. Meerhoff, B. A. Jacobsen, R. S. Hansen, M. Søndergaard, J. P. Jensen, T. L. Lauridsen, N. Mazzeo & C. W. C. Branco, 2007a: Restoration of shallow lakes by nutrient control and biomanipulation - the successful strategy varies with lake size and climate. Hydrobiologia 581: 269−285.

Jeppesen, E., M. Sondergaard, M. Meerhoff, T. L. Lauridsen & J. P. Jensen, 2007b. Shallow lake restoration by nutrient loading reduction - some recent findings and challenges ahead. Hydrobiologia 584: 239–252.

Jeppesen, E., B. Kronvang, M. Meerhoff, M. Sondergaard, K. M. Hansen, H. E. Andersen, T. L. Lauridsen, L. Liboriussen, M. Beklioglu, A. Ozen & J. E. Olesen, 2009. Climate change effects on runoff, catchment phosphorus loading and lake ecological state, and potential adaptations. Journal of Environmental Quality 38: 1930−1941.

Kankaala, P., J. Huotari, E. Peltomaa, T. Saloranta & A. Ojala, 2006. Methanotrophic activity in relation to methane efflux and total heterotrophic bacterial production in a stratified, humic, boreal lake. Limnology and Oceanography 51:1195–1204.

Litchman, E., 1998. Population and community responses of phytoplankton to fluctuating light. Oecologia 117: 247−257.

Lydersen E., K. J. Aanes, S. Andersen, T. Andersen, P. Brettum, T. Bækken, L. Lien, E. A. Lindstrøm, J. E. Løvik, M. Mjelde, T. J. Oredalen, A. S. Lyche, R. Ropmstad, B. Rørslett & T. M. Saloranta, 2003. THERMOS-projektet: Fagrapport 1998-2002. NIVA rapport 4720 (in Norwegian): 119 pp.

Løvstad, Ø. & T. Krogstad, 1993. Årungen 1992. Eutrofiering, plantenæringsstoffer og blågrønnalger. Institutt for Jordfag, Norges Landbrukshøgskole, Ås, Norway: 17 pp.

Mooij, W. M., D. Trolle, E. Jeppesen, G. Arhonditsis, P. V. Belolipetsky, D. B. R. Chitamwebwa, A. G. Degermendzhy, D. L. DeAngelis, L. N. D. Domis, A. S. Downing,

J. A. Elliott, C. R. Fragoso, U. Gaedke, S. N. Genova, R. D. Gulati, L. Hakanson, D. P. Hamilton, M. R. Hipsey, J. 't Hoen, S. Huelsmann, F. H. Los, V. Makler-Pick, T. Petzoldt, I. G. Prokopkin, K. Rinke, S. A. Schep, K. Tominaga, A. A. Van Dam, E. H. Van Nes, S. A. Wells & J. H. Janse, 2010. Challenges and opportunities for integrating lake ecosystem modelling approaches. Aquatic ecology 44: 633−667.

Nickus, U., K. Bishop, M. Erlandsson, C. D. Evans, M. Forsius, H. Laudon, D. M. Livingstone, D. Monteith & H. Thies, 2010. Direct impacts of climate change on freshwater ecosystems. In Kernan, M., R. W. Battarbee & B. Moss (eds), Climate change impacts on freshwater ecosystems. Wiley-Blackwell, West Sussex: 38−64.

Nõges, P., M. Kägu & T. Nõges, 2007. Role of climate and agricultural practice in determining matter discharge into large, shallow Lake Võrtsjärv, Estonia. Hydrobiologia 581: 125−134.

Otnes, J. & E. Ræstad, 1978. Hydrologi i praksis. Ingeniørforlaget, Oslo.

Padisak, J., E. Hajnal, L. Naselli-Flores, M. T. Dokulil, P. Nõges & T. Zohary, 2010. Convergence and divergence in organization of phytoplankton communities under various regimes of physical and biological control. Hydrobiologia 639: 205−220.

Reynolds, C. S., S. W. Wiseman, B. M. Godfrey & C. Butterwick, 1983. Some effects of artificial mixing on the dynamics of phytoplankton populations in large limnetic enclosures. Journal of Plankton Research 5: 203−234.

Riley, M. J. & H. G. Stefan, 1988. MINLAKE: a dynamic lake water quality simulation model. Ecological Modelling 43:155–182.

Saloranta T. M., 2006. Highlighting the model code selection and application process in policy-relevant water quality modelling. Ecological Modelling 194: 316–327.

Saloranta T. M. & T. Andersen, 2007. MyLake - A multi-year lake simulation model code suitable for uncertainty and sensitivity analysis simulations. Ecological Modelling 207: 45–60.

Saloranta T. M., M. Forsius, M. Järvinen & L. Arvola, 2009. Impacts of projected climate change on the thermody- namics of a shallow and deep lake in Finland: model simulations and Bayesian uncertainty analysis. Hydrological Research 40: 234–248.

Schindler, D. W., 1977. Evolution of phosphorus limitation in lakes. Science 195: 260−262.

Smith, V. H., G. D. Tilman & J.C. Nekola, 1999. Eutrophication: impacts of excess nutrient inputs on freshwater, marine, and terrestrial ecosystems. Environmental Pollution 100: 179−196.

Straile, D., K. Johnk & H. Rossknecht, 2003a. Complex effects of winter warming on the physicochemical characteristics of a deep lake. Limnology and Oceanography 48, 1432-1438.

Straile, D., D. M. Livingstone, G. A. Weyhenmeyer & D. G. George, 2003b. The response of freshwater ecosystems to climate variability associated with the North Atlantic Oscillation. In Hurrell, J. W., Y. Kushnir, G. Ottersen & M. Visbeck (eds), The North Atlantic Oscillation: Climate significance and environmental impact. American Geophysical Union, Washington DC: 263-279.

Tirok, K. & U. Gaedke, 2007. The effect of irradiance, vertical mixing and temperature on spring phytoplankton dynamics under climate change: long-term observations and model analysis. Oecologia 150: 625−642.

Trenberth, K. E., P. D. Jones, P. Ambenje, R. Bojariu, D. Easterling, A. K. Tank, D. Parker, F. Rahimzadeh, J. A. Renwick, M. Rusticucci, B. Soden & P. Zhai, 2007. Observations: Surface and Atmospheric Climate Change. In Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, M. Tignor & H. L. Miller (eds.), Climate Change 2007: The

Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge: 235−336.

Ulén, B., M. Bechmann, J. Fölster, H. P. Jarvie & H. Tunney, 2007. Agriculture as a phosphorus source for eutrophication in the north-west European countries, Norway, Sweden, United Kingdom and Ireland: a review. Soil Use and Management 23: 5−15.

Weyhenmeyer, G. A., R. Adrian, U. Gaedke, D. M. Livingstone & S. C. Maberly, 2002. Response of phytoplankton in European lakes to a change in the North Atlantic Oscillation. Verhandlungen des Internationalen Verein Limnologie 28: 1436−1439.

Whitehead, P., R. L. Wilby, R. W. Battarbee, M. Kernan & A. J. Wade, 2009. A review of the potential impacts of climate change on surface water quality. Hydrological Sciences Journal 54: 101−123.

Winder, M. & D. A. Hunter, 2008. Temporal organization of phytoplankton communities linked to physical forcing. Oecologia 156: 179–192.

Wilhelm, S. & R. Adrian, 2008. Impact of summer warming on the thermal characteristics of a polymictic lake: Consequences for oxygen, nutrients and phytoplankton. Freshwater Biology 53: 226-237.

Zohary, T., J. Padisak & L. Naselli-Flores, 2010. Phytoplankton in the physical environment: beyond nutrients, at the end, there is some light. Hydrobiologia 639: 261−269.

Figure captions

Fig. 1 Map of catchment draining into (1) Lake Årungen (59°41'18"N, 10°44'38"E), with the (2) weather station at Ås, (3) Lake Østensjøvannet and (4) the Skuterud monitoring station. Runoff data from the Skuterud sub-catchment (indicated by dark shading) are scaled up according to land area and usage of the rest of the catchment to estimate the total loading to Lake Årungen

Fig. 2 Simulated (line) and observed (circles) lake state variables for water temperature, TP, SRP, and chlorophyll *a* concentrations at seven depths

Fig. 3 Input variability shown as standard deviations on a water year scale (day-by-day, year-to-year variation, n =16, curves), with the overall 16-year standard deviations indicated by horizontal lines

Fig. 4 Output variability shown as standard deviations on a water year scale (day-by-day, year-to-year variation, n =16) for scenarios A (top of the forward diagonal band), B (bottom of the forward diagonal band), C (top of the backward diagonal band), and D (bottom of the backward diagonal band). See Table 3 for scenario configurations

Fig. 5 Principal component analysis (PCA) loadings for the two greatest components, together with the scores for the two components for 16 water years (letters). The arrows are scaled and therefore only meaningful for comparison within PCA axes

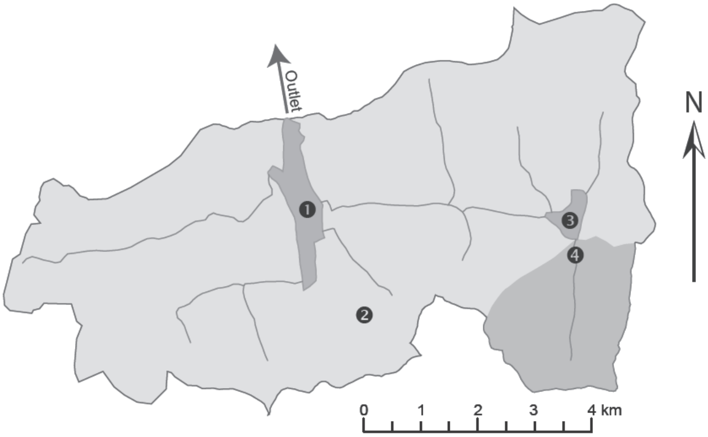
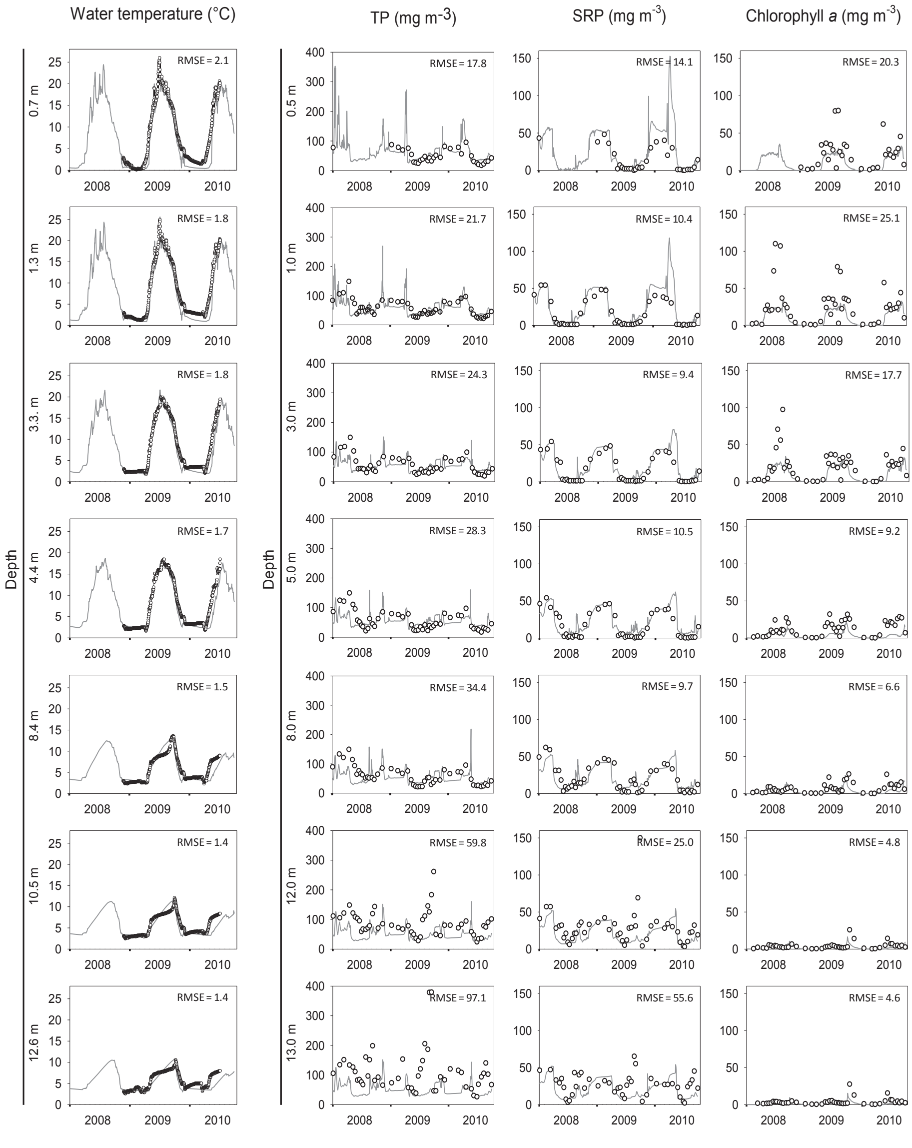


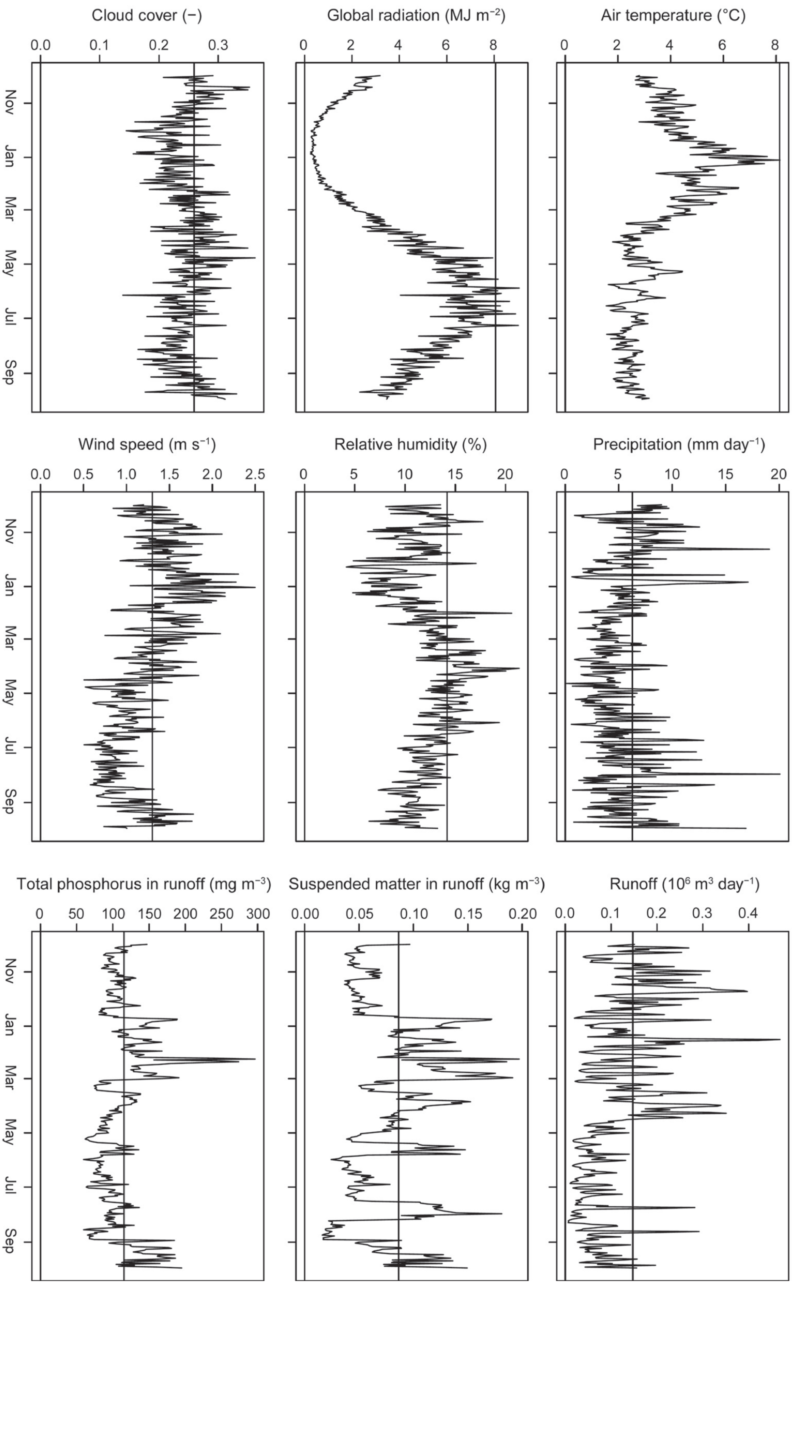
Fig. 1

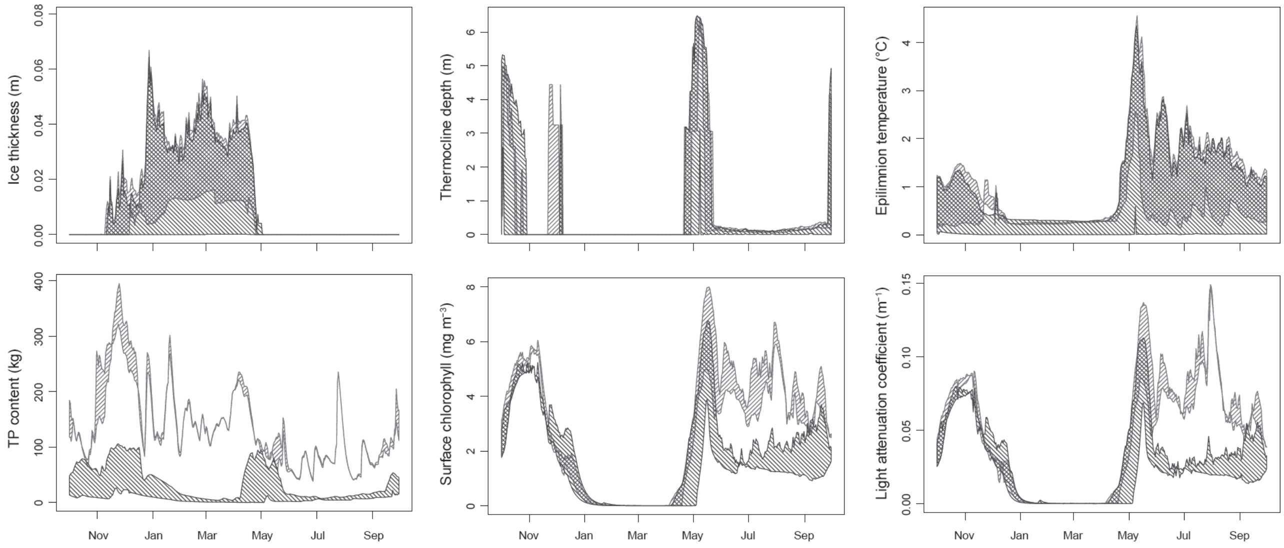


Water temperature (°C)
TP (mg m-3)
SRP (mg m-3)
Chlorophyll a (mg m-3)
Depth
0.7 m
1.3 m
3.3. m
4.4 m
8.4 m
10.5 m
12.6 m
0.5 m
1.0 m
3.0 m
5.0 m
8.0 m
12.0 m
13.0 m
RMSE = 2.1
RMSE = 1.8
RMSE = 1.8
RMSE = 1.7
RMSE = 1.5
RMSE = 1.4
RMSE = 1.4
RMSE = 17.8
RMSE = 21.7
RMSE = 24.3
RMSE = 28.3
RMSE = 34.4
RMSE = 59.8
RMSE = 97.1
RMSE = 14.1
RMSE = 10.4
RMSE = 9.4
RMSE = 10.5
RMSE = 9.7
RMSE = 25.0
RMSE = 55.6
RMSE = 20.3
RMSE = 25.1
RMSE = 17.7
RMSE = 9.2
RMSE = 6.6
RMSE = 4.8
RMSE = 4.6
2008
2009
2010

Fig.2

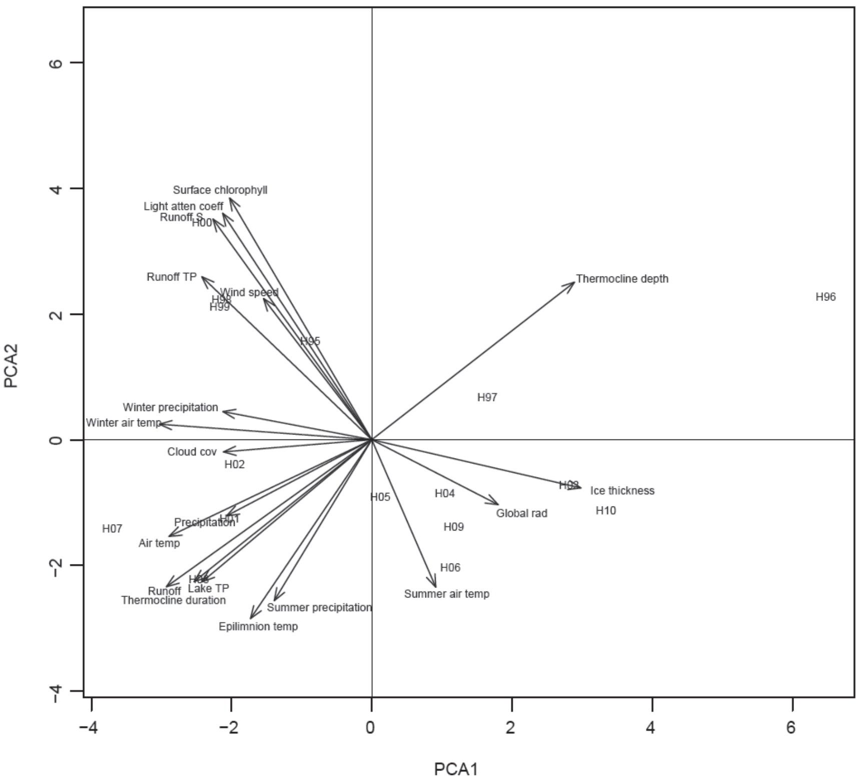
Fig. 3


Cloud cover (−)
Global radiation (MJ m−2)
Air temperature (°C)
Wind speed (m s−1)
Relative humidity (%)
Precipitation (mm day−1)
Total phosphorus in runoff (mg m−3)
Suspended matter in runoff (kg m−3)
Runoff (106 m3 day−1)
Nov
Jan
Mar
May
Jul
Sep



Ice thickness (m)
0.00
0.02
0.04
0.06
0.08
Thermocline depth (m)
0
1
2
3
4
5
6
Epilimnion temperature (°C)
0
1
2
3
4
TP content (kg)
0
100
200
300
400
Surface chlorophyll (mg m⁻³)
0
2
4
6
8
Light attenuation coefficient (m⁻¹)
0.00
0.05
0.10
0.15
Nov
Jan
Mar
May
Jul
Sep

Fig. 4



Surface chlorophyll
Light atten coeff
Runoff S
H00
Runoff TP
Wind speed
H98
H99
H95
Thermocline depth
H96
H97
Winter precipitation
Winter air temp
Cloud cov
H02
H03
Ice thickness
H05
H04
Global rad
H10
H07
Precipitation
H01
H09
Air temp
H06
H08
Runoff
Lake TP
Thermocline duration
Summer precipitation
Summer air temp
Epilimnion temp
PCA1
PCA2

Fig. 5

Table 1 Input and output data, and observed lake data for the calibration of the MyLake model and statistics for the ANOVA and PCA analyses

MyLake inputs	MyLake outputs (selected)	Observed lake data
Meteorological[a, f]	*Calibration purpose*[a]	*Calibration purpose*
Air temperature	(every 0.5 m by depth)	(at 7 depths)
Global radiation	Water temperature	Water temperature[a]
Cloud cover	TP concentration	TP concentration[b]
Precipitation	SRP concentration	SRP concentration[b]
Relative humidity	Chlorophyll concentration	Chlorophyll *a* concentration[b]
Wind speed		
	Statistics calculated for PCA	
Runoff [a, g]	(volume weighted mean 0-3.0 m)	
Flow volume	TP content	
Water temperature	Mean surface chlorophyll	
Suspended matter flux	Light attenuation coeficient	
TP flux	Thermocline depth	
	Mean epilimnion temperature[e]	
Statistics calculated for PCA	Ice thickness	
Global radiation[c]		
Cloud cover[c]		
Air temperature[c]		
Wind speed[c]		
Precipitation[c]		
Flow volume[c]		
Winter air temperature[d,h]		
Summer air temperature[d,i]		
Winter precipitation[d,h]		
Summer precipitation[d,i]		
Suspended matter flux[d]		
TP flux[d]		

[a]Daily data
[b]Biweekly data
[c]Annual mean
[d]Water year basis (October through September)
[e]Volume weighted above thermocline depth
[f]Inferred with data from Ås meteorological station (http://www.umb.no/fagklim/artikkel/meteorologiske-data-for-as)
[g]Inferred with data from Skuterud monitoring station and land use
[h]December through March, mean
[i]June through September, mean

Table 2 Parameters involved in calibration based on two-stage Markov chain Monte Carlo (MCMC) application (first stage for three parameters using 2,000 MCMC steps with 1,000 steps for burn-in and second stage for eight parameters using 30,000 MCMC steps with 10,000 for burn-in). MyLake equation numbers refer to the original model description (Saloranta & Andersen, 2007). Median values were chosen among the posterior parameter distribution

	Value	MyLake equation	Unit
Physical parameters			
Open-water vertical diffusion coefficient	5.00E-01	eq. 10	m^2 day^{-1}
Wind sheltering coefficient	7.96E-02	eq. 13	—
Minimum possible stability frequency	9.31E-05	eq. 10	sec^{-2}
Biological and chemical parameters			
PAR saturation level for photosynthesis	2.04E-04	eq. 29	mol quanta m^{-2} sec^{-1}
Particle resuspension mass transfer coefficient	2.94E-05	§ 2.7	m day^{-1}, dry sediment
Settling velocity for suspended matter	1.38E+00	eq. 20	m day^{-1}
Settling velocity for chlorophyll	7.31E-02	eq. 20	m day^{-1}
Specific mortality rate of phytoplankton	1.86E-01	eq. 26	day^{-1}
Max specific growth rate of phytoplankton	1.76E+00	eq. 27	day^{-1}
Half saturation inorganic phosphorus concentration for Langmuir isotherm	9.99E+02	eq. 24	mg m^{-3}
Saturation level for inorganic phosphorus isotherm	4.96E+04	eq. 24	mg kg^{-1}

Table 3 Model scenarios. The scenarios comprise either original input data (denoted O), pseudo repeated average year based on 16-years of input data (denoted R), or a combination of O and R

	Model scenarios					
Model inputs	A	B	C	D	Dt	Dp
Weather						
Global radiation	**O**	R	**O**	R	R	R
Cloud cover	**O**	R	**O**	R	R	R
Relative humidity	**O**	R	**O**	R	R	R
Wind speed	**O**	R	**O**	R	R	R
Air pressure	**O**	R	**O**	R	R	R
Air temperature	**O**	R	**O**	R	**O**	R
Precipitation	**O**	R	**O**	R	R	**O**
Runoff						
Flow volume	**O**	**O**	R	R	R	R
Suspended matter flux	**O**	**O**	R	R	R	R
Inflow water temperature	**O**	**O**	R	R	R	R
TP flux	**O**	**O**	R	R	R	R

Table 4 Summary results for six two-factor within-subject ANOVA (n = 16 × 2 × 2). Significance of additive and interactive effects of weather (two levels, original O or repeated average R) and loading (two levels, original O or repeated average R) inputs on the six selected model outputs. High *P*-values for interactive effects for all six tests indicate pure additive two-factor model and test for each factor separately. *P*-values below 0.05 are highlighted (shown bold) and indicate significant differences between corresponding sample means

	***P*-value**		
Model outputs	**W**	**L**	**W × L**
Ice thickness, m	**< 0.001**	0.839	0.560
Thermocline depth, m	0.281	0.218	0.398
Epilimnion temperature, °C	**0.014**	0.135	0.771
TP content, kg	0.365	0.088	0.726
Surface chlorophyll, mg m^{-3}	0.699	**< 0.001**	0.791
Light attenuation coefficient, m^{-1}	0.360	**< 0.001**	0.836

II

Melting and warming of northern lakes

Running head: melting and warming of northern lakes

Koji Tominaga[*1,2], Anders Gravbrøt Finstad[3], Øyvind Kaste[4] and Tom Andersen[2]

[1]Norwegian Institute for Water Research, Gaustadalléen 21, 0349 Oslo, Norway

[2]Department of Biosciences, University of Oslo, P.O. Box 1066 Blindern, 0316 Oslo, Norway[3]Norwegian Institute for Nature Research, P.O. Box 5685 Sluppen, 7485 Trondheim, Norway

[4]Norwegian Institute for Water Research, Southern Branch, Jon Lilletuns vei 3, 4879 Grimstad, Norway

*Corresponding author: e-mail: koji.tominaga@ibv.uio.no, phone: +47 22 85 56 00; fax +47 22 85 47 26;

Key-words: lake model, climate change, impacts, thermal response, ice cover, Fennoscandia

Type of paper: Primary research article

Abstract

Northern lakes are expected to respond strongly to changing climate, with severe consequences both for lakes as habitats and as providers of ecosystem services. Inferences about thermal responses of lakes to climate change based on extrapolation from empirical statistical relationships can obscure important interactions between atmospheric forcing and water body. Here, we present projections for thermal response of generic lake types with representative physical features based on a process-based lake thermodynamic model. As model input, we used regional synthetic weather for the reference period (1970 – 2000) and scenario period (2070-2100) based on a regional climate model under the IPCC A1B emission scenario. The simulations were applied to a grid covering major latitudinal, altitudinal, and oceanic climate gradients of the northern hemisphere, using Fennoscandia (Finland, Norway, and Sweden) as an example study region. Our simulations indicate that during the 21st century ice cover duration will shorten on average by 53 days, while summer mean temperature will increase by 2.4-2.5 ˚C in the epilimnion and 2.0 to 2.3 ˚C in the hypolimnion. These changes will be pan-regional and substantially faster than the already extensive change observed in the 20th century, with threshold impacts such as ice disappearance affecting many parts of the region. Our simulations also indicate that the magnitude of projected climate change impact is highly dependent on the physical feature of the lake, with depth and surface area being more important than water transparency.

Introduction

Sensitivity of freshwater lakes to changing climate has been repeatedly demonstrated by significant trends in historical time series of seasonal phenology, water temperature, stratification patterns, and ice cover (e.g. Magnuson et al. 2000; Adrian et al. 2009; Weyhenmeyer *et al.* 2011). Such changes in thermal conditions are likely to have severe consequences for lakes as habitats and as providers of a vast array of ecosystem service, including fisheries, drinking water, local and global climate control, and recreational uses (e.g. Shuter *et al.* 2012; Helland *et al.* 2011; Schröter et al. 2005). Thermal changes in lakes can also affect primary production and other light-dependent ecological processes (O'Brien 1979; Finstad *et al*. 2004; Smol *et al*. 2005). These conditions signify that assessment of future climate change impact on lakes is becoming an increasingly relevant and urgent task under projected change in climate during this century.

Prediction of lake responses to long-term alteration of atmospheric conditions can be based on empirical relationships dependent on climatic statistics (i.e., average weather) or on process-based physical models that operate on daily weather conditions including extreme events. Empirical approaches provide advantages in the form of robustness, particularly when used on multi-lake estimations, such as regional approaches (e.g. Weyhenmeyer *et al.* 2011). On the other hand, because daily weather conditions control many physical mechanisms by which lakes respond, process-based models capable of representing intricate physical mechanisms can be advantageous. For example, deep, large, or clear lakes store more heat in the summer due to greater area-specific heat capacity, greater wind-driven vertical mixing, or deeper penetration of short-wave radiation, than shallow, small, or humic ones (Blenckner 2005), and such processes can be influenced by daily weather progression. Furthermore, process-based models are more immune to artifacts due to temporal data aggregation. For example, empirical models may suffer from multicollinearity among predictors, which can mask regional variation in climate factors.

In the present study, we present projected impacts of the climate change for generic lake types based on a process-based lake thermodynamics model (Saloranta & Andersen 2007), using atmospheric forcing by synthetic weather from an European regional climate model. We used Fennoscandia (Finland, Norway, and Sweden) as the example study area because the region has a compact geographical extent and at the same time broad gradients in boreal, oceanic, and altitudinal levels. Throughout this region, we systematically explored the significance of both physical lake types and geographical variations in atmospheric forcing on lake thermodynamics. Owing to Fennoscandia's climate diversity, our conclusions should also pertain to boreal lakes in other regions of the northern hemisphere.

Materials and methods

The MyLake model. For simulating thermal responses of lakes to climatic forcing we employed a one-dimensional, process-oriented model for simulating mean diurnal vertical distribution of water temperature and density, and evolution of ice and snow cover (MyLake; Saloranta & Andersen 2007). The processes included in MyLake (e.g., surface heat balances, shortwave radiation penetration, vertical diffusion, wind mixing, and ice formation and breakup) are based on physical first principles; this design enabled mechanistic representation of the ways weather conditions influence the distribution of heat in lakes. The MyLake model has been previously applied to a number of locations and its heat balance simulation is well tested against separately taken observations (Saloranta *et al.* 2009; Dibike *et al.* 2011, 2012). The model is capable of simulating phosphorus dynamics through a simple algal growth model, but in the present study all chemical and biological processes were turned off.

Likewise, we ignored the often negligible contribution by catchment runoff to the lake heat budget. These simplifications reduced the computational load considerably and allowed greater spatial and temporal scope to the study. Physically-based models such as MyLake are often calibrated for a site-specific application (Tarantola 2006), but site-specific accuracy was considered irrelevant for the present study.

Weather inputs. The weather simulation used as the input to the lake model was generated by a regional climate model (RCM) application from the ENSEMBLES project (Van der Linden & Mitchell 2009). We used the HIRHAM5 RCM application, with boundary conditions given by the ECHAM5 global circulation model (GCM), conducted by the Danish Meteorological Institute (Van der Linden & Mitchell 2009). The GCM is forced under the Intergovernmental Panel on Climate Change (IPCC) A1B emission scenario (Nakicenovic et al. 2000). In northern Europe, the A1B scenario is anticipated to lead to an annual mean air temperature increase by 2.3 to 5.3 ˚C, depending on the GCM (median 3.2 ˚C), and a 9 % increase in precipitation by 2080-2099 compared with 1980-1999 (Christensen et al. 2007). The air temperature increase is expected be even more pronounced in the north-eastern part of the region when compared spatially, and in the winter when compared seasonally.

Daily mean weather variables were calculated from instantaneous RCM outputs where applicable. We only used approximately a quarter of all the 25 x 25 km Gaussian grid points in Fennoscandia of the original RCM simulation, yielding a 50 x 50 km grid cell size. The use of synthetic daily weather data facilitated simulation of daily lake temperature during the critical phases of ice formation and establishment of summer stratification. The current study ran simulations in two 30-year scenarios: the reference scenario (1970-2000) and the future scenario (2070-2100). The same RCM was used for both scenarios to cancel out linear RCM biases in the scenario contrasts. The 30-year weather inputs were preceded by a spin-up consisting of the first year repeated three times in order to thermally equilibrate the modelled lake to the local climate.

Lake types. For each RCM grid cell, we simulated ice cover and daily vertical temperature profiles for 27 generic physical lake types (hereinafter simply ‘lake types’) comprising of all possible combinations of 3 levels of 3 key lake characteristics depth (6, 24, 96 m), surface area (1, 10, 100 km^2), and vertical light attenuation coefficient (0.18, 0.90, 4.50 m^{-1}) (Table 1). The chosen light attenuation levels approximately correspond to Secchi depths of 10, 2, and 0.4 m (Kirk 1994). Lake bathymetry was assumed to be an inverted cone, such that the mean depth was one-third of the maximum depth, simply referred to as 'depth' in this study. Surface area also determined the parameterisations for wind-driven mixing and vertical diffusion processes, based on empirical relationships from a regional study of lakes in Minnesota, USA (Table 1; Hondzo & Stefan 1993). The levels of the different factors were chosen to cover the range of lakes in the region according to the classification system of the European Union Water Framework Directive; the light attenuation coefficient levels correspond with the clear, humic, and polyhumic lake categories (Carvalho et al. 2008). Use of generic lake typology allowed orthogonal design of the study, which increased the statistical power for assessing the general impact of climate change on lakes of varying type and interaction effects between typological factors. In addition, the variability represented by these generic lake types covers most of the actual lake population.

Analyses and presentation of results. Our simulation study produced daily thermal profiles and ice and snow cover for 30 years using 27 lake types at 472 locations under 2 simulation periods (reference (1970-2000) and future (2070-2100)). Daily model outputs were aggregated as monthly arithmetic means for each individual simulation run, and stored for post-processing. Most results presented here use the 30-year median for modelling for regional gradients, as well as contrasts between lake types and scenarios. Calculation of solar irradiance penetrating the water surface included the albedo and shading effects of ice if present. Epilimnion temperature was defined for the top metre for the shallow lake configuration (6 m) or the top two metres for the mid and deep lake configurations (24 m and 96 m). Hypolimnion temperature was defined for the bottom metre for the shallow lake configuration or the bottom two metres for the mid and deep lake configurations. Each aggregated output variable had altogether $27 \times 472 \times 2 = 25\,488$ entries.

Ice-on day was defined as the first ice covered day after the open-water period with the highest surface water temperature. Ice-off day was defined as the last ice-covered day before such an open-water period, in the following year. Ice duration was defined to be the number of days inclusive between the ice-on and ice-off days of year. For a multi-year contiguous ice cover period, the last year of the period received the remainder of the total ice duration time with the preceding years receiving the full year ice duration. Because winter is spanned through two consecutive calendar years, only 29 sets of these ice phenology variables were available from each 30-year simulation run. Intermittent thawing of ice cover was also modelled and recorded, but due to the way ice phenology variables were calculated this does not affect our presentation.

Regional gradients of parameters affecting the surface heat fluxes of lakes were represented by principal component scores (PCs) of aggregations from the RCM-based weather data. The chosen climatic indicators variables for our context were: whole-year, summer (June and July), and winter (December and January) mean solar irradiance, air temperature and precipitation, and whole-year mean air pressure, cloud cover, humidity, and wind speed (Table 2). All variables were averaged location-wise over the reference period (1970-2000) before principal component decomposition of the resulting column-centred and unit-scaled 472 by 13 matrix.

Process-based lake responses were summarized by generalized additive models (GAMs; Wood 2006) using spline smoothers to represent nonlinearities in climatic gradient effects and linear orthogonal contrasts for lake type factors, simulation scenarios, and their interactions. Effect sizes were computed as differences in explained variance caused by omitting a factor from the full model that includes all secondary pair-wise contrast interactions. These differential variance contributions were then normalised to the null model variance of the corresponding response parameter. We used the R package mcgv to create additive models for predicting the response parameters.

Most of the computation was scripted and run on Amazon Elastic Compute Cloud under the Ubuntu Linux operating system, but our scripts do not use any operating system specific functions. Approximate scale of computation amounted to 850,000 MyLake simulation years, which alone took 1,500 CPU-hours at 2.5 GHz CPU-clock frequency (2.5 Elastic Compute Unit according to Amazon’s abstracted computing power metric). All simulations were performed deterministically, which ensures reproducibility of the results. The script files are available upon request.

Results

We assessed the realism of our simulations by comparing them against a proxy for observations: ice phenology projection by an empirical model (Weyhenmeyer *et al.* 2011).

The empirical model is based on geographical variation of recent northern hemisphere climate observations and statistically relates ambient air temperature and latitude to ice phenology. We applied the Weyhenmeyer model to the same synthetic weather data used as MyLake input, and compared the projected ice phenological events. Whereas this involves extrapolation beyond the temporal scope of the Weyhenmeyer model, the models correspond very well to each other (Fig. 1), with overlapping patterns under both simulation scenarios (1970-2000 and 2070-2100). The close relationship between our model predictions and an empirical model of ice phenology in Sweden supports the credibility of the un-calibrated MyLake model for this purpose. It also indicates that substantial parts of the residual variance in the Weyhenmeyer model, also discussed by the authors, can be explained mechanistically by taking additional lake attributes such as depth into account.

The first four principal components (Table 2) accounted for 90% of the total variance in the regional climatic gradients (Fig. 2). PC1 (41% of total variance) can be interpreted as the boreal temperature gradient with greater values indicating colder climate. The lower values in the future scenario are a consequence of increased air temperature. PC2 (25% of total variance) mostly represents precipitation and is a proxy for the oceanic humidity and precipitation gradient. The greater values in the future scenario are an indication of increase in precipitation. PC3 (17% of total variance) mostly represents the heat transport effect from the North Atlantic Current. In the future scenario the scores are greater, which may suggest that the North Atlantic Current heating effect will be less important. PC4 (7% of total variance) appears to represent high elevation climate, with lower values for the future scenario suggesting that high elevation characteristics will be less pronounced. The four smoothed PCs were all significant ($P < 0.001$) in all GAM models, signifying how the geographical patterns in lake ice and thermal dynamics are driven by the overlying climate. Significant coefficients for other independent variables are shown in Table 3.

Relative importance of explanatory variables (regional climatic gradients and orthogonal lake type contrasts) on monthly aggregated simulation outputs are represented as GAM effect sizes (Fig. 3), with significant contrasts between the reference and future scenarios for all variables. The regional climate gradient effects are stronger for irradiance (Fig. 3a) than for epilimnion and hypolimnion temperatures (Fig. 3b,c), across all months. This decreasing trend follows the proximity to the surface, and is also reflected in the phasing of their summer extremes (Livingstone & Lotter 1998), with maximum solar radiation at midsummer, maximum epilimnion temperature in July, and maximum hypolimnion temperature in August. It is also noticeable that regional climatic effects on hypolimnion temperature (Fig. 3c) were higher in circulation periods (spring and autumn), than in periods with density stratification (summer and winter) when the hypolimnetic environment is practically isolated from atmospheric forcing. Summer stratification stability, represented by the density difference between epilimnion and hypolimnion (Fig. 3d), was clearly more influenced by lake type, especially depth, than regional climatic gradients and climate change scenarios.

The GAM analyses can only explain variance due to additive contrasts between climate scenarios, while some effects are probably better modelled by non-linear terms. Fig. 4 shows the climate change effects in greater detail for one particular lake type with intermediate depth, area, and transparency (24 m, 10 km^2, 0.9 m^{-1}). While the climate change effect on ice duration is close to additive for this particular lake type (Fig. 4a), the effects on winter subsurface irradiance (Fig. 4b) and summer water column stability (Fig. 4d) are closer to multiplicative, and the summer epilimnion temperature effect is uni-modal (Fig. 3c). Ice cover duration is projected to shorten for all lakes in the region from the reference scenario (1970-2000) to the future scenario (2070-2100), with the greatest changes in lakes that currently have the shortest winters (Fig. 4a). Winter sub-surface irradiance is predicted to increase about 6-fold in the current darkest lakes, while the projected change will be less dramatic in lakes with better underwater light conditions in winter (Fig. 4b). Average summer epilimnion temperatures will increase less in the currently warmest lakes than the colder ones, with absolute increases up to 6°C at intermediate temperatures (Fig. 4c). Only a small fraction of the lakes will experience unchanged or reduced density stratification stability in the summer, while the majority is predicted to gain a 2.5-fold increase in the density difference between epi- and hypolimnion (Fig. 4d).

Our simulations indicate that under present-day climate (1970-2000) 99.5 % of all combinations of lake types and simulation grid points develop ice cover in winter (Fig. 5a). In the future scenario (2070-2100), 6.2 % of these lakes are expected to have open water throughout the winter in an average year, with the southernmost part of Sweden being most affected (Fig. 6a). The risk of becoming ice-free varies with depth, with the deepest configuration (96 m depth) being the most vulnerable. For this configuration, lakes with ice duration shorter than 100 days in 1970-2000 are likely not to form ice by 2070-2100 (Fig. 5b). The other two lake type factors, namely surface area and light attenuation coefficient, were less important in determining lake ice phenology. The remainder of the lakes (93.3 %) are simulated to remain winter-freezing by 2070-2100. However, timing of ice-on will on average be delayed by 27 days and ice-off advanced by 24 days, when all lake types are taken together (Table S1). This also entails a substantial increase in the amount of solar radiation that reaches the water phase (Fig. 6b), due to the general thinning of the ice cover in addition to the reduced duration. Winter underwater irradiance is thus predicted to increase 6-fold on average in the future scenario.

The climatic changes projected by the IPCC A1B scenario will also cause lake water temperature to increase throughout the region (Fig. 6c). In the spring, these increases can be attributed to the shortening of ice duration; the earlier ice break in 2070-2100 allows earlier heating of surface water in the spring. In the summer, lake water temperature is predicted to increase during the 21st century by 2.4 to 2.5 ˚C in the epilimnion and 2.0 to 2.3 ˚C in the hypolimnion (Table S1). This is mostly due to the increase in air temperature, which reduces long-wave radiative heat loss from the lake surface.

Discussion

Local climate had a high impact on the heat balance of lakes, showed by large-sized effects of regional climatic gradient in all GAM models (Fig. 3). However, when adjusted for these regional climatic differences, there still remained substantial variance explained by both lake type contrasts and climate scenario effects. The climate sensitivity and succeeding consequences for lakes as habitats and providers of ecosystem services are therefore likely to vary considerably among neighbouring lakes differing in morphometric characteristics.

The most important factor explaining seasonal cycles of physical dynamics among lake type variations was depth; the depth factor was as important as the simulation scenario contrast for explaining epilimnion temperature, and even more important for hypolimnion temperature. Future lake temperature changes will therefore be at least as large as current differences between nearby lakes of different depth classes. There is also a strong seasonal pattern in the effect sizes of lake depth (Fig. 3b,c). Deeper lakes provide greater heat storage during the summer (June-September), thereby increasing the total heat content of the lake. This delays the ice formation for deep lakes, causing significantly higher epilimnion and hypolimnion temperatures during winter (October-March). Spring turnover is mostly explained by climatic gradients and simulation scenario, and therefore apparently resets the differences caused by the depth factor (April-May). The effect of surface area was smaller than for depth, although the effect of increasing surface area on the seasonal pattern was similar to increasing depth. This is likely due to larger lakes having deeper mixing and accordingly also higher summer heat storage.

The effect of light attenuation was highest in summer, but always substantially lower than any of the other explanatory variables. Conceptually, one would expect that lakes with high light attenuation coefficient would have higher epilimnion and lower hypolimnion temperature than clearer ones, due to shallower penetration of solar radiant flux. Yet, relevant physical explanations are that water transparency only affects the visible part of the solar spectrum, while the near infrared component (usually close to 50% of solar energy flux) will in any case be dissipated very close to the surface. Heat is distributed vertically by both radiative and turbulent processes, whose relative importance scale with wind energy transfer efficiency, and thus with lake surface area. This offers an explanation why Fee et al. (1996) found stronger impact of water transparency on vertical temperature distribution in lakes smaller than 0.5 km^2 than in larger ones, and also why we see so little effect of water clarity in our simulations using a minimum lake size of 1 km^2. The literature documents increase in humic content of freshwater systems observed recently (Roulet & Moore 2006) and projected continuation of this trend in the future (Larsen et al. 2010). This brownification trend may have numerous biogeochemical and ecological consequences, but its impact on ice cover and daily mean temperature of lakes are likely to be negligible, with a possible expiation for very small water bodies not covered by the present study.

The process-based understandings from the present study offer an insight into the pronounced residual variability of the ice-on prediction in the Weyhenmeyer *et al.* (2011) model. Apart from the regional and temporal climatic contrasts, our GAM results indicate that lake depth explains more of the residual variability in ice-on time and ice cover duration than the other two lake type factors (Fig. 3), in accordance with the discussion by Weyhenmeyer *et al.* (2011). This implies that the timing of ice cover formation is most crucially dependent on the heat storage capacity of the water column. The main process for heat loss is long-wave back radiation, which depends on the temperature difference between air and water surface. The deeper and the more homogeneously heat is distributed, the less heat is lost by surface long-wave back radiation.

Because ice and snow albedo dramatically affects subsurface light penetration, decreased ice cover duration will most likely increase annual primary productivity in northern lakes, simply by increasing the length of the growing season. The relative impact of changes in ice phenology on primary productivity will be strongest in the lakes that currently have the most extensive ice cover, i.e., arctic and alpine regions. Palaeoecological evidence from arctic lakes indicates that even modest lengthening of the growing season may entail dramatic changes in plankton community structure, as well as in the relative importance of benthic versus pelagic primary production (Smol et al. 2005). Empirical data suggest that earlier ice-off may not only change the timing of the diatom spring bloom but also increase its magnitude (Adrian et al. 1999). The same authors also showed that the spring bloom timing affected the seasonal succession of rotifers and daphnids, but found no direct effects on the timing and maximum of other major groups of autotrophs and herbivores. Shortening of the ice-covered season may expand and dilate the post-spring bloom successional pattern by inducing phenological mismatches between prey and predator (Winder & Schindler 2004), extending phases of over-grazing (Scheffer *et al.* 2001), and prolonging summer blooms of nuisance algae (Hyenstrand et al. 1998).

The physical changes predicted in the present study will have dramatic consequences on lakes as habitats and ecosystems. Ice cover, water temperature, vertical density stratification, and underwater light field are important determinants of physical habitat. Northern freshwater lake ecosystems are populated by species with behavioural and physiological strategies shaped to deal with extended periods of ice cover. The projected change will progressively reduce the relative advantage that winter specialists have over more eurythermal species in interspecific interactions (Pörtner *et al.* 2006), decreasing local abundance and contracting distributional ranges (Finstad *et al.* 2011; Helland *et al.* 2011; Ulvan *et al.* 2012). The effect of shorter ice cover duration may also be intensified by difficulties cold-water fish such as salmonids will face under epilimnetic summer warming (Somero 2010), particularly in small and shallow lakes which may no longer offer hypolimnetic thermal refuges.

Our findings that are based on the currently best available regional climate model for northern Europe support the results of earlier lake temperature modelling studies in North America (Fang & Stefan 1998, Hostetler & Small 1999). Lake thermal balances are predicted to be changing faster than the rates recently documented for northern freshwater systems (Magnuson *et al.* 2000), with magnitudes that are certainly alarming from ecological and human utility points of view. We urge bridging studies that bring process-based simulations such as ours to more ecologically and biologically oriented models in order to assess the pertinent impacts of the climate change on lake biota.

Acknowledgements

We thank Tuomo Saloranta for advising on lake model customisation. This study is funded by Research Council of Norway (NFR 185109, 208418, 190028, 208279). The ENSEMBLES data used in this work were funded by the EU FP6 Integrated Project ENSEMBLES (Contract number 505539).

References

Adrian R, Walz N, Hintze T, Hoeg S, Rusche R (1999) Effects of ice duration on plankton succession during spring in a shallow polymictic lake. *Freshwater Biology*, **41**, 621-632.

Blenckner T (2005) A conceptual model of climate-related effects on lake ecosystems. *Hydrobiologia*, **533**, 1–14.

Carvalho L, Solimini A, Phillips G, *et al*. (2008) Chlorophyll reference conditions for European lake types used for intercalibration of ecological status. *Aquatic Ecology*, **42**, 203–211.

Christensen JH, Hewitson B, Busuioc A, *et al*. (2007) Regional climate projections. In: *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* (eds: Solomon S, Qin D, Manning M, *et al*.), pp847–940.

Dibike Y, Prowse T, Bonsal B, Rham L de, Saloranta T (2012) Simulation of North American lake-ice cover characteristics under contemporary and future climate conditions. *International Journal of Climatology*, **32**, 695–709.

Dibike Y, Prowse T, Saloranta T, Ahmed R (2011) Response of Northern Hemisphere lake-ice cover and lake-water thermal structure patterns to a changing climate. *Hydrological Processes*, **25**, 2942–2953.

Fang X, Stefan HG (1998) Potential climate warming effects on ice covers of small lakes in the contiguous US. *Cold Regions Science and Technology*, **27**, 119-140.

Fee EJ, Hecky RE, Kasian SEM, CruikshankDR (1996) Effects of lake size, water clarity, and climatic variability on mixing depths in Canadian Shield lakes. *Limnology and Oceanography*, **41**, 912-920.

Finstad AG, Forseth T, Jonsson B, *et al*. (2011) Competitive exclusion along climate gradients: energy efficiency influences the distribution of two salmonid fishes. *Global Change Biology*, **17**, 1703–1711.

Finstad AG, Forseth T, Næsje TF, Ugedal O (2004) The importance of ice cover for energy turnover in juvenile Atlantic salmon. *Journal of Animal Ecology*, **73**, 959–966.

Helland IP, Finstad AG, Forseth T, Hesthagen T, Ugedal O (2011) Ice-cover effects on competitive interactions between two fish species. *Journal of Animal Ecology*, **80**, 539–547.

Hondzo M, Stefan HG (1993) Lake water temperature simulation model. *Journal of Hydraulic Engineering*, **119**, 1251.

Hostetler SW, Small EE (1999) Response of north American freshwater lakes to simulated future climates. *Journal of the American Water Resources Association*, **35**, 1625-1637.

Hyenstrand P, Blomqvist P, Pettersson A (1998) Factors determining cyanobacterial success in aquatic systems: a literature review. *Archiv fur Hydrobiologie, Ergebnisse der Limnologie*, **51**, 41-62.

Kirk JTO (1994) *Light and Photosynthesis in Aquatic Ecosystems*. Cambridge University Press.

Van der Linden P, Mitchell J (2009) *ENSEMBLES: Climate Change and its Impacts: Summary of research and results from the ENSEMBLES project*.

Larsen S., Andersen T. And Hessen DO (2011) Climate change predicted to cause severe increase of organic carbon in lakes. *Global Change Biology*, **17**: 1186–1192.

Livingstone DM, Adrian R (2009) Modeling the duration of intermittent ice cover on a lake for climate-change studies. *Limnology and Oceanography*, **54**, 1709–1722.

Livingstone DM, Lotter AF (1998) The relationship between air and water temperatures in lakes of the Swiss Plateau: a case study with palæolimnological implications. *J. Paleolimnol*., **19**, 181–198.

Magnuson JJ, Robertson DM, Benson BJ, *et al*. (2000) Historical trends in lake and river ice cover in the northern hemisphere. *Science*, **289**, 1743–1746.

Nakicenovic N, Alcamo J, Davis G, *et al*. (2000) *Special report on emissions scenarios: a special report of Working Group III of the Intergovernmental Panel on Climate Change*. Pacific Northwest National Laboratory, Richland, WA (US), Environmental Molecular Sciences Laboratory (US).

O'Brien WJ (1979) The predator-prey interaction of planktivorous fish and zooplankton: Recent research with planktivorous fish and their zooplankton prey shows the evolutionary thrust and parry of the predator-prey relationship. *American Scientist*, **67**, 572–581.

Pörtner HO, Bennett AF, Bozinovic F, *et al*. (2006) Trade-offs in thermal adaptation: The need for a molecular to ecological integration. *Physiological and Biochemical Zoology*, **79**, 295–313.
Roulet N, Moore TR (2006) Environmental chemistry: Browning the waters. *Nature*, **444**, 283–284.
Saloranta TM, Andersen T (2007) MyLake - A multi-year lake simulation model code suitable for uncertainty and sensitivity analysis simulations. *Ecological Modelling*, **207**, 45–60.
Saloranta TM, Forsius M, Jarvinen M, Arvola L (2009) Impacts of projected climate change on the thermodynamics of a shallow and a deep lake in Finland: model simulations and Bayesian uncertainty analysis. *Hydrology Research*, **40**, 234–248.
Scheffer M, Straile D, Nes EH van, Hosper H (2001) Climatic warming causes regime shifts in lake food webs. *Limnology and Oceanography*, **46**, 1780–1783.
Schröter D, et al. (2005) Ecosystem Service Supply and Vulnerability to Global Change in Europe. *Science,* **25**, 1333-1337.
Shuter BJ, Finstad AG, Helland IP, Zweimüller I, Hölker F (2012) The role of winter phenology in shaping the ecology of freshwater fish and their sensitivities to climate change. *Aquatic Sciences,* **74**, 637 - 657
Smol JP, Wolfe AP, Birks HJB, *et al*. (2005) Climate-driven regime shifts in the biological communities of Arctic lakes. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 4397–4402.
Somero GN (2010) The physiology of climate change: How potentials for acclimatization and genetic adaptation will determine "winners" and "losers." *The Journal of Experimental Biology*, **213**, 912–920.
Tarantola A (2006) Popper, Bayes and the inverse problem. *Nature Physics*, **2**, 492–494.
Ulvan E, Finstad A, Ugedal O, Berg O (2012) Direct and indirect climatic drivers of biotic interactions: ice-cover and carbon runoff shaping Arctic char *Salvelinus alpinus* and brown trout *Salmo trutta* competitive asymmetries. *Oecologia*, **168**, 277–287.
Weyhenmeyer GA, Livingstone DM, Meili M, Jensen O, Benson B, Magnuson JJ (2011) Large geographical differences in the sensitivity of ice-covered lakes and rivers in the Northern Hemisphere to temperature changes. *Global Change Biology*, **17**, 268–275.
Winder M, Schindler DE (2004) Climate change uncouples trophic interactions in an aquatic ecosystem. *Ecology*, **85**, 2100–2106.
Wood SN (2006) *Generalized additive models: an introduction with R*. Chapman & Hall/CRC.

Supporting Information legend

The coefficients for the GAM models are available in the Supporting Information, Table S1.

Tables

Table 1. Lake typology levels and derived quantities. The wind sheltering coefficient (*) refers to Saloranta and Andersen (2007), equations 13 and 14, and Hondzo and Stefan (1993). The diffusion coefficient scaling parameter (**) refers to Saloranta and Andersen (2007), equation 10, and Hondzo and Stefan (1993). Secchi disk transparency (***) is not used as a MyLake input parameter, but was used as a basis to determine the light attenuation levels in the lake typology.

depth level	**d0**	**d1**	**d2**
depth (m)	6	24	96
depth resolution (m)	0.5	1.0	1.0
surface area level	**s0**	**s1**	**s2**
surface area (km2)	1	10	100
wind sheltering coefficient*	0.26	0.95	1
scaling of vertical diffusion coefficient**	0.007	0.026	0.093
light attenuation level	**a0**	**a1**	**a2**
light attenuation coefficient (1/m)	0.18	0.90	4.5
approximate Secchi depth (m)***	10	2	0.4

Table 2. Loadings for principal components of climatic indicators. Principal component decomposition of mean daily RCM-generated weather data for the reference period (1970-2000), standardised by mean centring and standard deviation scaling. Summer values refer to the arithmetic means for June and July, and winter values for December and January. The RCM results themselves are results from a European project ENSEMBLES.

	PC1	PC2	PC3	PC4
Radiation, annual	-0.37	0.24	-0.12	0.26
Radiation, summer	-0.41	0.08	-0.11	0.01
Radiation, winter	-0.19	0.25	0.36	0.47
Cloud cover, annual	0.21	-0.25	0.40	-0.29
Air temperature, annual	-0.39	-0.03	0.27	-0.11
Air temperature, summer	-0.32	-0.22	0.30	0.11
Air temperature, winter	-0.37	0.10	0.19	-0.27
Humidity, annual	0.28	-0.15	0.40	0.06
Air pressure, annual	-0.26	-0.29	0.15	-0.41
Wind speed, annual	-0.09	0.28	-0.33	-0.54
Precipitation, annual	0.15	0.45	0.24	-0.15
Precipitation, summer	0.21	0.43	0.08	-0.01
Precipitation, winter	0.02	0.42	0.36	-0.21

Figure legends

Figure 1. Ice phenology simulated by MyLake compared with predictions based on the empirical model by Weyhenmeyer *et al.* (2011). Comparisons were made for the three variables most relevant in relation to historical records (Magnuson *et al.* 2000): **a**, ice-on day of year, **b**, ice-off day of year, and **c**, ice duration (days). Matches were made for 472 locations under 2 simulation scenarios (1970-2000 and 2070-2100), using 2 lake types, i.e., shallow (6 m) or deep (96 m) configurations with the intermediate levels of surface area (10 km^2) and light attenuation coefficient (0.9 m^{-1}). Different colours were used for contrasting simulation period and depth. The Weyhenmeyer model is not applicable for semi-freezing lakes that experience frequent thawing during the winter (Livingstone & Adrian 2009), but we nevertheless present the comparison (using open symbols) since MyLake also calculates ice phenology values for such lakes.

Figure 2. Scores of the four first principal components of weather data for the reference scenario (1970-2000) together with scatter-plots showing the expected change in these gradients during the 21st century. Variable loadings are shown in Table 2. The future scenario (2070-2100) principal component scores were calculated with the same centring and scaling of variables as for the reference scenario in order to facilitate direct comparison.

Figure 3. Relative importance of factors influencing underwater irradiance, water temperature, and water column stability. Monthly effect sizes expressed as fractions of explained variance in generalized additive models (GAMs) of 4 state variables: **a**, sub-surface solar radiation adjusted for ice cover albedo, **b**, epilimnion temperature, **c**, hypolimnion temperature, and **d**, density difference between epilimnion and hypolimnion. Effect sizes are represented non-cumulatively by fractions of explained variance on a square-root scale, with

inner circles spaced at every 0.2 unit. The outer thick line is the total explained variance by the full model.

Figure 4. Predicted effects of climate change during the 21st century on underwater light and water columns stability for all RCM grid locations (n = 472). Scatter plots of the future scenario (2070-2100) against the reference scenario (1970-2000) for: **a**, ice cover duration, **b**, sub-surface solar radiation in winter (log-scaled), **c**, epilimnion temperature in summer, and, **d**, density difference between epilimnion and hypolimnion in summer. Winter values refer to averages for December and January, while summer values are averages for June and July. Only values for the lake type combining the intermediate levels of depth (24 m), surface area (10 km^2), and light attenuation coefficient (0.9 m^{-1}) are shown. Dots are plotted with reduced opacity to give more weight to overlapping data points. Dashed lines indicate (a) 50 days absolute decrease, (b) 10-fold relative increase, (c) 5 ˚C absolute increase, and (d) 2.5-fold relative increase.

Figure 5. Ice cover durations under current and future climate, grouped by lake depth (n = 472). Ice durations for all lakes within a depth category were averaged and ranked among RCM grid locations. The resulting ranking was used to draw lines representing the day of year for ice-on (left curve) and ice-off (right curve), in the reference period (1970-2000; panel **a**) and the future climate scenario (2070-2100; panel **b**).

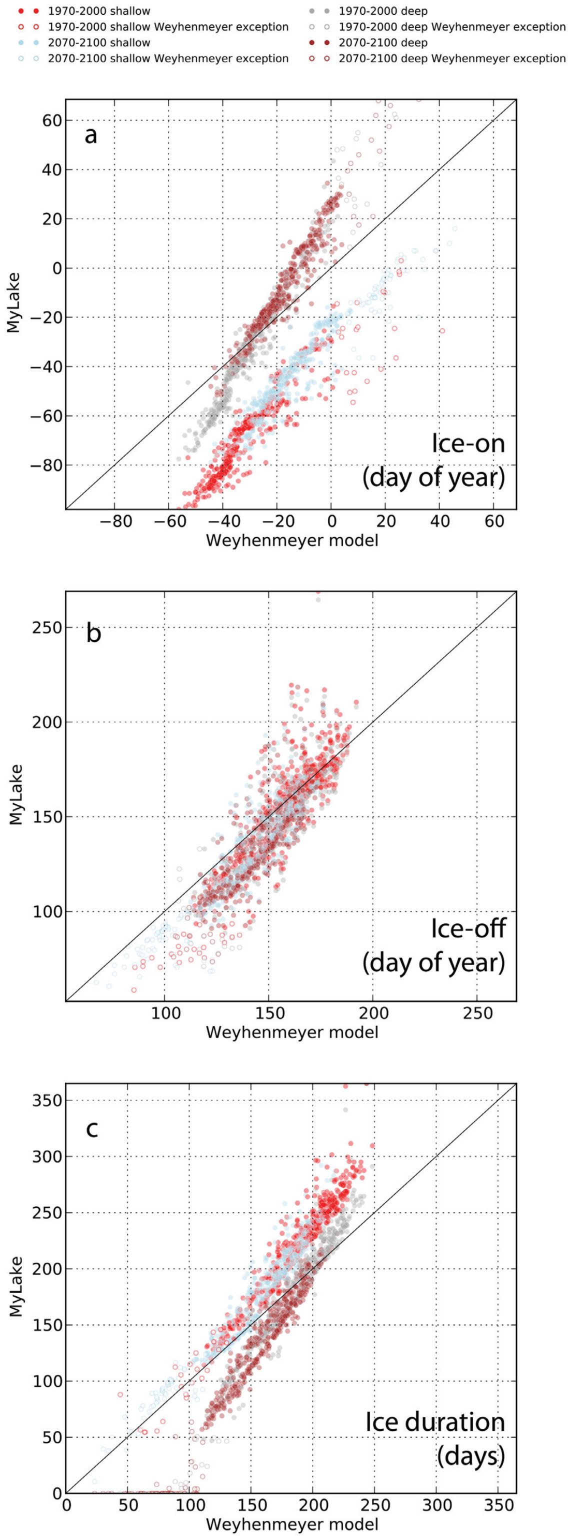
Figure 6. Spatial distribution of predicted impacts of climate change during the 21st century on three key variables related to the heat balance of lakes. a, annual ice duration, **b**, mean winter (December and January) sub-surface solar radiation, and **c**, mean summer (July and August) epilimnion temperature, under the reference (1970-2000) and the future

(2070-2100) scenarios shown in the left and right panel, respectively. Only values for the lake type combining the intermediate levels of depth (24 m), surface area (10 km^2), and light attenuation coefficient (0.9 m^{-1}) are shown. Maps were generated by interpolating the 472 RCM grid vertices by simple kriging with a linear semi-variogram model.

Author contributions

T.A. conceived the idea of connecting climate change forecast to lake physical response using a process-based model. All authors contributed to scientific design and interpretation of results. K.T. is responsible for the technical design and implementation, actual simulation runs, and post-simulation analyses including figure productions. K.T. took the lead in writing the manuscript, with all other authors substantially contributing at every stage of writing.

Figure 1


1970-2000 shallow
1970-2000 shallow Weyhenmeyer exception
2070-2100 shallow
2070-2100 shallow Weyhenmeyer exception
1970-2000 deep
1970-2000 deep Weyhenmeyer exception
2070-2100 deep
2070-2100 deep Weyhenmeyer exception
a
Ice-on
(day of year)
MyLake
Weyhenmeyer model
b
Ice-off
(day of year)
MyLake
Weyhenmeyer model
c
Ice duration
(days)
MyLake
Weyhenmeyer model

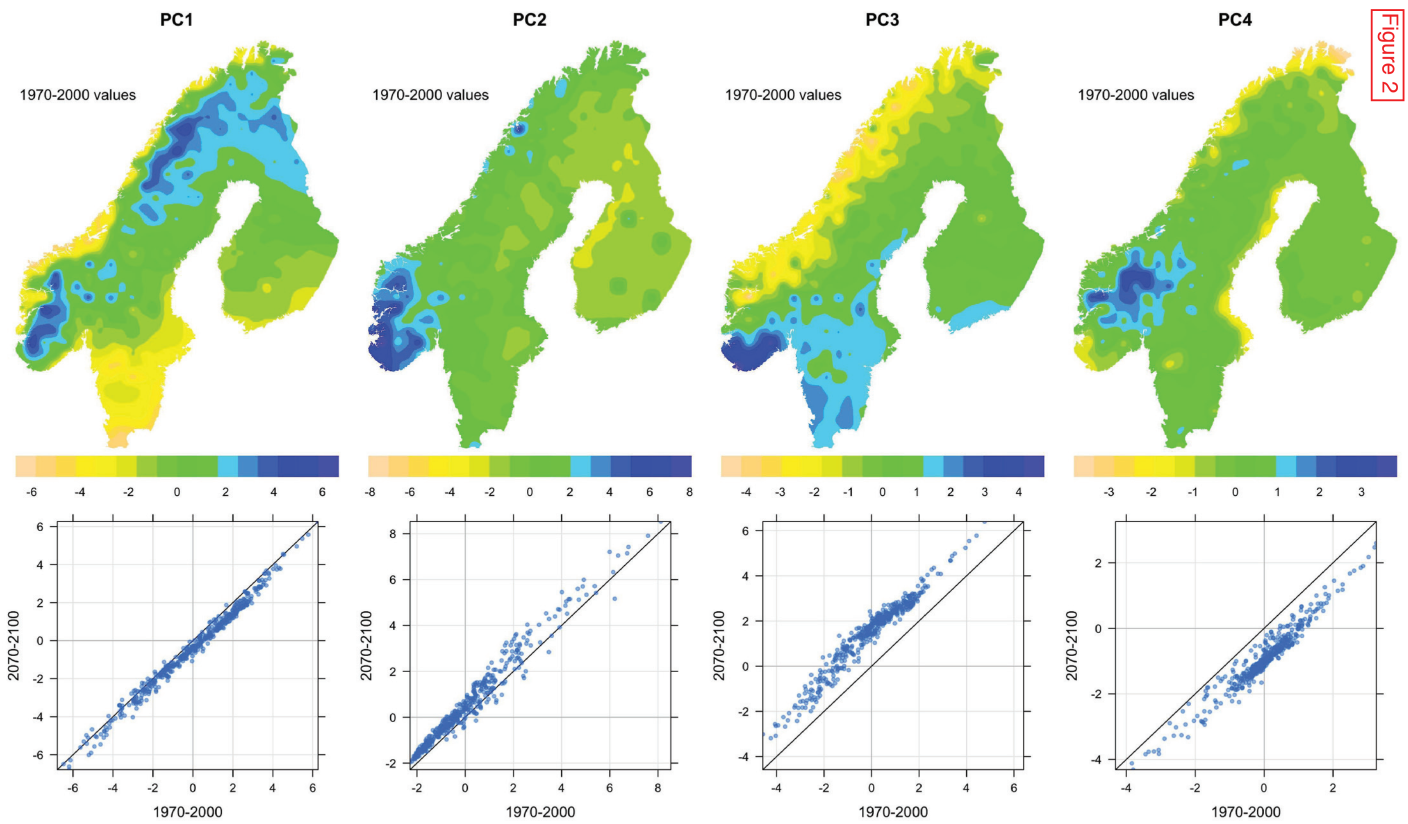
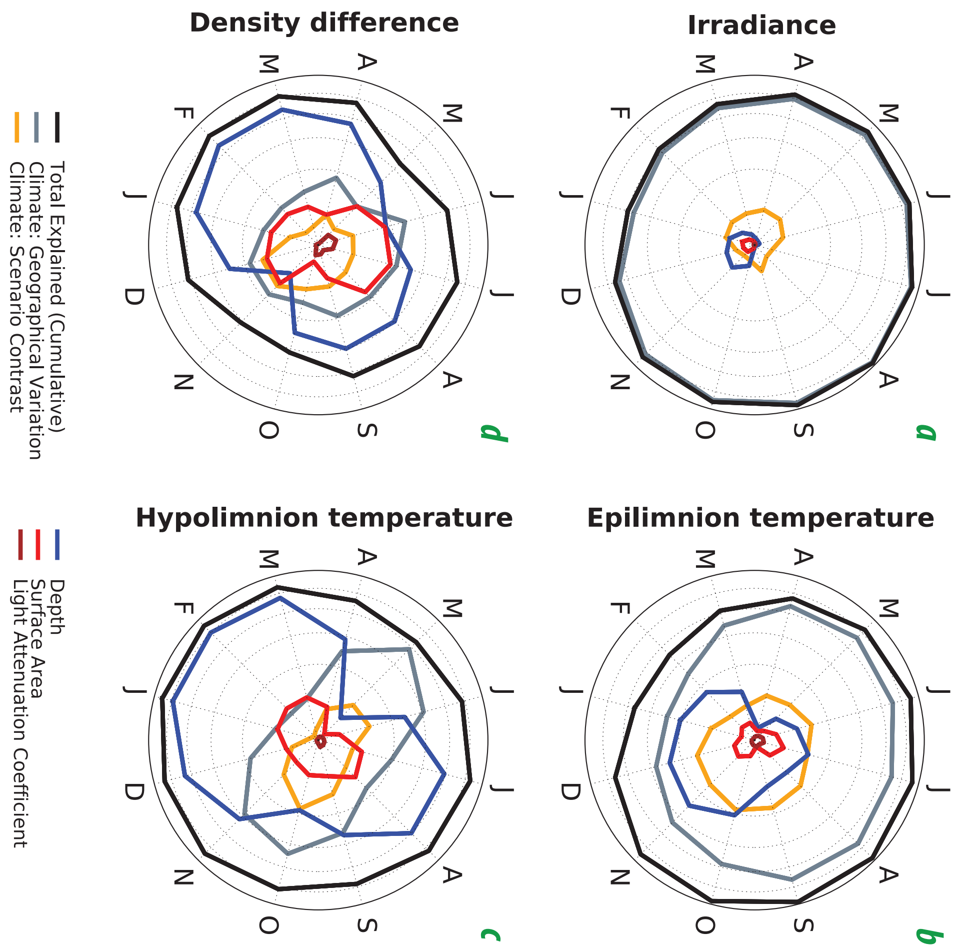


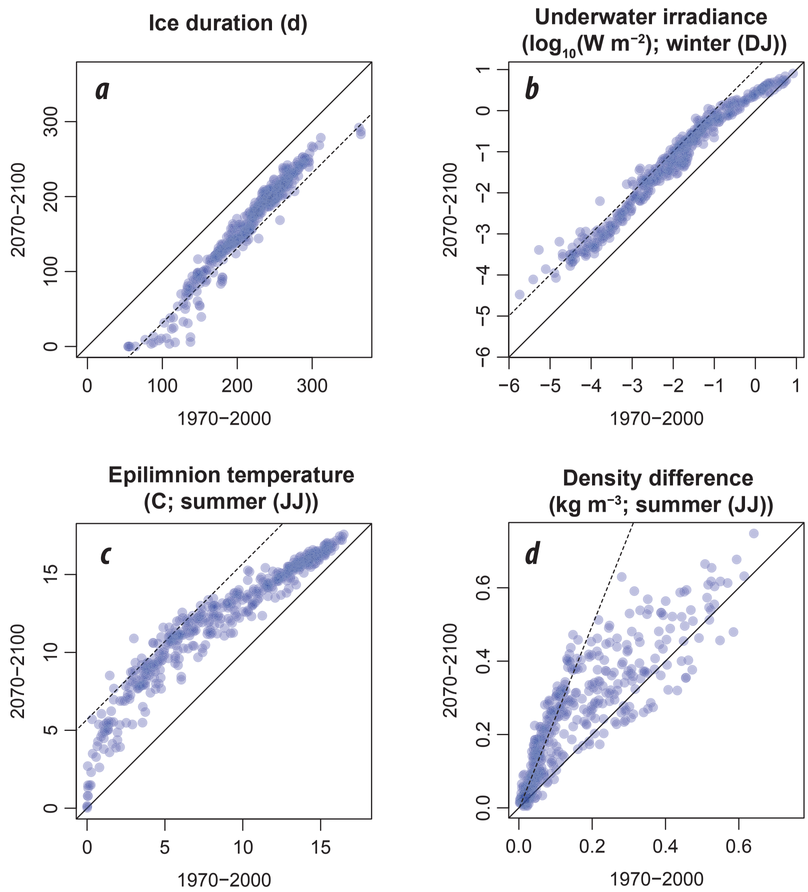
Figure 2

Figure 3



Irradiance
Epilimnion temperature
Hypolimnion temperature
Density difference
a
b
c
d
J
F
M
A
M
J
J
A
S
O
N
D
Total Explained (Cumulative)
Climate: Geographical Variation
Climate: Scenario Contrast
Depth
Surface Area
Light Attenuation Coefficient

Figure 4



Ice duration (d)
a
2070−2100
1970−2000
0
100
200
300
Underwater irradiance
(log$_{10}$(W m^{-2}); winter (DJ))
b
−6
−5
−4
−3
−2
−1
0
1
Epilimnion temperature
(C; summer (JJ))
c
0
5
10
15
Density difference
(kg m^{-3}; summer (JJ))
d
0.0
0.2
0.4
0.6

Figure 5

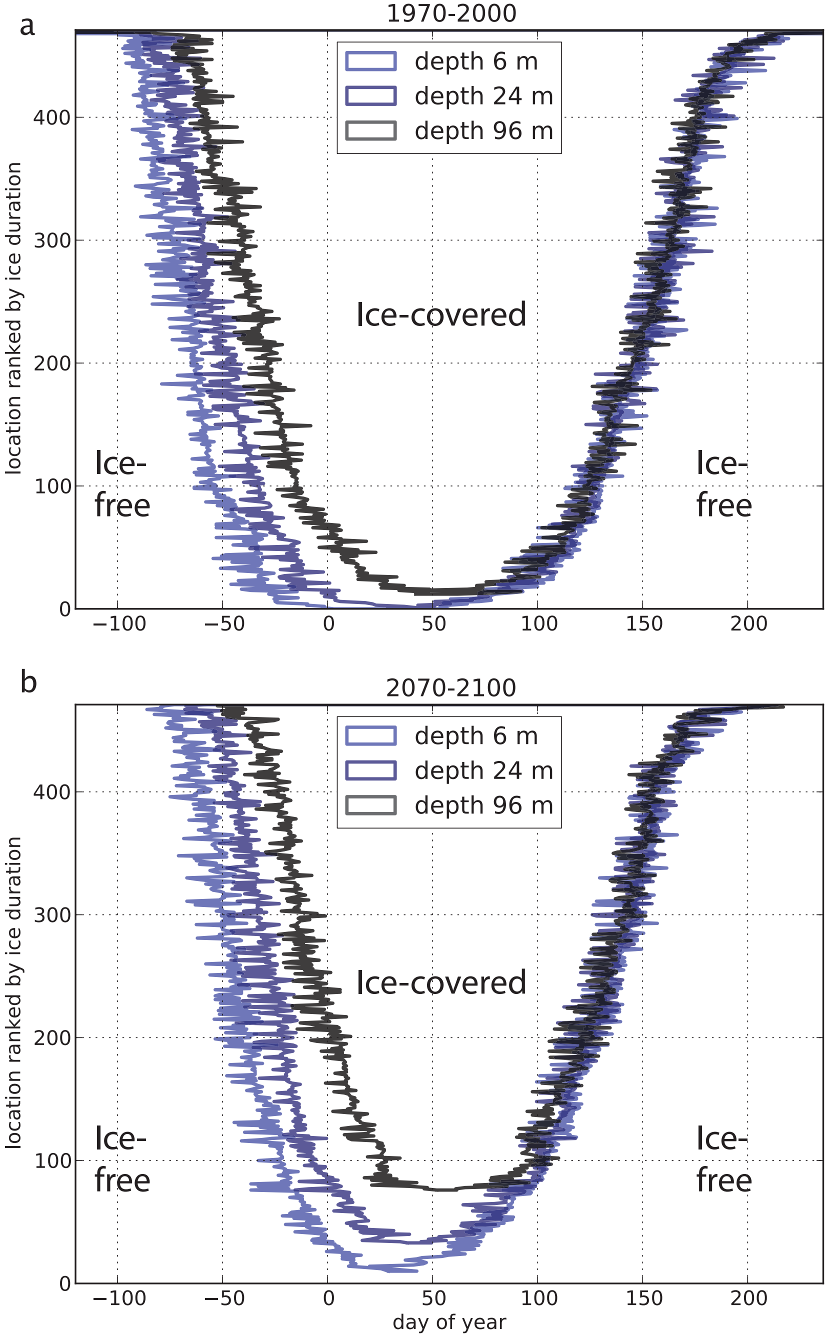

a
1970-2000
depth 6 m
depth 24 m
depth 96 m
Ice-covered
Ice-free
Ice-free
location ranked by ice duration
b
2070-2100
depth 6 m
depth 24 m
depth 96 m
Ice-covered
Ice-free
Ice-free
location ranked by ice duration
day of year

Figure 6

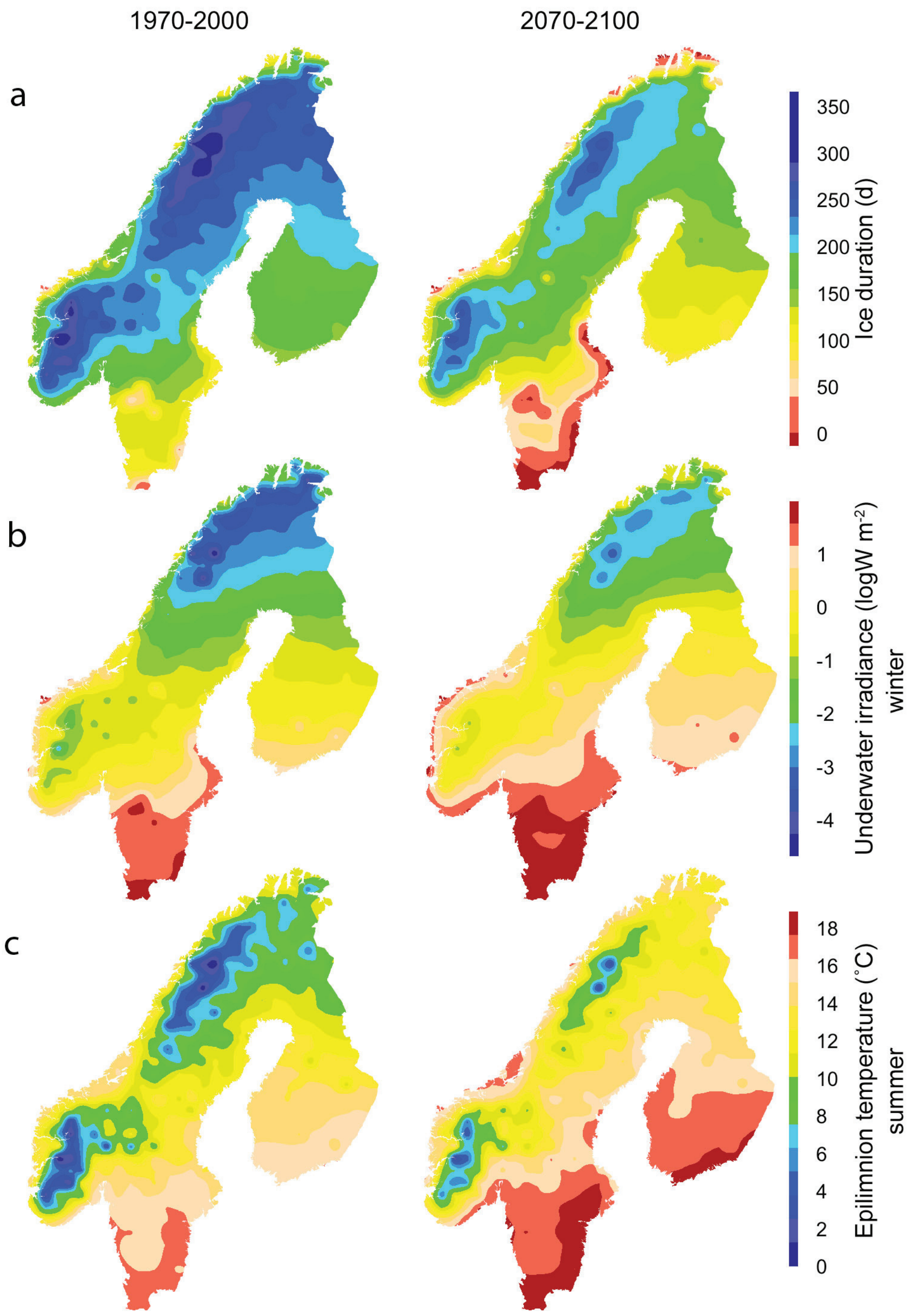

1970-2000
2070-2100
a
b
c
350
300
250
200
150
100
50
0
Ice duration (d)
1
0
-1
-2
-3
-4
Underwater irradiance (logW m-2)
winter
18
16
14
12
10
8
6
4
2
0
Epilimnion temperature (˚C)
summer

Table S1. GAM effect sizes on selected lake simulation variables. Only significant coefficients were shown ($P < 0.05$). The zero values are shown where coefficients were significant but were small enough to be rounded to zero at the significant digits. Intercept coefficient is the expected value for the reference scenario (1970-2000) for the shallow (6 m), small (1 km^2), and clear (light attenuation coefficient at 0.18 m^{-1}) lake type. Single-factor contrast levels were represented by the following: 2070 for the future scenario (2070-2100) vs. present; a1 for medium coloured lakes (light attenuation coefficient 0.90 m^{-1}) vs. clear; a2 for highly coloured lakes (light attenuation coefficient (4.50 m^{-1}) vs. clear; s1 for medium-sized lakes (10 km^2) vs. small; s2 for large lakes (100 km^2) vs. small; d1 for medium-depth lakes (24 m) vs. shallow; and d2 for deep lakes (96 m) vs. shallow. Interaction effects were denoted by the colon symbol (:). For example, the regional average (i.e., adjusted for climatic gradient contributions) for the predicted July hypolimnion temperature in a clear lake (a0) with an intermediate size (s1) and a depth (d1) in 1970-2000 will be 11.9 + 0.2 - 6.2 + 3.6 = 9.5 ˚C, and the average change for the same lake type from 1970-2000 to 2070-2100 will be 2.0 + 0.5 - 0.2 = 2.3 ˚C.

Non-linear geographical effects were all significant ($P < 0.001$). Deeper lakes may appear colder than shallower lakes in the summer (June-August) but this is not an indicative of less heat content in deep lakes. Epilimnion temperature is colder for deep lakes because of the vertical diffusion into deeper part of the lake. Hypolimnion is colder because of the definition of the hypolimnion temperature. The higher temperature for the autumn (October-November) rather indicates greater heat content stored in deep lakes compared with shallower lakes.

Modelled variable		Single-factor effects							Interaction effects																	
	Intercept	2070	a1	a2	s1	s2	d1	d2	2070:a1	2070:a2	2070:s1	2070:s2	2070:d1	2070:d2	a1:s1	a2:s1	a1:s2	a2:s2	a1:d1	a2:d1	a1:d2	a2:d2	s1:d1	s2:d1	s1:d2	s2:d2
Ice phenology																										
First ice-on doy	-65.7	24.0			1.0	1.1	10.7	21.8			0.7	0.9	2.4	6.1				-0.7				1.5	3.9	5.7	11.3	15.6
Last ice-off doy	150.7	-24.1		1.0			-1.5	-2.4						1.6										0.6	-0.6	-0.7
Number of ice-free days within ice duration	1.4	0.4					-0.9	-1.4					-0.6	-0.3												
Ice duration (days)	217.1	-49.5		1.0			-12.6	-24.2			-0.9	-1.2	-2.2	-4.5								-1.2	-3.3	-4.8	-11.5	-15.8
Epilimnion temperature (˚C)																										
Jan	0.1				-0.1	-0.1	0.2	0.5			0.1	0.2	0.1	0.7								0.0	-0.1	-0.2	0.1	0.2
Feb	0.1	0.1			0.0	-0.1	0.2	0.5			0.0	0.1		0.2									-0.1	-0.2	-0.1	-0.1
Mar	0.1	0.2			0.0	0.0	0.2	0.5				0.0	-0.1	0.1								0.0	-0.1	-0.2	-0.1	-0.2
Apr	0.4	1.0		-0.1			0.1	0.3		-0.1			-0.2	-0.3								0.1	-0.1	-0.2	-0.1	-0.2
May	2.7	2.8		-0.3			-0.3	-0.3		-0.1	-0.2	-0.2	-0.5	-1.2				0.1				0.2	-0.4	-0.5	-0.7	-0.9
Jun	7.5	4.4		-0.6			-0.6	-0.8		-0.1	-0.4	-0.4	-0.5	-1.4				0.2					-0.7	-0.8	-1.9	-2.5
Jul	12.3	2.4	0.1	-0.5			-0.7	-1.3					0.5	0.1									-0.4	-0.5	-2.1	-3.0
Aug	10.8	2.5	0.2	-0.2	0.0	0.1	0.5	-0.2		0.1		0.0		-0.2	0.0	-0.1	-0.1	-0.1		0.1		0.3	0.3	0.4	-1.1	-1.6
Sep	6.7	2.5	0.1	-0.1	0.1	0.1	1.7	1.3		0.1			-0.2	-0.5	0.0	-0.2	-0.1	-0.2		0.1	0.1	0.5	0.4	0.4	0.1	
Oct	2.5	2.5			0.0	0.0	1.7	2.6		0.0			0.3	-0.4		-0.1		-0.2		0.1	0.1	0.4	0.1	0.2	0.9	1.2
Nov	0.7	1.2			-0.1	-0.1	0.7	1.7			0.1	0.2	0.5	1.0		-0.1		-0.1				0.2			1.1	1.4
Dec	0.2	0.2			-0.2	-0.2	0.3	0.6			0.3	0.3	0.3	1.3								0.1	-0.1	-0.1	0.6	0.8
Hypolimnion temperature (˚C)																										
Jan	0.6	0.1			-0.5	-0.5	1.7	3.0			0.0	0.1	-0.1	0.1								0.0	-0.5	-0.8	0.1	0.1
Feb	0.4	0.1			-0.3	-0.4	1.8	3.2					-0.2	-0.1								0.0	-0.7	-0.9	-0.1	-0.3
Mar	0.4	0.3			-0.2	-0.3	1.9	3.2					-0.3	-0.3								0.0	-0.7	-1.0	-0.3	-0.4
Apr	0.7	0.9			-0.2	-0.2	1.8	3.0			0.1	0.1	-0.6	-1.0								0.1	-0.7	-1.0	-0.3	-0.4
May	2.7	2.4		-0.1			0.3	1.1			0.3	0.3	-1.3	-2.3				-0.1		0.2		0.3	-0.3	-0.5	-0.3	-0.4
Jun	7.4	3.5	0.2	-0.2	0.2	0.2	-3.1	-3.4			0.6	0.7	-1.6	-3.5		-0.2		-0.3		0.5		0.4	1.2	1.7	-0.2	-0.2
Jul	11.9	2.0	0.2		0.2	0.2	-6.2	-7.7			0.5	0.7	-0.2	-2.1		-0.2		-0.4		0.7	-0.2	0.4	3.6	4.5	0.3	0.4
Aug	10.8	2.1	0.2				-3.7	-6.5			0.5	0.6	-0.4	-2.2		-0.4		-0.5		0.6	-0.1	0.3	4.3	4.5	1.0	1.3
Sep	6.8	2.2					0.5	-2.5		0.1	0.4	0.4	-0.4	-2.1		-0.4		-0.4		0.4		0.3	1.5	1.5	1.3	2.0
Oct	3.0	2.1			-0.4	-0.5	1.7	1.2			0.3	0.5	0.2	-1.5								0.2	0.2	0.2	1.5	2.4
Nov	1.6	0.9			-0.7	-1.0	1.4	2.3			0.4	0.6	0.2	-0.1								0.1			1.3	1.9
Dec	0.9	0.2			-0.7	-0.8	1.5	2.7			0.2	0.3		0.3								0.1	-0.3	-0.5	0.6	0.7
Incoming shortwave solar radiation on water surface (MJ m-2 day-1)																										
Jan	0.2	0.3									0.1	0.1	0.1	0.3											0.3	0.4
Feb	0.5	1.0									0.1	0.3	0.2	0.8											0.4	0.7
Mar	2.3	3.7										0.5	0.5	1.6											0.7	1.2
Apr	11.2	13.1												1.4											1.2	1.8
May	48.2	23.9					1.5	2.8																		1.7
Jun	97.8	18.3					2.0	4.0					-1.0	-2.4												
Jul	116.1	-6.4						1.1						-0.9												
Aug	79.4	-4.6																								
Sep	44.2	-4.6					0.3	0.3					-0.3	-0.3												
Oct	14.3	0.9			0.3	0.3	1.9	3.0			-0.4	-0.4	-1.7	-2.7									0.2	0.3	0.2	0.2
Nov	1.9	0.8			0.2	0.2	0.5	1.3			-0.2	-0.3	-0.2	-0.7									0.2	0.2	0.3	0.4
Dec	0.2	0.3					0.1	0.3				0.0		-0.1									0.0	0.1	0.2	0.3

III

Calibration uncertainty in sequential model application

Koji Tominaga
Department of Biosciences, University of Oslo
Norwegian Institute for Water Research

Alexander M. Engebretsen
Department of Chemistry, University of Oslo
Norwegian Institute for Agricultural and Environmental Research

Jostein Starrfelt
Norwegian Institute for Water Research

Øyvind Kaste
Norwegian Institute for Water Research

Rolf D. Vogt
Department of Chemistry, University of Oslo

Tom Andersen
Department of Biosciences, University of Oslo

Abstract

Probabilistic parameter estimation methods such as Markov chain Monte Carlo are increasingly used to estimate quantify uncertainty in poorly known parameters of dynamic process-oriented environmental models. Although these methods are useful in illustrating the parameter uncertainty, cascading impacts of the parameter uncertainty of sequentially connected models are poorly understood. We address this type of uncertainty first by classifying it into three categories, and then by performing a case study on eutrophication to demonstrate different sources of uncertainty. The results suggest that 1) calibration of the downstream model are influenced by the choice of upstream model simulation, that 2) input uncertainty

on the simulation of the downstream model can also be as important for the final forecast as the parameter uncertainty of the downstream model itself, and that 3) it is possible to combine these uncertainty types to produce overall parameter uncertainty of two connected models in model projections. The scenario analysis is demonstrated by forecasting the effectiveness of an economically expensive measure against the eutrophication problem. The present investigation forms a generic platform for carefully studying this type of sequential modelling uncertainty also in other application fields of environmental modelling.

1 Introduction

Methods for probabilistically determining parameter uncertainty of dynamic process-based environmental models became prominent in the past decades (see, for example, *Andrieu et al.*, 2003; *Beven and Binley*, 1992; *Tarantola*, 2006). These methods effectively summarize the covariance structure of a high-dimensional parameter space, provide less subjective calibration and reveal forcast uncertainty caused by ambiguously defined parameterization. They have been utilized in many application fields of environmental modelling, and they are increasingly favoured by resource managers for forecasting future changes that are ultimately linked to probabilistic cost-benefit analyses (*Barton et al.*, 2012).

Many of these methods work within an inverse problem, in which parameter values of a model is estimated so that, with given forcing inputs, model output reproduces independently taken observed measurements (*Beven and Freer*, 2001). It is increasingly recognized that multiple combinations of parameter values can be considered equally likely to be a true candidate, because they result in similar overall proximity to the target observations (i.e. equifinality).

Recognition of the equifinality phenomenon has been an important step forward in parameter uncertainty study, though cascading impacts of this phenomenon on the true parameter uncertainty are often poorly treated. This is especially pertinent when two or more models are connected and applied sequentially and their parameters were calibrated individually. Connecting one environmental model to another is now ubiquitous. For example, with the forthcoming climate change (*Solomon et al.*, 2007), forecasts by climate models can be used as inputs to for example models of surface waters, forestry or agricultural systems to determine possible impacts of the climate change on these landscapes. Connecting models is also a common practice on a regional or local scale. For example in order to simulate the water quality in lakes with short water renewal time, it is often necessary to model major processes governing temporal variability in fluxes from the catchment. Outputs from such catchment models are then used as an input to a lake model that eventually provides the ultimate simulation results. In these cases, properly treating cascading impacts of the equifinality phenomenon can be crucial.

Here, we first define three forms of uncertainty in relation to equifinality of connected models. Thereafter we discuss the extent of these uncertainties in an example case of catchment and lake modelling in a eutrophication study.

2 Typology of calibration uncertainties in sequantial model application

2.1 Own parameterization uncertainty

Probabilistic calibration methods such as Markov chain Monte Carlo (MCMC) reveal sets of calibrated model parameter values yielding simulation outputs that are deemed equally likely. This uncertainty is commonly referred to as "parameter uncertainty" or "calibration uncertainty." These designations are however not precise and specific enough for the present discussion, and therefore we call it own parameterization uncertainty (OPU) in the present study.

Let the model be a differential equation

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = f_x(\mathbf{x}(t), \mathbf{u}(t)|\mathbf{p}_x),$$

where $\mathrm{x}(t)$ is a vector of state variables at time t, $\mathrm{u}(t)$ is a vector of forcing input variables at t, and p_x is parameterization of model dynamics f_x (Figure 1). Probabilistic calibration methods provide a set of equally likely parameter vectors $\mathbf{P}_x = \{\mathbf{p}_x\}_i$. Then we make a set of equally likely simulations $\{\mathbf{x}(t, \mathbf{u}(t), \mathbf{p}_{x,i})\}_i$. OPU is contained in this simulation set and is thus quantified. Arbitrarily chosen quantiles such as the 95-percentile might be used to produce a confidence envelope of the simulation.

2.2 Calibration input uncertainty

As described earlier, modelling of nutrient budgets in freshwater systems may require connecting a hydrologically upstream model to a downstream model. However, use of output from one model as input to another model creates new sources of uncertainty. When both models require calibration, the upstream model's OPU causes ambiguity in the inputs for the downstream model. At the same time, calibration and hence the posterior parameter space of the downstream model depend on the input data during calibration. In the present study, the upstream model (catchment model or lake model) provides nutrient and suspended particle loadings to the downstream model (lake model). In one case we may choose a particular model run of the upstream model that has simulated high nutrient loading to use as the input for lake model during calibration. The lake model calibration then needs to strengthen scavenging processes such as particle sinking in order to match with the target P concentration measured in the lake. In an alternative case, we could have chosen another catchment model run which provides low P loading, which is yet equally likely as input for lake model calibration. The lake model calibration would then respond by reducing the particle sinking velocity. This creates a different but equally likely calibration. This type of impreciseness in the downstream model calibration is not only inconvenient, but potentially has a significant impact for the final simulation. We will call this type of uncertainty due to ambiguity in the calibration input calibration input uncertainty (CIU).

CIU accounts for that the input to the downstream model is a result of the upstream model that requires calibration and hence contains probabilistic ambiguity. Let the upstream model be a differential equation

$$\frac{\mathrm{d}\mathbf{u}}{\mathrm{d}t} = f_u(\mathbf{u}(t), \mathbf{v}(t)|\mathbf{p}_u),$$

where $\mathbf{u}$ is the state variables to be used as input to the downstream model, $\mathbf{v}$ is the forcing variables, and $\mathbf{p}_u$ is parameterization of model dynamics f_u (Figure 1). Let us say that a calibration of the upstream model yielded a set of parameter vectors $\mathbf{P}_u = \{\mathbf{p}_u\}_j$ with corresponding simulations $\{\mathbf{u}(t, \mathbf{v}(t), \mathbf{p}_{u,j})\}_j$. This implies that the set of model parameters from the calibration of the downstream model is a function of the particular parameterization j of the upstream model such that

$$\mathbf{P}_x(j) = \{\mathbf{p}_{x,i}(\mathbf{u}_j)\}_i.$$

The ambiguity originating from the choice of the upstream model calibration j influences the calibration of the downstream model, and this is the CIU.

CIU emerges also when one needs to make a choice for the input data among several sources. As an example, the present study compares inputs from two catchment models of different complexity which give different load estimates as outputs. We are interested in how the difference in loading is realized in final lake model outputs. The downstream lake model calibration is optimized to reproduce lake observation as close as possible, regardless the loading input provided. The calibrated set of parameters of the lake model will thus be different when different input sources are used. Details are discussed in chapter 3.

2.3 Simulation input uncertainty

Input ambiguity becomes even more critical when a simulation (e.g., hindcast, forecast, or hypothetical management or climate scenarios) is run using a chosen calibrated model. The ordinary uncertainty propagation strategy is to include all combinations of uncertainty at each iteration. However, simply combining OPU of the upstream model and OPU of downstream model that is based on a single choice of upstream model run input will misrepresent the actual uncertainty. As discussed, we have multiple equally likely calibrations, each with its own calibration input. Ideally, the corresponding calibration input $\mathbf{u}_j$ or its variation $\mathbf{u}_j^*$ should be used together with the matching calibration $\mathbf{p}_x(\mathbf{u}_j)$ to produce a simulated output $\mathbf{x}$. This demands that many sets of calibrations are run to couple with a variety of inputs. This is unfortunately not practical due to the computational requirement of each calibration. In this study we call this uncertainty stemming from the multiple choices of the input data that feed into a single already calibrated model, the simulation input uncertainty (SIU). Although the parameterizations of the two models do not correspond to each other, studying how the downstream model reacts to varying runs from the upstream model is still useful in order to quantify how the CIU of the upstream model can be substantiated in the output of the downstream model.

Let us select a single calibration of the downstream model $\widehat{\mathbf{p}_x} \in \mathbf{P}_x(j) = \{\mathbf{p}_x(\mathbf{u}_j)\}_i$ for simplicity. This reduces the CIU (a single choice over j) and the OPU (a single choice over i). For each calibration of the upstream model $\mathbf{p}_{u,j} \in \mathbf{P}_u$, there is a corresponding simulation $\mathbf{u}(t, \mathbf{v}(t), \mathbf{p}_{u,j})$ and $\mathbf{x}(t, \mathbf{u}(t, \mathbf{v}(t), \mathbf{p}_{u,j}), \widehat{\mathbf{p}_{x,i}})$

The aim may be to study the effectiveness of a hypothetical management scenario. Depending on the type of changes required in the hypothetical scenario we modify either the parameters $\mathbf{p}_u$ or the time series input $\mathbf{v}(t)$ to implement the scenario. In this paper we take as an example the vegetative filter strip management (VFS) where streams are enveloped with grassland to reduce phosphorus in surface runoff from the agricultural field, described more in detail below. The modelled effect of this can be realized by changing the parameters $\mathbf{p}_{u,j}$ to $\mathbf{p}_{u,j}^*$, where parameters that are not affected by the VFS management option are inherited. This change creates an alternative hypothetical simulations $\mathbf{u}^*(t, \mathbf{v}(t), \mathbf{p}_{u,j}^*)$ and $\mathbf{x}^*(t, \mathbf{u}^*(t, \mathbf{v}(t), \mathbf{p}_{u,j}^*), \widehat{\mathbf{p}_{x,i}})$. If we instead need to run a hypothetical simulation for example under different weather inputs to simulate the impact of climate change, then we create corresponding simulations $\mathbf{u}^*(t, \mathbf{v}^*(t), \mathbf{p}_{u,j})$ and $\mathbf{x}^*(t, \mathbf{u}^*(t, \mathbf{v}^*(t), \mathbf{p}_{u,j}), \widehat{\mathbf{p}_{x,i}})$, where $\mathbf{v}^*(t)$ is the change in forcing inputs. For either situation the SIU underlines that the choice of parameterization $\mathbf{p}_{u,j}$ of the upstream model generates ambiguity in the downstream simulation $\mathbf{x}$.

3 Case study: Nutrient status of lake Vansjø

Here we take as an example case an uncertainty assessment of the eutrophication status of Vansjø in southeast Norway for the period 1993-2010 (Figure 2). The target waterbody for the present case was the western basin (Vanemfjorden), which was monitored for water quality and modelled by MyLake (*Saloranta and Andersen*, 2007). Upstream of this basin, 4 other hydrological elements (Vanemfjorden local catchment, Storefjorden lake basin, Storefjorden local catchment, and river Hobøl, Figure 3) were modelled using catchment models INCA-P (*Wade et al.*, 2002) and SWAT (*Gassman et al.*, 2007), or the lake model MyLake (Table 1). The uncertainties in these upstream models cascaded over to MyLake at the Vanemfjorden basin (Figure 3). This setup thereby examines CIU, OPU and SIU of modelling of the Vanemfjorden basin, described in detail below. See Appendices for the details of the site description, models, available data and calibration methods. We can then denote the status of the Vanemfjorden basin as $\mathbf{x}(t, \mathbf{u}(t, \mathbf{v}(t), \mathbf{p}_{u,j}), \widehat{\mathbf{p}_{x,i}})$, where t is days from 1 January 1993 to 31 December 2010, $\mathbf{u}(t, \mathbf{v}(t), \mathbf{p}_{u,j})$ is the daily input into Vanemfjorden, a function of the choice of the model and the choice of calibration j, and $\widehat{\mathbf{p}_{x,i}}$ is a particular own calibration i. After preparing calibration, we made multiple different model configurations (MCs) to produce multiple final simulations in the Vanemfjorden basin. Various comparisons of these MCs were then made (Table 2), with each comparison scheme having an intention of studying a particular uncertainty.

From the perspective of the Vanemfjorden basin, there are many sources of uncertainty in inputs to the basin that are related to the flux of P and suspended solids (Table 3). The variety of equally likely inputs has implications to CIU and SIU of MyLake simulations at the Vanemfjorden basin. For CIU we made two calibrations, instead of calibrating MyLake for every possible combinations of inputs. These two calibrations are produced using two different inputs that are created for the input from the Vanemfjorden catchment by two different models (SWAT and INCA-P). The models have varying complexity levels and hence require different amounts of effort for setting up. They also diffe a r in capability in implementing certain land management practices. For example, SWAT enables policy related assessment of land management options that is not possible with INCA-P. We restricted SWAT to the Vanemfjorden catchment and used it in a very detailed manner while the rest of the catchment was taken care of by INCA-P, such that we were able to set up the model for the larger catchments to Storefjorden. The motivation of separating the catchment into two groups with different levels of details was also supported by the expectation that the upstream Storefjorden basin most likely functions as a sedimentation basin, which stabilses the Storefjorden-origin particle loading into the Vanemfjorden basin. The lower complexity does note imply that the INCA-P is less accurate because model complexity does not directly translate to accuracy or accountability, primarily due to inflation in the number of model parameters needed to produce plausible simulations. In addition, INCA-P's calibrated parameters for the river Hobøl at the Kure weir (Figure 2) were imported to the Vanemfjorden local catchment to produce a parallel input to SWAT's input. Thus, INCA-P's simple simulation for the input from local

Vanemfjorden catchment provides an alterntiave to a detailed SWAT simulation.

Below, we describe the specifics of the model comparison schemes.

The comparison scheme I studies the OPU of SWAT and how this realizes in the Vanemfjorden basin as the SIU of MyLake. The comparison scheme II studies the same uncertainties for INCA-P connected to MyLake for the same basin. Three variations of each of SWAT and INCA-P were taken to represent their respective parameter uncertainty. These three represents variation in total P loading due to uncertainty in parameterization for each model.

The comparison scheme III was designed to study the uncertainty associated with the use two different models for different groups of catchments. MC 2 represents the "reference" or "best research effort" configuration using SWAT-based input for calibration and the best SWAT parameterization for simulating input for the whole 18-year period. MC 5 uses INCA-P-based input for calibration and the median INCA-P run as input for the whole period, representing the "low research effort" situation. MCs 2 and 5 use their respective MyLake calibrations, and this should therefore reflect the comparison based on different research effort levels. MC 7 is a combination of INCA-P inputs and MyLake's "foreign" calibration using SWAT input, and adds another dimension to the comparison. We expected that the MC 7 to perform poorer during MyLake's calibration period (2005-2010), that the difference between MC 2 and MC 7 be greater than the difference between MC 5 and MC 7 in the earlier hindcast period, and that the difference between MC 5 and MC 7 account for the inconsistency between the input used to calibrate the model and input used to make simulations (namely the inconsistency between $\mathbf{u}_1$ and $\mathbf{u}_2$ in simulation $\mathbf{x}(t, \mathbf{u}_2(t), \mathbf{p}_x(\mathbf{u}_1)))$.

The comparison scheme IV studies the uncertainty associated with the loading input from Sundet, i.e., the water input from the Storefjorden basin. The Storefjorden basin water is used as input to the Vanemfjorden basin, but modelling of Storefjorden basin itself depends on the upstream model INCA-P for loading. From the perspective of studying the SIU of the Vanemfjorden basin, we decided to only represent the OPU of MyLake at the Storefjorden basin, although one could have examined the CIU and SIU of the connection between the "grandparent model" INCA-P at Hobøl and the Storefjorden local catchment and MyLake in the Storefjorden basin. In particular, the INCA-P simulation with median P loading was used to calibrate MyLake in Storefjorden basin, and the the same INCA-P run was used for hindcast to complete the 18-year period.

The comparison schemes V and VI take hypothetical approaches; this makes them model based experiments rather than uncertainty studies. As such, they make a strong contrast to the comparison schemes I to IV, where combinations of MCs were made to highlight a particular uncertainty. The comparison scheme V studies the effectiveness of vegetative filter strip (VFS) in reducing the loading of phosphorus to the lake. See the relevant Appendix for the implementation of VFS in the current study. The comparison scheme VI simulates a hypothetical situation where the two runoff contributions to the basin, namely the Vanemfjorden local catchment and the water flow from the Storefjorden basin at Sundet, are separated and used individually. We designed this experiment in order to elucidate the role each loading source had on the phosphorus loading in a qualitative manner. For example, we expected from available monitoring data that the input from Sundet water often has lower concentration of phosphorus than the local Vanemfjorden runoff, while the volume of water is approximately a magnitude order greater. The Sundet water input should therefore make the water displacement much faster, and is expected to reduce the impact of local runoff events on the basin.

For all these comparisons, MyLake's OPU in the Vanemfjorden basin was taken care of by sampling the same 60 randomly sampled posterior calibration parameter combinations (i.e., $i = 1, 2, ..., 60$), specific to each of the two calibrations: those using SWAT-based input and those using the INCA-P-based input. The median and the 5- and 95-percentiles are taken to aggregate the 60 samples of i.

4 Results

4.1 Calibration of MyLake

Because we assumed a Gaussian normal distribution for the observation error for each observation series, the sum of the squares of error

$$\sum_{k=1}^{n_o} (\widehat{x_{o,k}} - x_{o,k}(\mathbf{p}_x))^2,$$

where $x_{o,k}$ is the k-th observation in the observation series o of length n_o, and $\widehat{x_{o,k}}(\mathrm{p}_x)$ is the corresponding model simulation given parameters p_x, was used during the calibration. For each type of observation o, an error variance σ_o^2 was drawn using Gibbs sampling, see Appendix for details. The formal likelihood was therefore calculated as following:

$$\mathscr{L}(\hat{\mathbf{x}}|\mathbf{p}_x) = \prod_o \prod_{k=1}^{n_o} \frac{1}{\sqrt{2\pi\sigma_o^2}} \times \exp\left(-\frac{(\widehat{x_{o,k}}(\mathbf{p}_x) - x_{o,k})^2}{2\sigma_o^2}\right).$$

For presentation purpose we used instead root mean square error (RMSE) as model performance statistic:

$$\mathrm{RMSE}_o = \sqrt{\frac{1}{n_o} \sum_{k=1}^{n_o} (\widehat{x_{o,k}}(\mathbf{p}_x) - x_{o,k})^2}.$$

RMSE has the same unit as the observed variable, and is thus an indication of the average deviation of the model predictions from the observations (*Janssen and Heuberger*, 1995).

4.1.1 Storefjorden basin

The calibration of MyLake at the Storefjorden basin captured well the seasonal variation of Chlorophyll and phosphate-phosphorus concentration (Figure 4). The model missed the impact of a landslide event early in 2008 that can be seen as high concentrations of total phosphorus and particulate phosphorus (*Skarbøvik and Bechmann*, 2010). This was a response to INCA-P based input that also missed the event. Although INCA-P simulates erosion and mobilization of particles, larger scale erosion events such as landslides are not captured. Furthermore, this event occurred downstream of Kure where observation data used for calibrating INCA-P were collected.

4.1.2 Vanemfjorden basin

Compared with the calibration at the Storefjorden basin, calibration was more challenging for both cases at the Vanemfjorden basin (one using SWAT-based input and the other using INCA-P-based input) (Figures 5 and 6). For example the algal bloom in the summer 2006 was not produced. Stratification that was observed at the deepest part of the basin was not simulated either. However, the simple model structure of MyLake provides plausible explanations for this discrepancy and hence gives suggestion regarding how the next research efforts could be allocated, see Discussion.

The parameter variance structure of the two MyLake calibrations at the Vanemfjorden basin appeared to be very similar, despite the difference in input from SWAT and INCA-P (Figure 7). For both calibrations, parameters p_2 and p_3 (Table 4) that together control the net sinking loss of suspended inorganic particles were positively correlated. The scope of these parameters is slightly different: p_2 applies to all depths but p_3 applies only to epilimnion. p_5 and p_6 were also positively correlated as they have a cancelling effect on each other. p_8 and p_9 were negatively correlated as they have complementing tendencies to each other. These correlations were expected, but these parameters were included for calibration due to prior knowledge that they represent processes that are dependent on different state variables. For example, the value of p_5 affects the impact of algae mortality given a temperature and algae concentration. But p_6 represents algae growth and it affects the impact of light and available P in addition to temperature but not existing algae concentration; algae concentration indirectly reduces the available light however. Given this information, our results illustrate the relative significance of the various parameters. Between the two calibrations at Vanemfjorden basin, the only clear difference in the parameter space was the mode and median for p_2, such that the INCA-P-based input promoted particle resuspension from the sediments more than the SWAT-based input (Figure 8). Despite the lower loading of suspended particles by INCA-P-based input during the calibration period (Figure 9), the calibration method settled to keep suspended particle from quickly sedimenting in order to sustain the phosphorus longer, as less phosphorus is provided by INCA-P-based input. Compared with these two calibrations at the Vanemfjorden basin, the calibration at the Storefjorden basin had a more complex parameter covariance structure (positive correlation between p_2 and p_4 (Pearson product-moment correlation coefficient $r = 0.65$) and negative correlation between p_1 and p_5 ($r = -0.76$)), in addition to the aforementioned correlation pairs (p_2 and p_3 ($r = 0.80$), p_5 and p_6 ($r = 0.89$), p_8 and p_9 ($r = -0.51$)). This elucidates dissimilar dominant processes that govern the physicochemical state of the lake water between the two basins. The complex parameter structure at the Storefjorden basin indicates more intricate process combinations, and this resulted in apparently better model performances, i.e., smaller RMSE values. The model calibrations at the shallower Vanemfjorden basin suggests that the basin is more hydrologically governed, and the basin's fully mixed nature during the summer in which calibration took place left less room for the lake model to take control of in-lake dynamics that depend on depth-gradient.

4.2 Uncertainty analyses: comparisons I-IV

From the management point of view the summer total P concentration is the most operationally relevant variable as it is considered a good proxy for bioavailable phosphorus which is one of the necessary premises for nuisance algae blooms. We therefore review how the difference in P loading due to various uncertainties (Figure 9) are realized in the lake summer total P concentration (Figure 10), although other notable differences will be discusses in addition. In all cases OPU here represent the uncertainty due to MyLake's parameter equifinality in the Vanemfjorden basin.

4.2.1 Comparison I: SIU of SWAT input on MyLake

The best SWAT parameter set was identified by calibration at the Guthus subcatchment (MC 2). We also used two SWAT parameter sets that gave the 97.5- and 2.5-percentile phosphorus loading (MC 1 and MC 3, respectively). MC 2 was close to lower range with respect to total phosphorus loading but was close to higher range with respect to suspended particle loading.

There was little SIU for suspended inorganic particle concentration as all MCs 1, 2 and 3 gave similar summer concentrations for all simulation years. However, OPU was high for suspended inorganic particle concentration in the summer. All other concentrations were predicted to be the highest for MC 1. In case of the summer total phosphorus concentration, there was little overlapping between MC 1 and MC 2, or MC 1 and MC 3. This suggests that the lake simulation may become different depending on the SWAT parameter set. However, such difference could be made less pronounced if MyLake had been calibrated using the MC 1 input or the MC3 input. In some years MC 2 forecast higher concentrations than MC 3 but the difference is small relative to the OPU. This contrast between MC 2 and MC 3 may be due to the differences in the sedimenting particle loading, despite that the model calibration suggested little P sink in the Vanemfjorden basin. The decision of ranking the input simulations according to phosphorus alone might reveal its limitation here because the loading of sedimenting inorganic particles are also important in lake phosphorus simulations in the lake water. However, when different parameters are compared, resulting loading input of phosphorus and that of inorganic particles do not strongly correlate for either model. Therefore, representation of the true runoff variability would have required attentions for both variabilities for total P and sedimenting inorganic particles.

4.2.2 Comparison II: SIU of INCA-P input on MyLake

INCA-P gave little parameter uncertainty after calibration for the Vanemfjorden local catchment, and therefore lake simulations yielded little differences. Because the calibration is different, the size of OPU is different from Comparison I. This is described in the next section

4.2.3 Comparison III: CIU of input model choice of MyLake at the Vanemfjorden basin

For either input model choice (by SWAT or by INCA-P), the MyLake model were calibrated during the period 2005-2010 such that the correspondence between the two calibrations is tighter in this interval than earlier in the simulation run. Overall, SWAT-based-calibration (MC 2) simulated higher total P concentrations before 2005 than INCA-P-based calibration (MC 5). OPU was greater for summer total P concentration for MC 5 than that for MC 2. This might reflect the simpler model structure of INCA-P and the use of externally calibrated parameters from Hobøl. The combination of the SWAT-based calibration together with the INCA-P-based input (MC 7) created the simulation that is furthest from the MC 2 for summer total P concentration. This was expected as the calibration is not native to the simulation input, see “comparison scheme III” in section 3. The size of OPU for MC 7 resembled MC 2, following the uncertainty in calibration, rather than the uncertainty in simulation input.

4.2.4 Comparison IV: SIU of loading from Sundet

The quantiles of total P loading from Sundet (input from the Storefjorden basin) were chosen to represent the SIU from Sundet: median (MC 2), 5-percentile (MC 8) and 95-percentile (MC 9). The ranking of the three according to total P loading (MC9, MC2, and then MC8) did not correspond with the ranking of sedimenting particle loading (MC9, MC8 and then MC2). There was little difference in the summer total P concentration among the MCs, but the other fractions of phosphorus were simulated differently.

4.3 Model experiments: Comparisons V and VI

4.3.1 Comparison V: Effectiveness of VFS under OPU and SIU

Both the OPU (shades in Figure 10) and SIU of SWAT (contrast among MCs 1, 2, and 3, or among MCs 1v, 2v, and 3v) revealed comparable extent of uncertainty. The effectiveness of the VFS implementation was contrasted for summer total P concentration in 2001, which is a year after mild winter flooding that caused high P loading input, and 2010, which is a recent year exhibiting one of the lowest levels of summer total phosophorus concentration. Figure 11 shows the probability distribution of the summer total P concentration when the combinations of the OPU of MyLake ($i = 1 \ldots 60$) and the SIU for SWAT-MyLake connection ($j = 1 \ldots 3$) were combined. According to the simulations, the summer total P concentration should have been reduced upon implementation of VFS in 2001. However, the effect size of VFS is small compared with the uncertainty range given by OPU and SIU. In year 2010, when OPU and SIU were small, reduction of summer total P concentration by VFS implementation appeared minor. A pair-wise comparison illustrates that VFS will never increase lake total P concentration at any time of the year and will always reduce it when it goes above 50 μg P L^{-1} (Figure 12). This means that VFS will systematically improve the basin water quality above a certain threshold total P concentration, but there is a wide uncertainty in the abosolute value in the simulations.

4.3.2 Comparison VI: Relative importance of loading from the Vanemfjorden local catchment and loading from Sundet

Much higher concentrations of total P, chlorophyll, and phosphate-P were simulated for MC 10, which shut out the water flow from the Storefjorden basin at Sundet and only uses the Vanemfjorden local catchment input modelled by SWAT. P loading in MC 10 is lowest among the MCs in comparison here, but the low water flow increased the residence time considerably. This observation suggests that the Vanemfjorden basin will always have a condition that is worse than the Storefjorden basin, because the water from the Storefjorden basin works as a diluting medium for the Vanemfjorden basin (c.f., MC 11).

5 Discussion

The two calibrations of MyLake gave different simulations at the Vanemfjorden basin. The differences were greater during the simulation period (before 2005) than in the calibration period (2005-2010). As described in sections 2.2 and 2.3, the choice of the basis input is crucial, and the importance is further pronounced when the model is used for forecast. In the present study, MC 5 represents the low research effort situation, but it depends on more assumptions than MC 2. First, landuse-type-specific characteristics of two catchments Hobøl and Vanemfjorden local catchment are assumed to be very similar so that export of the parameters from Hobøl is acceptable. The fractions of landuse types are changed accordingly, however. Second, INCA-P is a stream- or river-based model with surrounding land that feeds into the stream. However, the Vanemfjorden local catchment does not have major streams and most runoff is through other channels such as agricultural drainage pipes and surface overflow. Furthermore, there is an island of a considerable size in the middle of the basin that does not have easily visible streams. Therefore, application of INCA-P in the Vanemfjorden local catchment will have to assume that the model will perform similarly well even in a catchment where the runoff is not accumulated into well defined outlet streams or rivers. Ideally, multiple calibrations accounting for variation in input should be performed in order to reduce this uncertainty. The results also suggest the importance of consistency between the calibration basis input and the simulation forecast input. Furthermore, model assumptions need to be communicated explicitely.

Poorer model performances in the Vanemfjorden basin, compared with the calibration in the Storefjorden basin, can be explained from the process-oriented point of view. First, the shallow bathymetry and the short residence time of the Vanemfjorden basin constrained its capacity as a phosphorus sink, which can be clearly contrasted with the Storefjorden basin calibration. The model did not manage to reproduce the thermal stratification in the summer 2010 at depth 8.3 and 9.4 metres. This implies underrepresentation of the net phosphorus sedimentation, but the actual effect of it cannot be very large because only 5 percent of the surface area is deeper than the 2010 thermocline depth. Therefore, as far as the total phosphorus is concerned, the Vanemfjorden basin is primarily controlled by catchment hydrology, and the runoff input almost deterministically translates to the concentrations in the lake. This is in particular reflected in the very small MyLake parameter uncertainty as shown by the narrowness in the gray bands in Figures 5 and 6.

Second, properly calibrating the catchment model to represent the magnitude of each rainfall event is difficult due to low frequency of stream and river samples that availed during the calibration (the Guthus catchment and the Kure observation site on the river Hobøl, see Appendices). Because these samples were taken quasisystematically following the calendar weeks, they can only represent instantaneous snapshots of the catchment loading. Because of this observation stochasticity, calibration of the catchment models using many instantaneous observations may have represented the mean responses to rainfall events, which may be temporally least variable. Thus, the calibrated paramter spaces of these catchment models may emphasize different processes that respond differently to environmental conditions. For example, exaggeration of runoff to a particular rainfall event may be compromized by making another rainfall event less exaggerated by adjusting another process. Such confusion by the catchment model could be improved for example by deploying a flow weighted automatic sampler, which produces a more precise and accurate representation of the actual loading flux.

Third, the ecological reason for an algae bloom in the summer 2006 is still not understood well, and as expected, models did not reproduce the algae bloom. Although the prediction of algae bloom will be conducted using another tool (statistical model based on MyLake prediction of physicochemical status of the lake), the discrepancy implies under representation of this event. The Figures 5 and 6 also shows that the year-to-year variability of the seasonal succession was not large, suggesting that some other processes that are controlling the chlorophyll concentration were not captured.

6 Perspectives

The present study tackles some of the previously overlooked uncertainties when a chain of multiple calibrated models are connected. The example quantifications presented here illustrated that these uncertainties now need a proper treatment. The assessment of the effectiveness of catchment management implementation such as VFS in reducing total P concentration in the lake needs to be contrasted with the full range of various uncertainties in order to provide realistic conclusion. Transparent uncertainty communication is paramount in modelling applications where public policy will depend on the results. Uncertainty quantification such as the one presented in the current study also facilitates connection to probabilistic knowledge aggregation tools such as Bayesian belief networks. Technical analogies can be drawn in any other application field of dynamic modelling.

Appendix A: Site descriptions

The study areas are located in the southeastern part of Norway (Figure 2). The Vansjø-Hobøl catchment area is 690 km^2, and its land use is dominated by agriculture (16 %) and forestry (80 %) (*Bouraoui et al.*, 2009). The agricultural production around the Storefjorden basin consists mostly of grain with a smaller fraction devoted to grass. Most of the catchment lies below 200 m elevation and is covered by marine sediments deposited after the last graciation.

The surface of the lake is situated 25 m above the sea level. Mean annual rain fall is 810 mm and the specific runoff is 14.4 L s^{-1} km^2. The main inlet river to the Storefjorden basin, Hobøl, has a catchment area of 337 km^2, whereas the entire catchment area of the eastern basin totals 580 km^2.

The total area of the Vanemfjorden local catchment is 83.3km^2, in which 13.5 % of the area is agricultural land, 61.8 % is forested, 8.9 % urban and 15.7 % lake and water. The top soils in the agricultural areas of the Vanemfjorden local catchment consist of 73 % clay or loamy clay soils and 22 % sand. Sandy soils in the south-western part of the sub-catchment originate from a huge end moraine which was formed under the last ice age when a colder period made the retreating ice front stagnant for a long time. The bedrock in the Vanemfjorden local catchment is mainly Pre-Cambrian bedrock, predominately gneiss. In the forested areas of the Vanemfjorden local catchment, the hill tops and slopes have thin sandy beach deposits or thin humic cover, while the valley bottoms mainly have thick clay marine deposits. About 3.2 % of the forested area is comprised of bogs or wetlands. The areas dominated by clay loam soils are used mostly for cereal production, whereas the sandy end moraine has a dominance of potato and vegetable production in addition to cereal production. Animal production is limited in the catchment. The agricultural areas are systematically tile drained and most of the subsurface drainage installed in the 50s and the 60s.

Due to the low prices of P fertilizers between the 1950s and 1990s and their positive effect on crop production, an increased usage of P fertilizers has resulted in P acculation in agricultural soils in most of the Western countries. In Norway this practice has led to an increase in biologically available and mobile P in cultivated soils during the past few decades (*Øgaard*, 1995). The buildup of large P-pools in agricultural soils increases the export of P to receivingt waters (*Sims and Sharpley*, 2005). Abatement measures have been implemented in the Vanemfjorden catchment. Most of these measures were implemented in 2007: no autumn tillage in areas with high erosion risk, installation of sedimentation ponds in selected sub-catchments, vegetative filter strips along open water and a 47 % reduction in P fertilization (*Skarbøvik and Bechmann*, 2010).

Lake Vansjø (59.42° N, 10.86° E) has a total area of 37 km^2 and has complex morphology with several distinct basins. In the current paper we divided the lake into two model unit basins, Storefjorden and Vanemfjorden (Figure 3). The water level of the lake is managed at a dam approximately 4 km downstream the outflowing river named Mosselva. We assume that the channel connecting the Storefjorden basin and the Vanemfjorden basin, referred to here as Sundet, has a unidirectional flow, such that water in the Storefjorden basin always flows into the Vanemfjorden basin. The Storefjorden basin (25 km^2) has a maximum depth of 41 m and a mean depth of 10 m, where as the Vanemfjorden basin (12 km^2) has a maximum depth of 19 and a mean depth of 5 m. The theoretical water residence time for the whole lake is 277 days, while the basin specific residence times are 242 days for the Storefjorden basin, and 41 days for the Vanemfjorden basin. Other bathymetric and morphological information has been published elsewhere (*Saloranta*, 2006). Vansjø, in particular the Vanemfjorden basin, is eutrophic and has experienced nuisance cyanobacteria blooms over the past decades. Because the lake and the downstream Mosselva has significant utility values (driniking water supply, bathing, fishing, recreation), toxic algal blooms are considered a critical pollution issue in the watershed.

Appendix B: Models and associated data

INCA-P

INCA-P is a process-based mass balance model designed for simulating the P dynamics in catchments through the accumulation and export of P in the plant and soil systems of different land use types (*Wade et al.*, 2002). The water containing both P and suspended particles is then routed downstream in the catchment while also including effluent discharges and in-stream sedimentation. The input fluxes (farmyard manure, fertilizers, livestock wastes etc) and P addition and removal processes are differentiated by land use type and varied according to environmental conditions (e.g. soil moisture and temperature). The model also accounts for accumulated pools of inorganic and organic P in the soil (in readily available and firmly bound forms), groundwater and streams.

INCA-P requires daily time series of hydrological forcing, with hydrologically effective rainfall and soil moisture deficit. These timeseries were generated using a newly developed hydrological model PERSiST, set up for the catchment using daily flow at the gauging station 3.22.0.1000.1 Høgfoss obtained from the Norwegian Water Resources and Energy Directorate (NVE). Water chemistry data come from the MORSA monitoring programme, conducted by Norwegian Institute for Agricultural and Environmental Research (Bioforsk) and Norwegian Institute for Water Research (NIVA). Total phosphorus and suspended sediment data from the monitoring station Kure were used for calibration of INCA-P. General land cover data are obtained from the Norwegian Forest and Landscape Research Institute, whereas more detailed information about land use fertilization regimes on agricultural fields are provided by Bioforsk. Nutrient outputs from sewage treatment plants are obtained from Statistics Norway and the database KOSTRA. Outputs from scattered dwellings are provided by Bioforsk's information system GISavløp.

INCA-P simulates the daily dynamics of phosphorus mobilization and transport in the catchment and was set up to predict daily discharge of water, phosphorus, suspended sediment in addition to temperature of the discharged water as input into the lake. Hobøl river was set up with 5 separate hydrological reaches (river sections with associated catchment drainage area), while using observed water chemistry at the 4th reach for calibration. Separate simulations for the Storefjorden local catchment and the Vanemfjorden local catchment were also set up assuming parameterization similar to the upper parts of the Hobøl river. INCA-P includes a large number of parameters (ca. 840 for this setup, much less with same parameter values for different reaches and landuse classes). A selection of sensitive or uncertain parameters mainly involved in partitioning and mobilization of phosphorus were estimated using Bayesian inference and inverse modelling, similar to the methods presented here for MyLake, see below.

SWAT

The Soil and Water Assessment Tool (SWAT) is a semi-distributed watershed-scale model which runs in continuous time, on a daily time step (*Gassman et al.*, 2007; *Arnold et al.*, 1998). SWAT is based on physical processes and is developed to quantify the impact of land management practices in large, complex watersheds. Information about weather, soil properties, topography, vegetation, and land management practices in the watershed are required to run the model. The physical processes associated with water movement, sediment movement, crop growth, and nutrient cycling are directly modeled by SWAT using these input data. The SWAT model is widely used to evaluate cultivate practices at watershed scale and several hundred peer reviewed articles have been published of different SWAT applications since the early 1990s. The present study utilizes SWAT's ability to simulate improvements due to Vegetative Filter Strips (VFS) (*White and Arnold*, 2009). The VFS model in SWAT is able to predict sediment, nitrogen and phosphorus retention under uniform sheet flow conditions as well as taking into account concentrated flow conditions. The VFS routines were implemented at hydrological response unit level in SWAT.

Calibration of the SWAT model was performed using the SWAT Calibration and Uncertainty Program (SWAT-CUP) with the Sequential Uncertainty Fitting (SUFI-2) procedure (*Abbaspour et al.*, 1997, 2007). In SUFI-2, uncertainty of parameter distributions are depicted as uniform distributions while output uncertainty is depicted by percentile aggregations of output variables through a stage-based repeated Monte Carlo simulation using Latin hypercube sampling. In this study, 500 Latin hypercube sampled parameter sets were obtained (i.e. the model was run 500 times per SUFI-2 iteration). SUFI-2 repeatedly performs Monte Carlo simulations first by assuming a large parameter range, and narrowing the range at each stage until the percentile aggregations were less than the sample standard deviation of the observed data and signficant correlations were obtained. The best parameter set was determined by the greatest sum of weighted Nash-Sutcliffe model efficiency coefficients for three observation variables, among the sampled parameter sets in the last iteration of Monte Carlo simulation:

$$\sum_{o=1}^{3}\left(w_o\left(1-\frac{\sum_{k=1}^{n_o}(u_{o,k}-\widehat{u_{o,k}}(\mathbf{p}_u))^2}{\sum_{k=1}^{n_o}(u_{o,k}-\overline{u_o})^2}\right)\right),$$

where $\overline{u_o}$ is the sample mean of the observation series u_o, and weights u_o of value 0.7 for flow volume and suspended inorganic particle concentration and the value 1.0 for total phosphorus concentration.

MyLake

The MyLake (Multi-year Lake) model is a one-dimensional process-oriented model of daily vertical distribution of heat and basic phosphorus dynamics (*Saloranta and Andersen*, 2007). It also simulates a simple sediment-water interaction and ice and snow cover of the lake. The ability of the lake model to simulate ice cover has been tested (*Dibike et al.*, 2012, 2011; *Saloranta et al.*, 2009) and this made it the strongest candidate among many other water quality models (*Mooij et al.*, 2010) for application at lake Vansjø, which is winter-freezing and dimictic. Unlike many other lake models, MyLake does not consider food chain dynamics, or competition among algal species. MyLake takes a minimaslistic approach in model design, so that important physical laws and only selected aggregated biological and chemical processes are included. We consider that this makes the model accountable for model deviation when the modelled value and the target observation values do not coincide.

The model requires daily weather variables and runoff inputs to the lake. Weather observation data of the Norwegian Meteorological Institute at the Rygge Airport (59.38° N, 10.79° E) located between the two basins were used as the common atmospheric forcing throughout the study. Missing data for cloud cover and wind speed were complemented using historical monthly means. Other missing data, which were not frequent, were linearly interpolated between the available observations. Runoff data into the Storefjorden basin, excluding preciptation on the lake surface itself, were provided by the INCA-P model. In the Vanemfjorden basin, runoff data from the upstream Storefjorden through Sundet were provided by MyLake, whereas runoff from the Vanemfjorden local catchment were provided by INCA-P and SWAT. We assumed that Sundet allows water flow approximately down to 4 m depth with linearly diminishing flow for each metre below the surface. For the lake as a whole, the vertical dimension was discretized into 1-m layers. Within these layers, water quality determinands were considered uniform across the horizontal spread.

According to an earlier chemical analysis of sedimenting minerals of Vansjø in the sediment, 1 g of suspended inorganic particles contains 0.7 mg of phosphorus. It gives nevertheless a minimum estimate of the P content of sedimenting inorganic particles, and thus also the P burial rate. One can compare 0.07 % P of dry weight on inorganic particle with up to 1.25 % P of dry weight in organic particles, assuming the C:P ratio of 40:1 by weight and 50 % C of dry weight. In order to take this into account, the runoff input of a particular day that receives an excess amount of the suspended inorganic particles more than the maximum amount implied by the ratio (0.7 mg P in 1 g suspended inorganic particles) is removed from the input. This is expected to occur during an event of high discharge in which most of large inorganic particles with low surface to volume ratio and correspondingly low P adsorption capacity enters the lake. When conducting the whole 18-year simulation (1993-2010), the year 1993 was repeated three times before starting the simulation from 1993, in order to spin up the model.

10 MyLake parameters (Table 4) were calibrated at each basin. In Storefjorden, we in addition calibrated a temporarily constant value to what the model regards as “dissolved organic phosphorus (DOP)” concentration. DOP is the fraction of allochthonous phosphorus species of unknown or variable identity, which were assumed not to be involved in transformations of P within the lake. Since little is known about the dynamics of this fraction of P, the assumption of DOP as a conservative substance is purely operational and therefore only provisional. The same concentration of DOP as applied for the Storefjorden simulation was used as input for the Vanemfjorden basin. Finally, MyLake assumes that 1 g of chlorophyll a is produded for every g P retained by algae. This relationship has been empirically noted for many decades (*Sakamoto*, 1966; *Vollenweider*, 1976), but it also reflects MyLake’s principle of keeping the model simple; its design choice excluded representing light acclimation of chlorophyll content or luxury consumptionof P, for example. As described earlier, two calibrations were produced for the Vanemfjorden basin: one using the best SWAT simulation together with the median Storefjorden simulation as runoff input, and the other using the median INCA-P simulation together with the median Storefjorden simulation. Calibration followed a variation of Markov chain Monte Carlo methods that use multiple chains (DREAM, Differential Evolution Adaptive Metropolis) (*Vrugt et al.*, 2009). There were 16 chains in a calibration, and 200 iterations (Storefjorden basin) or 400 iterations (Vanemfjorden basin with either SWAT-based input or INCA-P-based inputs), where the first 100 or 200 iterations were considered as the burn-in and excluded from the posterior parameter space, respectively. This resulted in 1600 parameter set samples for the Storefjorden basin and 3200×2 parameter sets for the Vanemfjorden basin (3200 each for SWAT-based and INCA-P based) although the number of unique parameter combinations were less due to chain stagnation. In order to calculate the likelihood of a particular parameter set, we assumed that the model error (predicted minus observed) is distributed approximately according to the error distribution of the previous iteration (Gibbs sampling, see e.g., (*Gelman et al.*, 2003; *Vrugt et al.*,

2009)).

The MyLake model produces daily depth profiles of temperature and phosphorus concentrations. In-lake observed data were used to compare against the model simulations. Temperature measurements were available at 7 depths (Vanemfjorden) and 8 depths (Storefjorden) at every 30 min for May through August 2010. A preliminary assessment indicated that diurnal temperature fluctuations appeared to be mostly less than 1 degree and occasionally up to 2 degrees. Hence, daily mean temperature was used (n = 99). Chemical measurements were available for many of the 18 years of simulation, but in order to keep consistent measurement quality, only data from the period 2005-2010 (analysed at the Norwegian Institute for Water Research) were used for the calibration purpose. We noticed during the calibration that a model run may seem to capture the observed time series after 2005, or the one before 2005, but not both at the same time. We reckon that this is probably due to systematic bias between two data series, especially at the low concentrations. From the 2005-2010 data set, fortnightly observation during the most of the ice-free period for 5 determinands (total phosphorus, total particular phosphorus, chlorophyll a, PO_4-phosphorus, and suspended inorganic particles) with varying sample sizes ranging from 84 to 153 for the period 2005 through 2010 were used for calibration. These measurements were based on composite samples from 0-4 m from the surface at the deepest part of each basin (shown in Figure 2 as VAN1, and VAN2).

What is often known as model validation was not conducted for MyLake on the basis that it is not useful for validating the model, see also *Oreskes et al.* (1994). Model validation is a two iteration calculation where data available for calibration are split into two parts, one of them is used for calibration, and the other for post-calibration comparison. We reckon that the model validation methodology leaves too much room for arbitrary judgement for the lake model, and the methodology was in particular not relevant for the present study.

Acknowledgments

This work has been supported by the EUTROPIA project (Reseach Council of Norway), the EU FP7 project REFRESH, and the Strategic Institute Initiative funded by the Norwegian Institute for Water Research, and the Research Council of Norway. We thank Tuomo Saloranta for help with model customization, and Anne Bjrkenes Christiansen, Stein Beldring, and Deborah Lawrence for their help with hydrology modelling. We thank all people at the Norwegian Institute for Agricultural and Environmental Research, the Norweigian Institute for Water Research, the Norwegian Water Resources and Energy Irectorate, and the Department of Chemistry and Department of Biology at the University of Oslo who collected and analysed the monitoring samples from the lake and the streams.

References

Abbaspour, K. C., M. T. van Genuchten, R. Schulin, and E. Schläppi (1997), A sequential uncertainty domain inverse procedure for estimating subsurface flow and transport parameters, *Water Resour. Res.*, *33*(8), 1879–1892.

Abbaspour, K. C., J. Yang, I. Maximov, R. Siber, K. Bogner, J. Mieleitner, J. Zobrist, and R. Srinivasan (2007), Modelling hydrology and water quality in the pre-alpine/alpine thur watershed using SWAT, *J. Hydrol.*, *333*(2-4), 413–430.

Andrieu, C., N. de Freitas, A. Doucet, and M. I. Jordan (2003), An introduction to MCMC for machine learning, *Mach. Learn.*, *50*(1), 5–43.

Arnold, J. G., R. Srinivasan, R. S. Muttiah, and J. R. Williams (1998), Large area hydrologic modeling and assessment part i: Model development, *JAWRA J. Am. Water Resour. As.*, *34*(1), 73–89, doi:10.1111/j.1752-1688.1998.tb05961.x.

Barton, D. N., S. Kuikka, O. Varis, L. Uusitalo, H. J. Henriksen, M. Borsuk, A. de la Hera, R. Farmani, S. Johnson, and J. D. Linnell (2012), Bayesian networks in environmental and resource management, *Integrated Environ. Assess. Manag.*, *8*(3), 418–429.

Beven, K., and A. Binley (1992), The future of distributed models: Model calibration and uncertainty prediction, *Hydrol. Process.*, *6*(3), 279–298.

Beven, K., and J. Freer (2001), Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology, *J. Hydrol.*, *249*(1–4), 11–29.

Bouraoui, F., B. Grizzetti, G. Adelsköld, H. Behrendt, I. d. Miguel, M. Silgram, S. Gómez, K. Granlund, L. Hoffmann, B. Kronvang, S. Kværnø, A. Lázár, M. Mimikou, G. Passarella, P. Panagos, H. Reisser, B. Schwarzl, C. Siderius, A. S. Sileika, A. a. M. F. R. Smit, R. Sugrue, M. VanLiedekerke, and J. Zaloudik (2009), Basin characteristics and nutrient losses: the EUROHARP catchment network perspective, *J. Environ. Monit.*, *11*(3), 515–525.

Dibike, Y., T. Prowse, T. Saloranta, and R. Ahmed (2011), Response of northern hemisphere lake-ice cover and lake-water thermal structure patterns to a changing climate, *Hydrol. Process.*, *25*(19), 2942–2953.

Dibike, Y., T. Prowse, B. Bonsal, L. d. Rham, and T. Saloranta (2012), Simulation of north american lake-ice cover characteristics under contemporary and future climate conditions, *Int. J. Climatol.*, *32*(5), 695–709.

Gassman, P. W., M. R. Reyes, C. H. Green, and J. G. Arnold (2007), *The Soil And Water Assessment Tool: Historical Development, Applications, And Future Research Directions*, Center for Agricultural and Rural Development, Iowa State University.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin (2003), *Bayesian Data Analysis, Second Edition*, 2 ed., Chapman and Hall/CRC.

Janssen, P., and P. Heuberger (1995), Calibration of process-oriented models, *Ecol. Model.*, *83*(1–2), 55–66.

Mooij, W. M., D. Trolle, E. Jeppesen, G. Arhonditsis, P. V. Belolipetsky, D. B. R. Chitamwebwa, A. G. Degermendzhy, D. L. DeAngelis, L. N. D. S. Domis, A. S. Downing, J. A. Elliott, C. R. Fragoso, U. Gaedke, S. N. Genova, R. D. Gulati, L. Hakanson, D. P. Hamilton, M. R. Hipsey, J. 't Hoen, S. Huelsmann, F. H. Los, V. Makler-Pick, T. Petzoldt, I. G. Prokopkin, K. Rinke, S. A. Schep, K. Tominaga, A. A. Van Dam, E. H. Van Nes, S. A. Wells, and J. H. Janse (2010), Challenges and opportunities for integrating lake ecosystem modelling approaches, *Aquat. Ecol.*, *44*(3), 633–667.

Øgaard, A. F. (1995), Effect of phosphorus fertilization and content of plant-available phosphorus (p-AL) on algal-available phosphorus in soils, *Acta Agr. Scand. B-S. P.*, *45*(4), 242–250.

Oreskes, N., K. Shrader-Frechette, and K. Belitz (1994), Verification, validation, and confirmation of numerical models in the earth sciences, *Science*, *263*, 641–646.

Sakamoto, M. (1966), Primary production by phytoplankton community in some japanese lakes and its dependence on lake depth, *Arch. Hydrobiol.*, *62*, 1–28.

Saloranta, T. M. (2006), Highlighting the model code selection and application process in policy-relevant water quality modelling, *Ecol. Model.*, *194*(1–3), 316–327.

Saloranta, T. M., and T. Andersen (2007), MyLake - a multi-year lake simulation model code suitable for uncertainty and sensitivity analysis simulations, *Ecol. Model.*, *207*, 45–60, 1.

Saloranta, T. M., M. Forsius, M. Jarvinen, and L. Arvola (2009), Impacts of projected climate change on the thermodynamics of a shallow and a deep lake in finland: model simulations and bayesian uncertainty analysis, *Hydrol. Res.*, *40*, 234–248, 2-3.

Sims, J. T., and A. N. Sharpley (2005), *Phosphorus: Agriculture and the Environment*, illustrated edition ed., American Society of Agronomy-Crop Science Society of America-Soil Science Society of America.

Skarbøvik, E., and M. Bechmann (2010), Some characteristics of the Vansjø-Hobøl (Morsa) Catchment, *Bioforsk Report 128, vol. 5*, Bioforsk Soil and Environment, Ås, Norway.

Solomon, S., D. Qin, M. Manning, M. Marquis, K. Averyt, M. M. B. Tignor, H. L. Miller, and C. Zhenlin (2007), *Climate Change 2007: The Physical Science Basis*, Intergovernmental Panel on Climate Change.

Tarantola, A. (2006), Popper, bayes and the inverse problem, *Nat. Phys.*, *2*(8), 492–494.

Vollenweider, R. A. (1976), Advances in defining critical loading levels for phosphorus in lake eutrophication, *Mem. Ist. Ital. Idrobiol.*, *33*, 53–83.

Vrugt, J. A., C. J. F. ter Braak, C. G. H. Diks, B. A. Robinson, J. M. Hyman, and D. Higdon (2009), Accelerating markov chain monte carlo simulation by differential evolution with self-adaptive randomized subspace sampling, *J. Nonlin. Sci. Num.*, *10*(3), 273–290.

Wade, A. J., P. G. Whitehead, and D. Butterfield (2002), The integrated catchments model of phosphorus dynamics (INCA-P), a new approach for multiple source assessment in heterogeneous river systems: model structure and equations, *Hydrol. Earth Syst. Sc.*, *6*, 583–606, 3.

White, M. J., and J. G. Arnold (2009), Development of a simplistic vegetative filter strip model for sediment and nutrient retention at the field scale, *Hydrol. Process.*, *23*(11), 1602–1616.

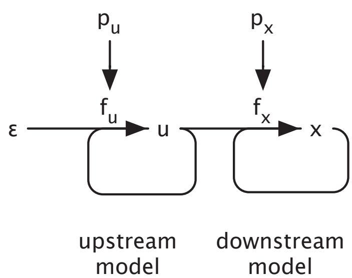


Figure 1: Diagram showing how the mathematical notations are linked to each other.

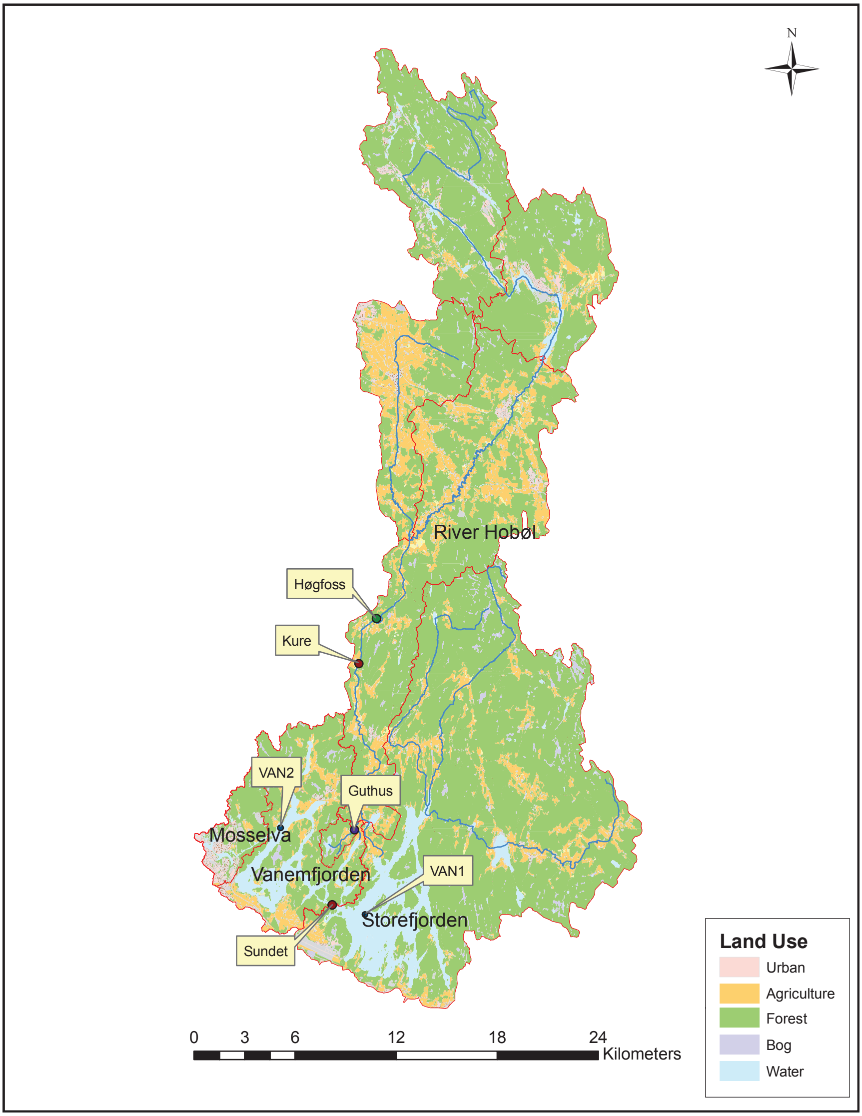


Figure 2: Map showing the relevant hydrological elements of the Vansjø watershed in the case study. Red lines are the deliniations of the catchment domains used by the catchment models. Lake Vansjø is shown near the bottom of the map and includes three sub-waterbodies with the same water level (Mosselva, Vanemfjorden and Storefjorden), although it could be divided in more or less arbitrary choices of basin demarcation schemes.

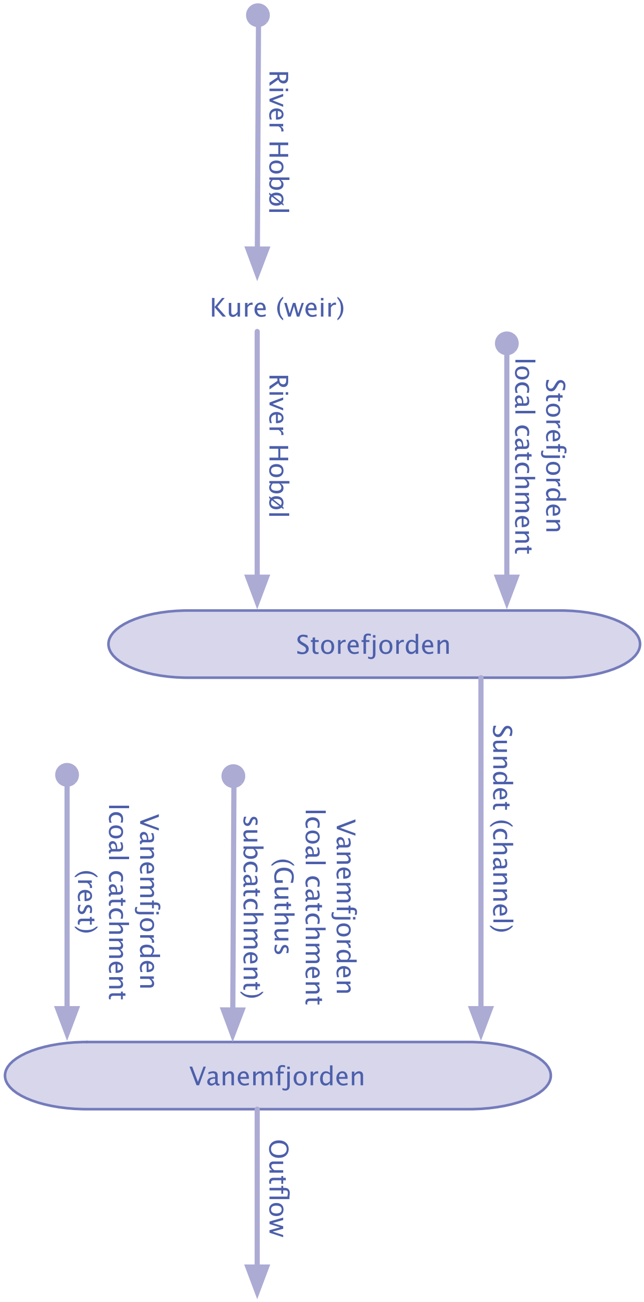


Figure 3: Diagram showing the hydrological connection of relevant waterbodies in the case study.

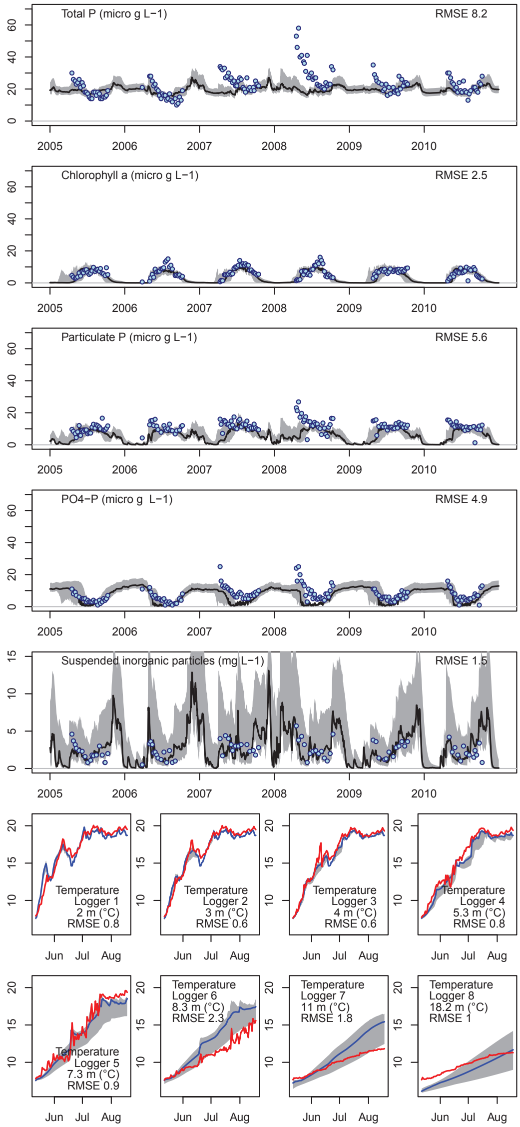


Figure 4: Calibration performance of MyLake at the Storefjorden basin using INCA-P input, for all 13 variables that the formal likelihood calculations were based on. The posterior parameters were resampled (n = 60) and daily quantile statistics (median, shown as solid black line, and 10- and 90-percentiles, shown as gray shaded area) together with observation data shown in blue circles or blue daily line are shown.

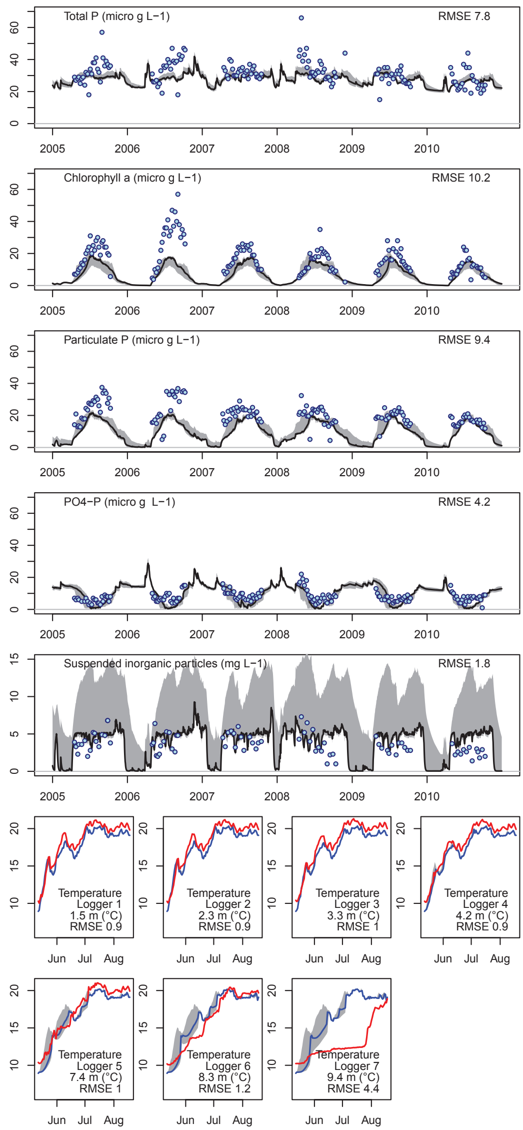


Figure 5: Calibration performance of MyLake at the Vanemfjorden basin using SWAT input, for all 12 variables that the formal likelihood calculations were based on. The posterior parameters were resampled (n = 60) and daily quantile statistics (median, shown as solid black line, and 10- and 90-percentiles, shown as gray shaded area) together with observation data shown in blue circles or blue daily line are shown.

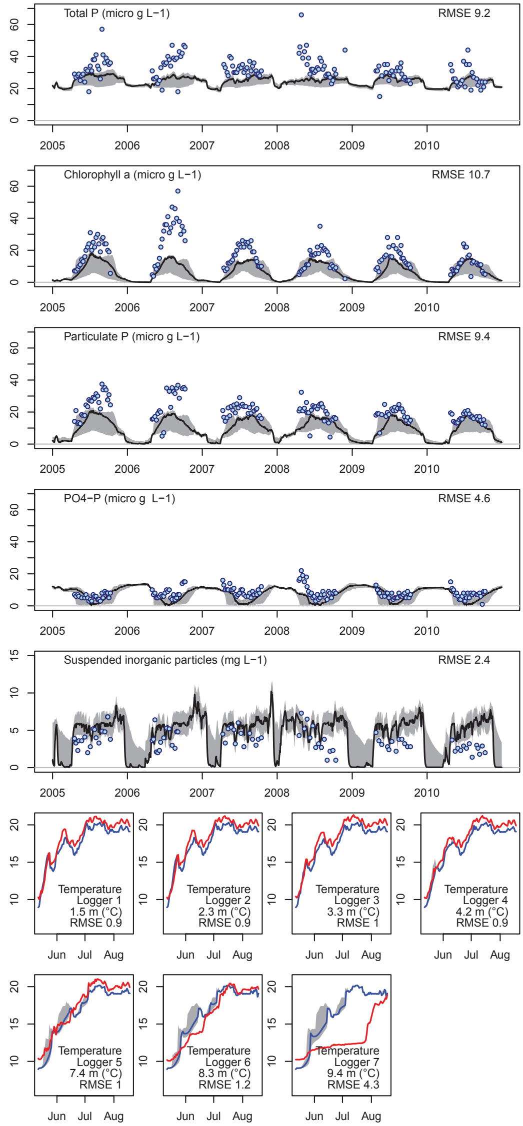


Figure 6: Calibration performance of MyLake at the Vanemfjorden basin using INCA-P input, for all 12 variables that the formal likelihood calculations were based on. The posterior parameters were resampled (n = 60) and daily quantile statistics (median, shown as solid black line, and 10- and 90-percentiles, shown as gray shaded area) together with observation data shown in blue circles or blue daily line are shown.

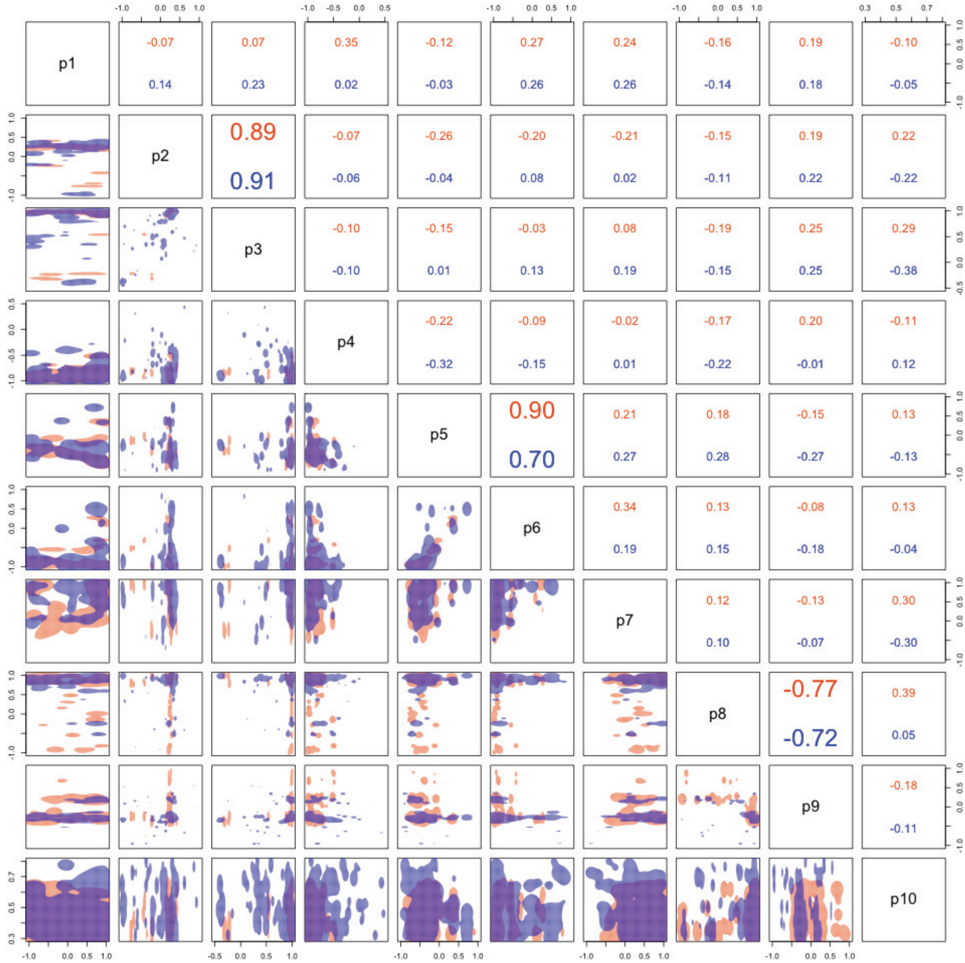


Figure 7: Parameter covariance structure of two calibrations at the Vanemfjorden basin: calibration using SWAT-based input (orange), and INCA-P-based input (blue). Pair-wise correlation coefficients of the parmaeters are shown for both calibration in the top-right triange of the figure. Posterior contour lines are based on two-dimensional kernel densities. Only one contour line that represents the probability density that is one-fifth of the prior density (two-dimensional uniform distribution) is shown.

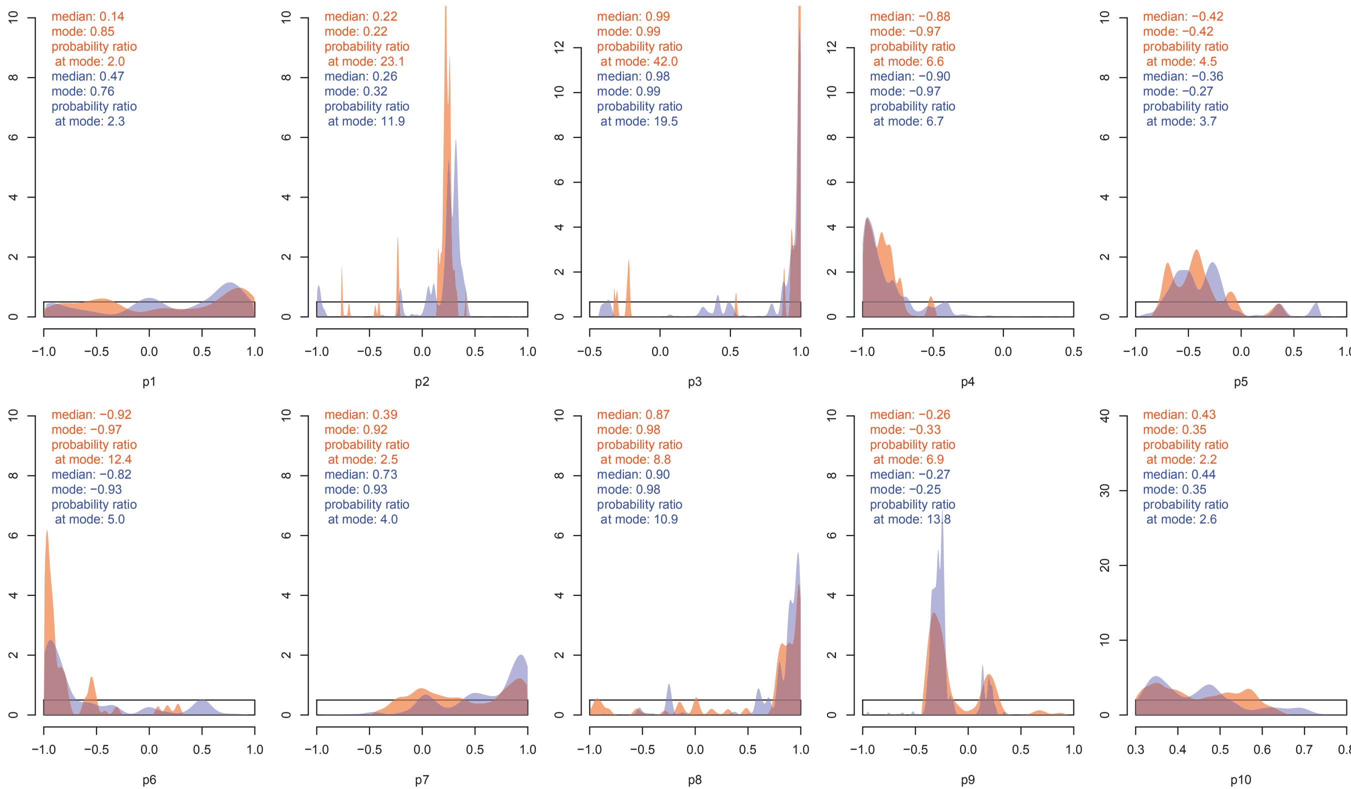


Figure 8: Parameter density distributions of two calibrations at Vanemfjorden: calibration using SWAT-based input (orange), and INCA-P-based input (blue). The black box represents the prior probability density. The colour of the in-plot statistics correspond with the colour of the probability density function.

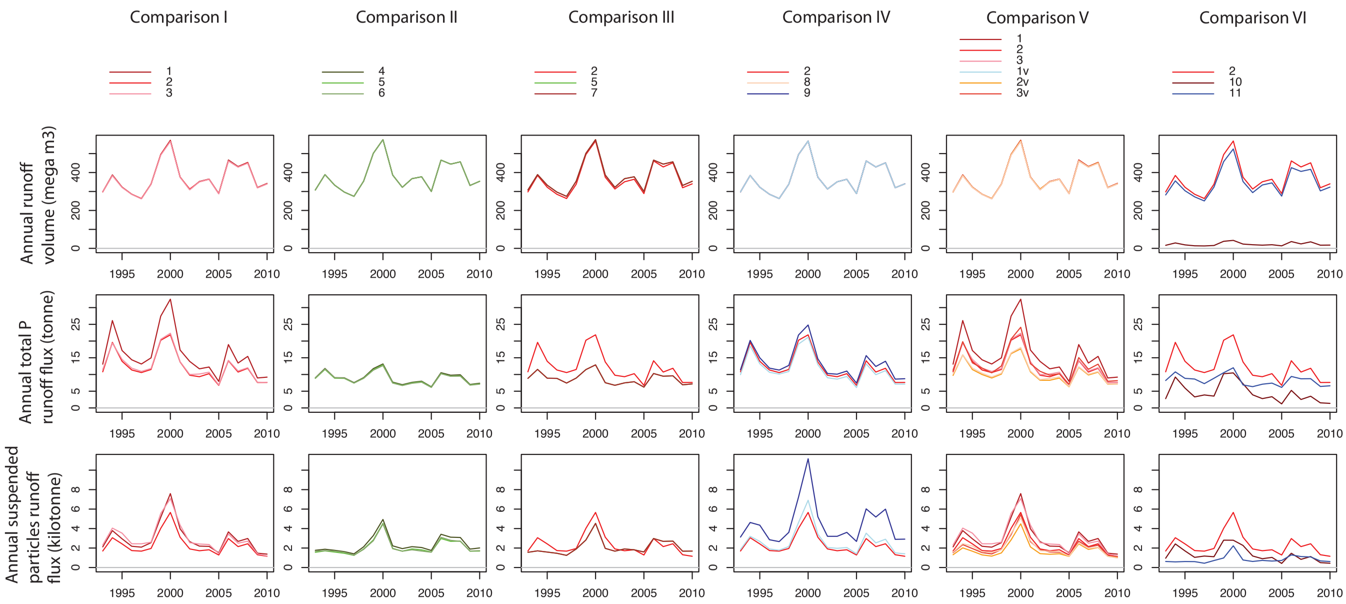


Figure 9: Annual aggregated input fluxes for three main variables that were taken from output of upstream model (INCA-P, SWAT or MyLake) to be used as input to MyLake for Vanemfjorden. Total of 14 different input combinations or model configurations (MCs) are shown by comparison schemes in order to clearly show the intended difference in model simulation input.

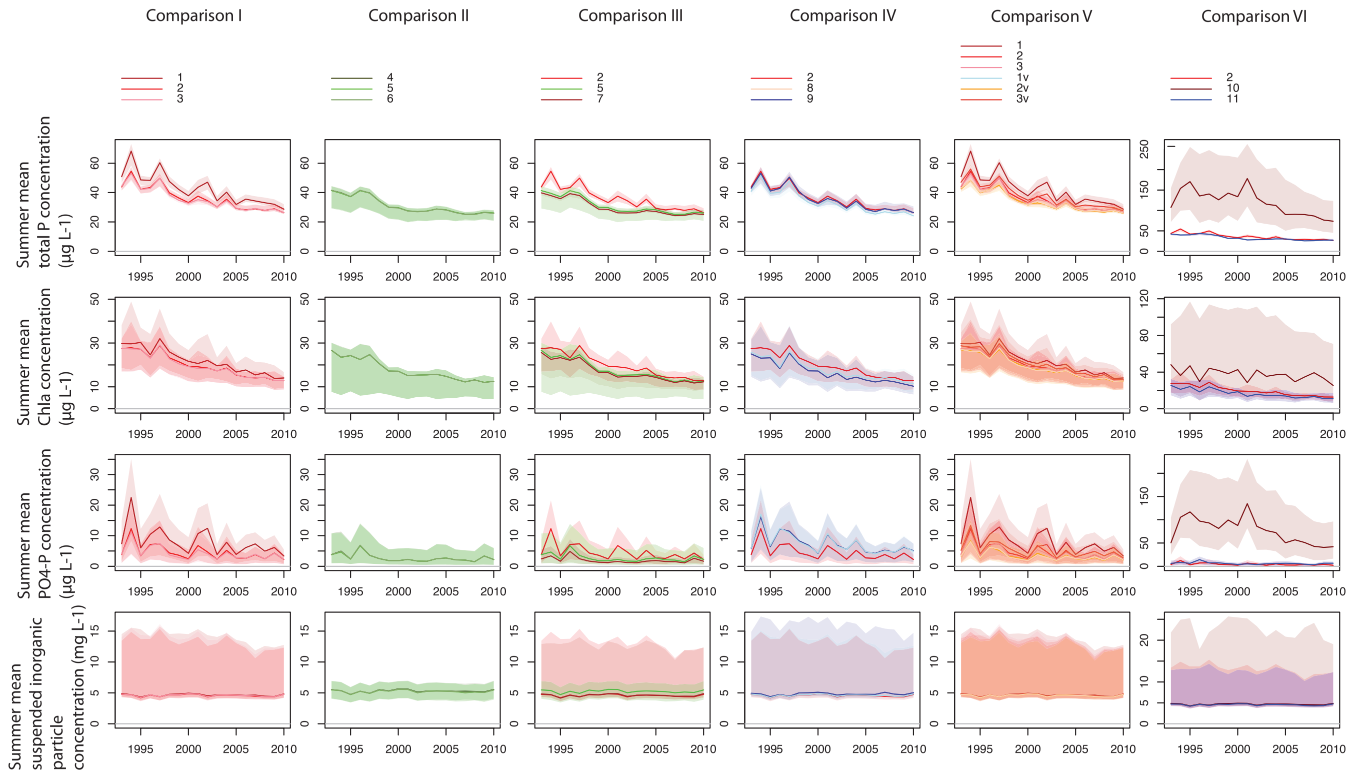


Figure 10: Comparison of output of MyLake at the Vanemfjorden basin using total of 14 model configurations (MCs). The solid lines are MyLake's median around the OPU. The shades are the 5- and 95-percentiles of the same OPU are shown by shaded area of the same colour. Summer mean concentration here is defined as the mean of daily values over the three months (JJA).

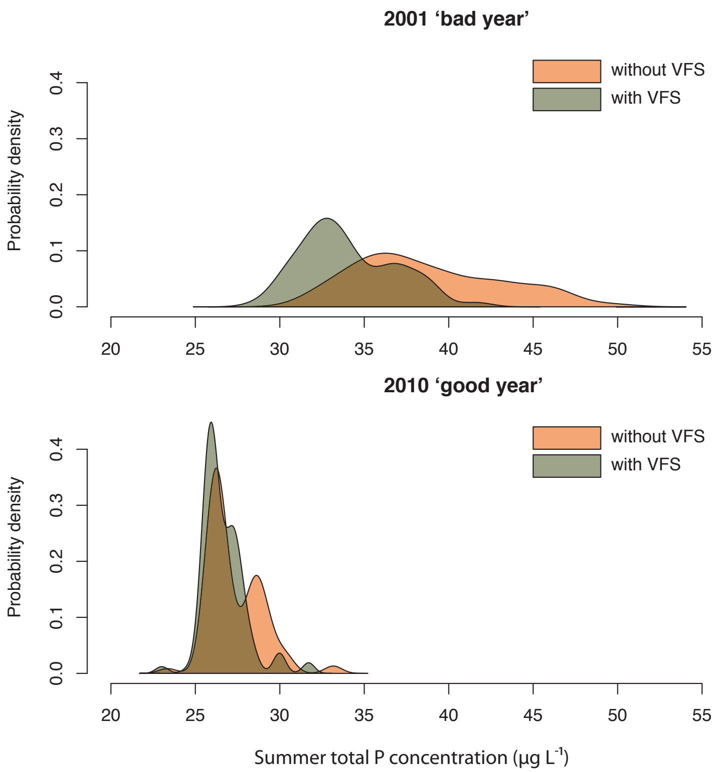


Figure 11: Effectiveness of VFS as realized in the mean summer (JJA) total P concentration (μgL^{-1}) in the Vanemfjorden basin for 2001 and 2010.

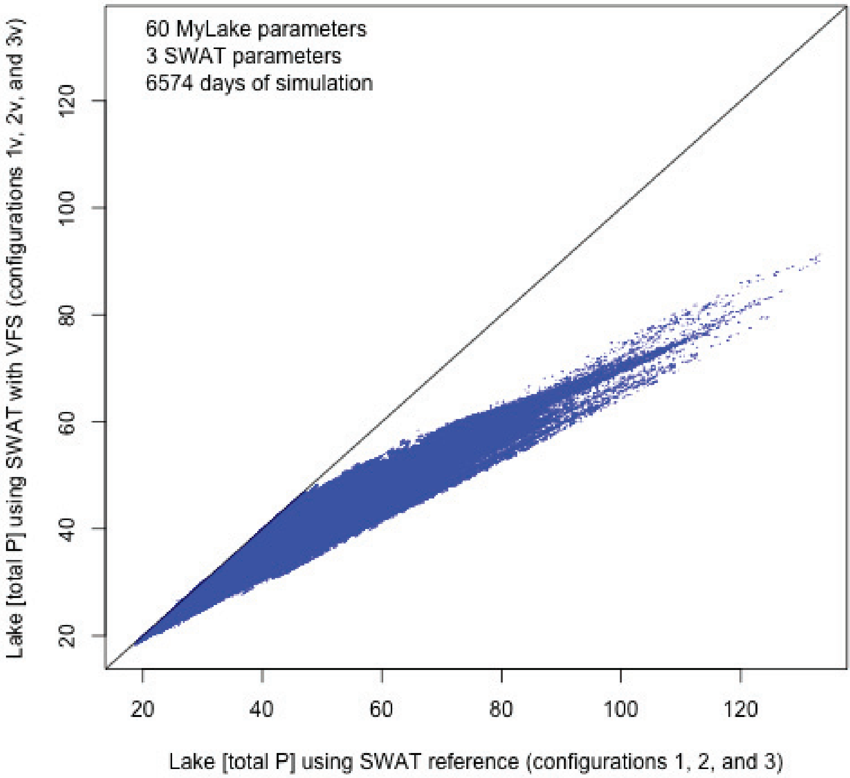


Figure 12: Scatterplot comparing daily total P concentration in the Vanemfjorden between SWAT inputs with VFS and SWAT input without VFS for 60 MyLake parameter variations (OPU) and 3 SWAT parameter variations (SIU).

Table 1: Summary of calibration configurations listed by waterbodies.

Model	Calibrated at	Observation type (period) used for calibration	Forcing data for calibration	Number of calibrated parameters	Calibration method	Waterbody the calibrated model was applied to	Simulatin period including hindcast/forecast
INCA-P	Weir at Kure	Hydrological (1993-1994)	Weather	28	DREAM-MCMC	Upper and lower River Hobl Storefjorden local catchment	1993-2010
		Chemical (1993-1994)	Calibrated PERSIST Land management				
MyLake	Storefjorden	Physical (2010)	Weather	8 chemical and 3 physical	DREAM-MCMC	Lake Storefjorden	1993-2010
		Chemical (2005-2010)	Calibrated INCA-P				
SWAT	Guthus subcatchment	Hydrological (2006-2008)**	Weather	60	MCMC	Whole Vanemfjorden catchment	1993-2010
		Chemical (2006-2008)**	Land management history				
MyLake	Vanemfjorden	Physical (2010)	Weather	7 chemical and 3 physical	DREAM-MCMC	Lake Vanemfjorden	1993-2010
		Chemical (2005-2010)	Calibrated SWAT* Calibrated MYLAKE at Storefjorden (taken top 4 metres)				

* an alternative is to extrapolate INCA-P to Vanemfjorden local catchment
** 'model validation' conducted for data 2008-2010

Table 2: Summery of model configurations (MCs) for MyLake at the Vanemfjorden calibration and contrasts among the MCs. Calibration A refers to calibration using the best SWAT input together with median Sundet whereas Calibration B refers to calibration using INCA-P's median together with median Sundet as the input.

				Uncertainty assessment				Experiment	
Model configuration	Sundet alternative	Vanemfjorden local catchment alternative	Vanemfjorden MyLake basis calibration inputs	I	II	III	IV	V	VI
1	Median	SWAT's 97.5-percentile	Calibration A	x				x	
2	Median	SWAT's best	Calibration A	x		x	x	x	x
3	Median	SWAT's 2.5-percentile	Calibration A	x				x	
4	Median	INCA-P's 95-percentile	Calibration B		x				
5	Median	INCA-P's median	Calibration B		x	x			
6	Median	INCA-P's 5-percentile	Calibration B		x				
7	Median	INCA-P's median	Calibration A			x			
8	5-percentile	SWAT's best	Calibration A				x		
9	95-percentile	SWAT's best	Calibration A				x		
1v	Median	SWAT's 95-percentile with VFS	Calibration A					x	
2v	Median	SWAT's best with VFS	Calibration A					x	
3v	Median	SWAT's 5-percentile with VFS	Calibration A					x	
10	Not used	SWAT's best	Calibration A						x
11	Median	Not used	Calibration A						x

Table 3: Scope of the uncertainties that influences MyLake simulations at the Vanemfjorden simulation in the case study.

Uncertainties in inputs to Vanemfjorden basin		
	Loading through Sundet	
		MyLake (Storefjorden) CIU
		MyLake (Storefjorden) OPU
		MyLake (Storefjorden) SIU
	Loading from Vanemfjorden local catchment	
		Model choice uncertainty (SWAT or INCA-P)
		Model OPU
Vanemfjorden basin		
	MyLake (Vanemfjorden) CIU, see above	
	MyLake (Vanemfjorden) OPU	
	MyLake (Vanemfjorden) SIU, see above	

Table 4: List of MyLake parameters calibrated for the Vanemfjorden basin in the present study and their corresponding processes they control

Name	**Parameter**	**Unit**	**Meaning of greater values**
p_1	PAR saturation	mol quanta m-2 sec-1	Greater photosynthesis amount given light intensity
p_2	Suspended inorganic particles resuspension in epilimnion	m sediment day-1	Faster resuspension from sediment
p_3	Suspended inorganic particles sinking speed	m day-1	Faster sinking speed
p_4	Algae sinking speed	m day-1	Faster sinking speed
p_5	Specific algae mortality rate	day-1	Faster mortality, given temperature
p_6	Algae growth rate	day-1	Faster growth, given conditions of light, temperature and available P
p_7	P-Langmuir isotherm half-saturation	mg P m-3	Dissolved inorganic P is more sticky to suspended inorganic particles
p_8	Vertical diffusion coefficient	unitless	Greater exchange of heat between neighboring horizontal layers
p_9	Wind sheltering coefficient	unitless	Greater physical mixing impact of a given wind speed
p_{10}	Snow albedo (absolute scale)	unitless	Faster melting of snow on lake surface when air temperature is greater than 0 C

IV

Technical Note: Simultaneous calibration of two sequentially connected models

Koji Tominaga
Department of Biosciences, Univesity of Oslo
Norwegian Institute for Water Research

Jostein Starrfelt
Norwegian Institute for Water Research

Øyvind Kaste
Norwegian Institute for Water Research

Tom Andersen
Department of Biosciences, University of Oslo

Abstract

Process-oriented environmental models rely on the calibration procedure in order to gain relevance and applicability, and parameter uncertainty surrounding the calibration can be effectively described using probability-based methods. However, the parameter uncertainty is not properly addressed when two models are calibrated separately but used together. Here, using a case study on freshwater eutrophication, we demonstrate how simultaneously calibrating two sequentially connected models better contained parameter uncertainty due to calibration. Use of models in sequence is a common practice in many application fields but simultaneous calibration is often underutilised in the literature. Because it implicates interpretation of the modelled system, we urge simultaneous calibration where applicable.

1 Introduction

Many process-oriented dynamic environmental models, for example hydrological and ecosystem models, are calibrated against field observations (Oreskes et al., 1994; Tarantola, 2006; Mooij et al., 2010). The main aim of calibration is to find one or more parameter sets that reproduce the target observations. The ability of reproducing observation is often qualified as being behavioural in the calibration context. Because multiple parameter sets can appear equally behavioural, known as the equifinality phenomenon (Beven and Freer, 2001), they are considered equally likely and should be given equal representation. Assembling all behavioural parameters as a multi-dimensional posterior parameter space is therefore useful when the model is used beyond the original physical and temporal scope (i.e., 'prediction' or 'forecast'). For example, calibration suggests a particular parameterisation A to represent the dynamic processes in the model, because the parameters could reproduce the target field observations as close as it could. Using the calibrated parameters, one could in turn produce a forecast. Often this forecast carries managerial significance and is the motivation of setting up the model. Now, equifinality implies that there is also another parameterisation B and even many more that come as close to the field observations as the parameterisation A. However, because the parameterisation B is different from the parameterisation A, derived forecast will also be different. Therefore, the managers should consider all forecasts that are derived from a variety of different but equally likely parameterisations.

The preciseness and accuracy of these forecasts from all posterior parameter sets depend on several factors, such as the completeness of model process representation of complex real world systems and consistency in the spatiotemporal scope between calibration and forecast. Another important factor that determines the preciseness and accuracy of the forecasts is quality of the target field data. Collecting relevant field data is crucial because what is being modelled needs to be accurately measured by field observation. In practice this is never perfect due to technical factors (e.g., laboratory errors) and stochasticity in nature not captured in sparse sampling (e.g., hourly varying state that are are sampled on an instantaneous basis). Inaccurate and irrelevant target observed data will naturally result in inaccurate calibration.

Here we study how this calibration uncertainty may be addressed when two models are sequentially connected and when each model requires calibration. As an example case, we take on a modelling task where a stream-based nutrient runoff model INCA-P (Wade et al., 2002) and a lake model MyLake (Saloranta and Andersen, 2007) were used in sequence to predict how the land-use and agricultural activities in the catchment influences the lake eutrophication status (total P concentration as the proxy). A common approach in such a sequential modelling task has often been to conduct calibration separately (e.g., Tominaga et al., unpublished manuscript):

i) first calibrate INCA-P using river observations,

ii) choose a singular forecast based on a particular parameterisation,

iii) pass on the derived forecast to use as the driving input for calibrating MyLake using lake observations, and

iv) produce forecast in the lake.

This approach is useful when expert users differ between various models. However, because typical INCA-P applications require calibration of multiple parameters, INCA-P is not free from the equifinality phenomenon. Therefore, such a strategy of reducing INCA-P's parameter uncertainty to a singular run to use as the input for MyLake is not ideal.

In order to address this problem, we instead calibrated both models simultaneously. In other words, INCA-P parameters and MyLake parameters are calibrated at the same time to target both stream data and lake data. The advantages of this arrangement are i) that INCA-P is now calibrated using more information (i.e., lake observations) that is less stochastic than the stream observation and ii) that MyLake is less obscured because the calibration uncertainty of INCA-P is included. The disadvantage is that INCA-P's calibration is influenced by MyLake's inacuracy and impreciseness because both models affect lake simulation. To put this into a generalised context, simultaneously calibrating two sequentially connected models should be most effective when a confidence in the calibration of upstream model is less (e.g., due to lack of relevant observation to use as the target during the calibration) than the calibration of downstream model. In our case of sequential calibration, we suspected that during the INCA-P-only calibration, INCA-P may have suffered from instanteneous and stochasic nature of stream chemistry data (taken fortnightly) that may not have represented the daily cumulative discharge that INCA-P simulates. We also suspected that this may have reduced accuracy in simulated loading to drive MyLake. On the contrary, lake P concentration is a temporarily aggregated data type and is less prone to data stochasticity.

Simultaneous calibration of two sequentially connected dynamic models has unfortunately been underutilised among the environmental models despite its anticipated benefits. We demonstrate here how parameterisation or posterior equifinal parameter space is changed from two separate model calibrations to a simultaneous calibration of both models. Hence, the aim of this study is to show how additional information in the downstream observation can be used to better condition the upstream model. The principles shown here are applicable in any model application in which multiple models are connected sequentially.

2 Methods

The Vansjø-Morsa catchment is located in the southeastern Norway and is one of the most cultivated areas in the country. Occasional toxic algae blooms in the lake Vansjø in the past decades brought the management to study the nutrient budget of the lake (Bouraoui et al., 2009). The present study focuses on the eastern basin of Vansjø called Storefjorden, which is the larger and upstream basin of the lake, representing approximately 90 percent of the water flowing out of the lake. The western basin is of greater management concern for eutrophication impacts and models are also being applied in the western catchment and the basin. But for simplicity the present study will consider only the connection between the eastern catchment modelled by INCA-P and the eastern lake basin modelled by MyLake. The temporal scope was 2002-2010.

The INCA-P model is a process-based mass balance model for simulating the mobilization and transport of phosphorus in a catchment (Wade et al., 2002). It calculates flow of water through the soil in different land use categories, routes the water with phosphorus and sediments into a river, generating concentrations of phosphorus and sediment in the water column as well as flow in a daily resolution. INCA-P requires hydrological forcing in terms of hydrological effective rainfall and soil moisture deficit and also uses phosphorus inputs in terms of sewage and effluents to the river as well as fertilizer applied to the land. The MyLake model is a one-dimensional process-oriented model for daily revolution of heat and phosphorus concentration gradient s(Saloranta and Andersen, 2007). It also simulates sinking of sedimenting particles and particle surface reactions, as well as lake surface heat balance. The model is capable of simulating ice formation and snow accumulation on the surface of the lake during the winter season. See Table 1 for the types of required data for these models and see Tominaga et al. (unpublished manuscript) for details of the study sites and other model specifics.

INCA-P was applied on the main river flowing into the lake (Hobølelva) and the rest of the catchment that also drains into the lake. The INCA-P only calibration was conducted using the river data which had been collected from the Kure dam located about 20 km upstream of the lake (Table 2). In order to relay the catchment runoff simulation to MyLake, the same INCA-P parameters were used downstream of the dam and the rest of the runoff that does not originate from Hobølelva. MyLake in turn takes these INCA-P outputs as an input to simulate the daily vertical profile of the lake. In-lake data were used to calibrate MyLake (Table 2). The MyLake-only calibration used the median INCA-P runoff simulation with respect to the total P loading among the posterior equifinality space as the runoff input. Simultanesous calibration of the two models was conducted by testing both INCA-P and MyLake parameters together to connect from the headwater to the lake in each iteration. Then the Kure dam data and in-lake data were together used as the target for calibration.

All our calibrations used the DREAM-MCMC (Differential Evolution Adaptive Metropolis Markov Chain Monte Carlo) algorithm (Vrugt et al., 2009). DREAM operates on multiple combination chains for each Markov chain iteration, unlike the singular combination chain of ordinary MCMC methods. DREAM searches the multidimensional parameter space more thoroughly by operating on multiple combination chains, and was appropriate for the present study. Refer to Table 3 for the list of calibrated parameters. The intial parameter combinations were chosen using Latin hypercube sampling to minimise blind space. The settings for DREAM were summarised in Table 4. For each type of observation an error variance was drawn using Gibbs sampling, and the formal likelihood was calculated according to the following:

$$\mathscr{L}(\hat{\mathbf{Y}}|\boldsymbol{\theta},\mathbf{X},\boldsymbol{\sigma}^2) = \prod_{o}\prod_{t=1}^{n_0} \frac{\exp(-\frac{(y_{o,t}(\mathbf{X},\boldsymbol{\theta})-\hat{y}_{o,t})^2}{2\sigma_o^2})}{\sqrt{2\pi\sigma_o^2}}, \tag{1}$$

where $\hat{\mathbf{Y}}$ is the model simulation, $\boldsymbol{\theta}$ is the parameters, $\mathbf{X}$ is the forcing data, $\boldsymbol{\sigma}$ is the error variance, n_o is the number of observation for observation type o, $y_{o,t}$ is the model simulated vaolue at observation time t, and $\hat{y}_{o,t}$ is the corresponding observation(Gelman et al., 2003). Due to the computational resource limitation, we could not observe that the DREAM calibrations converged according to the Gelman-Rubin method (Gelman et al., 2003). We therefore did not assume burn-in iterations and gathered all iteration distributions for use as the posterior distribution.

3 Results

The posterior parameter probability density functions differed between the calibrations (Figure 1). The group of INCA-P parameters that are conditioned (i.e., having a narrower and higher-reaching density function) was different between the two calibrations, namely the INCA-only calibration and the simultaneous calibration. The INCA-P only calibration mostly relied on parameters 11, 12, 14, and 16 that control surface erosion to achieve behavioural simulations. The simultaneous calibration constrained additional parameters, but did not sharply condition parameter 11. This indicates that different set of processes were emphasised depending on the calibration. The output variability due to parameters was negligible for discharge data within calibration due to parameter uncertainty and between calibrations, but for suspended solid and total P concentrations, the output variability between calibration became notable (Figure 2). High loading events were not consistently modelled by the two calibrations, and in case of total P concentration, the correspondence was poor between the calibrations, although both calibrations were poor at reducing model error for the total P concentration.

MyLake parameters were also differently conditioned depending on the calibration (Figure 1). Tighter calibration were concluded with the MyLake-only calibration than with the simultaneous calibration (e.g., parameters 3, 4, 8, and 9). A few explanations could be made for this difference. First, the parameter set of INCA-P that was chosen to make the input for MyLake during the MyLake only calibration might have been suitable for reproducing the observation, but only so suitable in a specific way. Second, in the simultaneous calibration, MyLake is more greatly influenced by the forcing inputs that were simulated by INCA-P and therefore more affected by the INCA-P parameters. The model performance of MyLake at the Storefjorden observation site support these explanations: the simultaneous calibration had a wider confidence envelope than the MyLake only calibration, although the simultaneous calibration seems to have encompassed the observations within the confidence envelope more thoroughly than the MyLake only calibration (Figure 3).

Covariance structure among the posterior parameters were also different between the calibrations (Figure 4). The correlation among the parameters illustrate processes that complement each other within the model; our results therefore indicates that the dominant combination of processes that operate together are different depending on the calibration. Fundamentally, we can arrive at different conclusions about the modelled system depending on the target observation for calibration. Furthermore, the complex covariance structure between the INCA-P parameters and MyLake parameters in the simultaneous calibration confirms that the parameter uncertainty of INCA-P is not negligible and needs a proper handling. The previously common method of modelling two sequentially connected models by calibrating them separately therefore would have fallen short in properly addressing the overall paraemter uncertainty.

4 Conclusions

Our illustration is a clear example case that strongly suggests a simultaneous calibration when two models are sequentially connected, in order to reduce the blind spot caused by paramter uncertainty. Similar conclusions might be made in other application fields of two sequentially connected environmental models. But we cannot generalise the benefits of connecting the models during calibration. In our case, the following conditions raised the value of a simultaneous calibration: i) the ultimate purpose of the modelling tasks was to simulate the lake water conditions, not the intermediate results in the river, ii) although the information was still relevant, the observation data at the Kure dam might have suffered from temporal stochasticity and discrepancy in time resolution, and iii) MyLake application in this lake was sensitive to the forcing data provided by INCA-P.

Acknowledgements

This work has been supported by the EUTROPIA project (Reseach Council of Norway), the EU FP7 project REFRESH, and the Strategic Institute Initiative funded by the Norwegian Institute for Water Research, and the Research Council of Norway. We thank Tuomo Saloranta for help with model customization, and Anne Bjrkenes Christiansen, Stein Beldring, and Deborah Lawrence for their help with hydrology modelling. We thank all people at the Norwegian Institute for Agricultural and Environmental Research, the Norweigian Institute for Water Research, the Norwegian Water Resources and Energy Irectorate, and the Department of Chemistry and Department of Biology at the University of Oslo who collected and analysed the monitoring samples from the lake and the streams.

References

Beven, K. and Freer, J.: Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology, J. Hydrol., 249, 11–29, 2001.

Bouraoui, F., Grizzetti, B., Adelsköld, G., Behrendt, H., Miguel, I. d., Silgram, M., Gómez, S., Granlund, K., Hoffmann, L., Kronvang, B., Kværnø, S., Lázár, A., Mimikou, M., Passarella, G., Panagos, P., Reisser, H., Schwarzl, B., Siderius, C., Sileika, A. S., Smit, A. a. M. F. R., Sugrue, R., VanLiedekerke, M., and Zaloudik, J.: Basin characteristics and nutrient losses: the EUROHARP catchment network perspective, J. Environ. Monit., 11, 515–525, 2009.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B.: Bayesian Data Analysis, Second Edition, Chapman and Hall/CRC, 2 edn., 2003.

Mooij, W. M., Trolle, D., Jeppesen, E., Arhonditsis, G., Belolipetsky, P. V., Chitamwebwa, D. B. R., Degermendzhy, A. G., DeAngelis, D. L., Domis, L. N. D. S., Downing, A. S., Elliott, J. A., Fragoso, C. R., Gaedke, U., Genova, S. N., Gulati, R. D., Hakanson, L., Hamilton, D. P., Hipsey, M. R., 't Hoen, J., Huelsmann, S., Los, F. H., Makler-Pick, V., Petzoldt, T., Prokopkin, I. G., Rinke, K., Schep, S. A., Tominaga, K., Van Dam, A. A., Van Nes, E. H., Wells, S. A., and Janse, J. H.: Challenges and opportunities for integrating lake ecosystem modelling approaches, Aquat. Ecol., 44, 633–667, 2010.

Oreskes, N., Shrader-Frechette, K., and Belitz, K.: Verification, Validation, and Confirmation of Numerical Models in the Earth Sciences, Science, 263, 641–646, 1994.

Saloranta, T. M. and Andersen, T.: MyLake - A multi-year lake simulation model code suitable for uncertainty and sensitivity analysis simulations, Ecol. Model., 207, 45–60, 1, 2007.

Tarantola, A.: Popper, Bayes and the inverse problem, Nat. Phys., 2, 492–494, 2006.

Vrugt, J. A., ter Braak, C. J. F., Diks, C. G. H., Robinson, B. A., Hyman, J. M., and Higdon, D.: Accelerating Markov Chain Monte Carlo Simulation by Differential Evolution with Self-Adaptive Randomized Subspace Sampling, J. Nonlin. Sci. Num., 10, 273–290, 2009.

Wade, A. J., Whitehead, P. G., and Butterfield, D.: The Integrated Catchments model of Phosphorus dynamics (INCA-P), a new approach for multiple source assessment in heterogeneous river systems: model structure and equations, Hydrol. Earth Syst. Sc., 6, 583–606, 3, 2002.

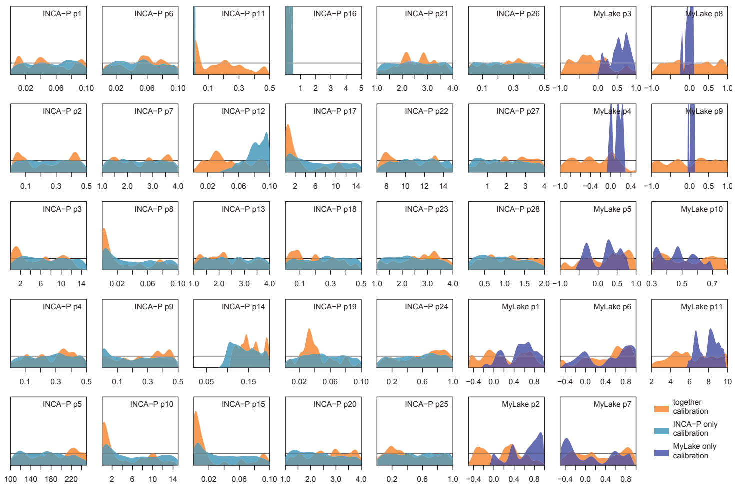


Figure 1: Posterior probability density functions. Calibration schemes are overlaid to emphasise the difference in the calibrated parameter density. Gaussian normal kernel density was assumed to produce the density functions. The prior uniform distributions are indicated here by the horizontal line.

Table 1: Type of forcing input data and other fixed model configurations.

Use	Category	Variables	Period
Input for both INCA-P and MyLake	Weather	Precipitation, air temperature, wind speed*, cloudiness*, air pressure*, relative humidity*	2002-2010
Input for INCA-P	Hydrology	Discharge	2002-2010
	Geography	Land use fraction, reach definition	n/a
Output from INCA-P (Input for MyLake)	Runoff	Flow volume, loadings of total P and suspended particles	2002-2010
Input for MyLake	Lake morphology	Bathymetry	n/a

* only needed for MyLake

Table 2: Summary of observation data that were used to calibrate the models.

Water body name	Category	Sample size	Period
Hobølelva at the Kure dam	Flow volume	74	2002-2004
	Total P concentration	74	2002-2004
	Suspended particle concentration	1461	2002-2005
Vansjø Storefjorden basin	Total P concentration	153	2004-2010 (without winter)
	Particulate P concentration	152	2004-2010 (without winter)
	Suspended particle concentration	84	2004-2010 (without winter)
	Chl-a concentration	153	2004-2010 (without winter)
	Dissolved inorganic P concentration	150	2004-2010 (without winter)
	Temperature at 8 depths	99 each	2010 (May through August)

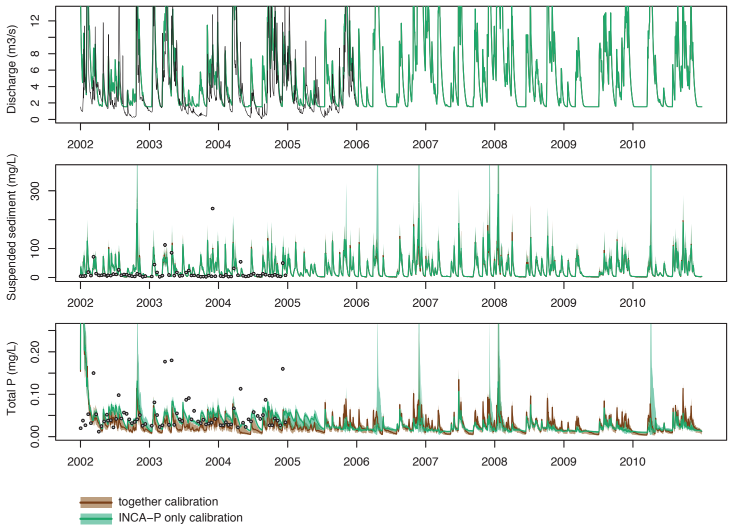


Figure 2: Model performance of INCA-P at the Kure observation station. Daily median among the posterior simulations is shown using a line, while the daily 10- and 90-percentiles are shown as the banded area. Observed values are shown with a line for the discharge data and gray dots for the other data.

Table 3: List of calibrated paraemters. All parameters have uniform prior parameter distributions.

Model	Parameter	Process	Prior minimum	Prior maximum
INCA-P	p1	Soil erosion land use 1	0.0001	0.1
	p2	Soil erosion land use 1	0.0001	0.5
	p3	Soil erosion land use 1	0.05	15
	p4	P mobilization (?) land use 1	0.005	0.5
	p5	P? mobilization all land uses	100	250
	p6	P? mobilization	0.002	0.1
	p7	Residence time soil water land use 1	1	4
	p8	Soil erosion land use 2	0.0001	0.1
	p9	Soil erosion land use 2	0.0001	0.5
	p10	Soil erosion land use 2	0.05	15
	p11	P mobilization (?) land use 2	0.005	0.5
	p12	P mobilization (?) land use 2	0.002	0.1
	p13	Residence time soil water land use 2	1	4
	p14	Hydrology, all 4 reaches.	0.02	0.2
	p15	Soil erosion land use 3,4, and 5	0.0001	0.1
	p16	Soil erosion land use 3,4, and 5	0.0001	5
	p17	Soil erosion land use 3,4, and 5	0.05	15
	p18	P mobilization (?) land use 3,4, and 5	0.005	0.5
	p19	P mobilization (?) land use 3,4, and 5	0.002	0.1
	p20	Residence time soil water land use 3	1	4
	p21	Residence time soil water land use 4	1	4
	p22	Flow erosion of bed sediment all reaches	7	15
	p23	Residence time soil water land use 5	1	4
	p24	Phosphorus dynamics in water	0.001	1
	p25	Phosphorus dynamics in stream bed	0.001	1
	p26	??	0.01	0.5
	p27	Phosphorus dynamics in water column/pore water	0.001	4
	p28	Scaling input data from sewage treatment works and scattered dwellings.	0.1	2
MyLake	p1	Light sensitivity in photosynthesis	-0.5	1
	p2	Sediment resuspension of inorganic particles	-0.5	1
	p3	Sinking of inorganic particles	-1	1
	p4	Sinking of algae cells	-1	0.5
	p5	Algae mortality	-1	1
	p6	Algae growth	-0.5	1
	p7	Inorganic P exchange on inorganic particles	-0.5	1
	p8	Vertical heat diffusion	-1	1
	p9	Wind-induced mixing	-1	1
	p10	Snow albedo	0.3	0.8
	p11	P speciation (inert P in lake)	2	10

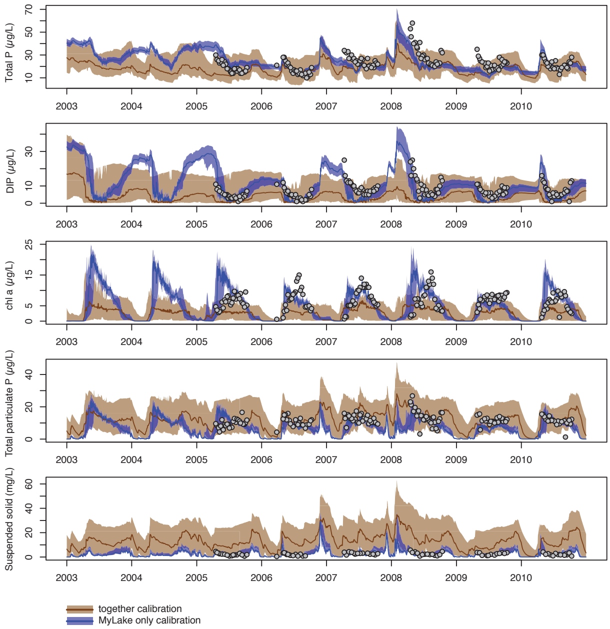


Figure 3: Model performance of MyLake at the observation station in the Storefjorden basin of the lake Vansjø. Daily median among the posterior simulations is shown using a line, while the daily 10- and 90-percentiles are shown as the banded area. Only the chemical measurements are shown. Observed values are shown using gray dots.

Table 4: DREAM specific configurations.

DREAM configuration	INCA-P only	MyLake only	INCA-P and MyLake together
number of parameters	28	11	39
number of chains	20	8	32
length of chains	640	400	400
number of crossover values	6	6	6
maximum number of pairs of chains from which proposals are affected	3	3	3
number of iterations between each use of full step size	5	5	5
width of uniform distribution of epslion	0.05	0.05	0.05
variance of normal distribution of epsilon	0.001	0.001	0.001

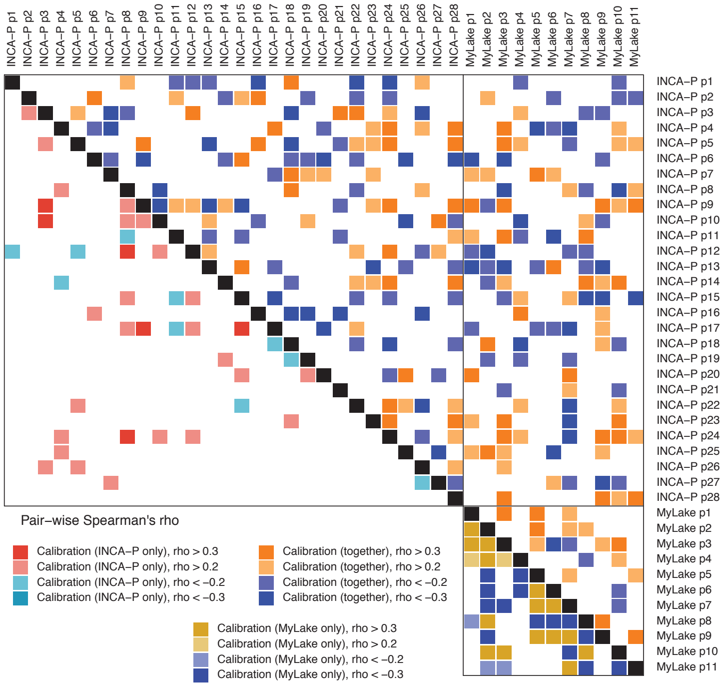


Figure 4: Correlation plot among the posterior parameters. Pair-wise Spearman's ρ is shown by colours. The top trianglular space is used for the calibration in which INCA-P and MyLake were used together, and the bottom two triangular spaces were for the separate calibrations of INCA-P and MyLake.