

DISS. ETH NO. 16833

**MODELLING FUNCTIONAL GROUPS OF PHYTOPLANKTON:
GROUP SELECTION AND EFFECT ON MODEL TRANSFERABILITY
TO DIFFERENT LAKES**

A dissertation submitted to

ETH ZURICH

for the degree of

Doctor of Sciences

presented by

JOHANNA MIELEITNER

Dipl.-Ing., Technical University of Munich (TUM)

born 30.12.1975

citizen of Germany

accepted on the recommendation of

Prof. Dr. Peter Reichert, examiner

Prof. Dr. Bernhard Wehrli, co-examiner

Prof. Dr. Jürgen Benndorf, co-examiner

2006

Summary

Many mechanistic lake models have been developed in the last 30 years. They represent simplified and partial descriptions of reality and play the role of archives of knowledge about mechanisms, process formulations, and parameter values. Comparisons of the results of such models with measured data can help to improve our understanding of lake ecosystem function. As the model represents hypotheses about lake ecosystem function a comparison of model results with data is a test of the hypotheses formulated in the model. Predictions made by such models under different management scenarios can support lake management.

However, due to severe simplifications that have to be made for model formulation, lack of knowledge about processes and their formal description, site-specific process formulation, not identifiable model parameters, and poorly known external influence factors, such predictions are very uncertain. Particular difficulties are finding universal process formulations and parameter values and choosing an adequate model complexity for a specified model use.

The **aim** of this study was to investigate and improve the transferability of a simple biogeochemical lake model to different lakes (i) by increasing the universality of process formulations and parameter values, and (ii) by improving the description of the phytoplankton growth with a more detailed model that distinguishes functional groups of species with similar properties. In order to fulfil the second aim a procedure to identify functional groups of phytoplankton was developed.

The transferability of the different models was tested by comparing simulations with data from three lakes of different trophic state. The three lakes selected for this project are Walensee (oligotrophic), Lake Zürich (mesotrophic), and Greifensee (eutrophic). The work is divided into three research tasks.

First the transferability of a biogeochemical lake model developed for Lake Zürich was analyzed by confronting it with data from two other lakes. An iterative procedure of: (1) initial simulation, (2) lake-specific calibration, (3) joint calibration for all lakes and (4) sensitivity and identifiability analysis of model parameters was applied with the goal of finding a model with a minimum number of lake-specific parameter values. The application of this procedure led to a revised model structure with recalibrated values of some model parameters. The agreement of the model results with data from all lakes is good. The revised model needs only three parameters with lake-specific values. As the trophic states of the three lakes are significantly different, this demonstrates a remarkably high generality of the model. The main reasons for needing different values of these parameters were identified. They are the coarse spatial resolution of the sediment, the simplified description of mineralization processes in the sediment and the lack of a description of predation of Chaoborus larvae and fish on zooplankton. Conclusions were drawn regarding which model extensions would make the model even more general.

In the **second** step data analysis was combined with biological knowledge to find a possible division of phytoplankton species into groups of similar behaviour. The goal of this step was to find functional groups of phytoplankton species which respond similarly to environmental conditions because they have similar properties. Functional groups of phytoplankton were identified using a combination of prior knowledge and statistical cluster analysis of long-term species-level data from the three lakes. The groups were identified by a procedure that (i) starts from a taxonomic classification, (ii) divides the taxonomic groups into subgroups by clustering species with similar yearly occurrence pattern based on empirical data, (iii) assigns properties responsible for the occurrence pattern to subgroups based on biological knowledge, and, finally, (iv) merges subgroups with similar properties and occurrence pattern across taxonomic boundaries to different aggregation levels. This led to suggestions for functional groups at multiple levels of aggregation.

In a **third** step one of the suggestions to divide phytoplankton into functional groups from step 2 was implemented in the biogeochemical lake model from step 1 and tested by comparing model results with data from the three lakes. The spatial resolution of the model was simplified for this step to increase computational speed. The implemented four groups correspond to the highest aggregation level of functional groups proposed for these three lakes. The groups are distinguished by different values of maximum specific growth rate ($k_{gro,ALG}$), half-saturation light intensity ($K_{I,ALG}$), edibility ($k_{gro,ZOO}$), sedimentation velocity (v_{sed}) and half-saturation concentration with respect to phosphate uptake ($K_{P,ALG}$). After calibration, the occurrence pattern of the functional groups in the three lakes can be qualitatively represented by the model. However some quantitative and systematic differences between model results and data remain. Furthermore, the division of total phytoplankton into functional groups makes these groups more sensitive to parameter changes than total phytoplankton was before. The sensitivity of total phytoplankton to parameter changes remains the same. This indicates a poor predictability of the occurrence of functional groups in these lakes but an equal predictability of total phytoplankton biomass when functional groups are considered. Strong phosphorus limitation during the summer could be the cause for this observation. The fit for total phytoplankton improves slightly if functional groups are considered.

General conclusions: It can be concluded that both the simple plankton model as well as the functional group model have a good transferability after having been calibrated to the data of the three lakes. A good fit was achieved for all three lakes with a minimum of differences in parameter values between the lakes. However, the relatively high sensitivity of model results to parameter values still makes predictions quite uncertain. The data analysis showed a high variety of occurrence patterns of different species in the lakes. This makes it difficult to identify good functional groups. It was still possible to derive a technique that combines clustering of occurrence pattern with biological knowledge to identify such functional groups. The consideration of functional groups in the lake model led to reasonable agreement of simulation results with the general patterns observed in the measurements, but it increased the sensitivity of the model to parameter values. Modelling functional groups of phytoplankton improves the fit for total phytoplankton slightly, but it decreases the identifiability of the model parameters if no data for the phytoplankton groups are available.

Zusammenfassung

In den letzten 30 Jahren wurden viele mechanistische Seemodelle entwickelt. Sie sind vereinfachte und unvollständige Beschreibungen der Realität und dienen als Wissensarchive über Mechanismen, Prozessformulierungen und Parameterwerte. Der Vergleich der Ergebnisse solcher Modelle mit Messdaten kann helfen, das Verständnis für die Funktionsweise des Ökosystems See zu verbessern. Der Vergleich der Modellergebnisse mit Messdaten ist ein Test der im Modell formulierten Hypothesen. Vorhersagen von Seemodellen für verschiedene Management Szenarien können das See-Management unterstützen.

Solche Modellvorhersagen sind allerdings sehr unsicher. Gründe dafür sind die starken Vereinfachungen, die bei der Modellformulierung gemacht werden müssen, fehlendes Wissen über Prozesse und ihre Formulierung, See-spezifische Prozessformulierungen, nicht-identifizierbare Parameter und nur ungenau bekannte externe Einflussgrößen. Besondere Schwierigkeiten sind das Finden universeller Prozess-Formulierungen und Parameterwerte und die Wahl einer angemessenen Modell-Komplexität für eine bestimmte Modellanwendung.

Das **Ziel** dieser Studie war es, die Übertragbarkeit eines einfachen Seemodells auf andere Seen zu untersuchen und zu verbessern. Das sollte einerseits durch eine Erhöhung der Universalität der Prozessformulierungen und andererseits durch die Verbesserung der Beschreibung des Phytoplankton Wachstums mit einem detaillierten Modell, das mehrere funktionelle Gruppen von Phytoplankton Spezies unterscheidet, erreicht werden. Um das zweite Ziel zu erreichen, wurde eine Methode zur Identifizierung funktioneller Gruppen des Phytoplanktons entwickelt.

Die Übertragbarkeit verschiedener Modelle wurde getestet, indem Simulationen mit den Daten von drei Seen unterschiedlichen trophischen Zustands verglichen wurden. Folgende Seen wurden für dieses Projekt ausgewählt: Der Walensee (oligotroph), der Zürichsee (mesotroph) und der Greifensee (eutroph). Diese Arbeit wurde in drei Teilprojekte unterteilt.

(1.) Zunächst wurde die Übertragbarkeit eines für den Zürichsee entwickelten biogeochemischen Seemodells auf zwei andere Seen getestet. Dazu wurde ein iteratives Verfahren angewendet, welches aus folgenden Schritten besteht: (1) Anfangssimulation, (2) See-spezifische Kalibrierung, (3) gemeinsame Kalibrierung für alle Seen und (4) Sensitivitäts- und Identifizierbarkeitsanalyse der Modellparameter. Ziel war es, ein Modell mit einer möglichst kleinen Zahl von See-spezifischen Parameterwerten zu finden. Dieses Verfahren führte zu einer veränderten Modellstruktur mit einigen neu kalibrierten Parameterwerten. Dadurch konnte eine gute Übereinstimmung der Modellergebnisse mit den Daten aller Seen erreicht werden. Nur für drei Parameter wurden See-spezifische Werte benötigt. Dies zeigt eine erstaunlich hohe Universalität des Seemodells, insbesondere wenn man berücksichtigt, dass der trophische Zustand der drei Seen verschieden ist.

Die wesentlichen Gründe für die Notwendigkeit verschiedener Parameterwerte konnten identifiziert werden. Sie liegen in der relativ groben räumlichen Auflösung des Sedimentmodells, in der stark vereinfachten Beschreibung der Mineralisierungsprozesse und in der fehlenden Beschreibung der Predation von Insektenlarven (Chaoborus) und Fischen auf Zooplankton. Daraus konnten Rückschlüsse gezogen werden, welche Modellverbesserungen nötig sind, um die Universalität des Modells weiter zu verbessern.

(2.) Im nächsten Schritt wurde eine Datenanalyse mit biologischem Wissen kombiniert, um eine Einteilung der Phytoplankton Spezies in funktionelle Gruppen zu finden. Das Ziel dieser Analyse war es, funktionelle Gruppen zu finden, die ähnliche Eigenschaften haben und daher in ähnlicher Weise auf unterschiedliche Umweltbedingungen reagieren. Durch eine Kombination von Vorwissen und statistischer Clusteranalyse von Langzeitdaten (Phytoplankton-Daten auf Spezies-Level) der drei Seen wurden solche funktionelle Gruppen gefunden. Die angewendete Methode besteht aus den folgenden Schritten: (i) Einteilung der Phytoplankton Spezies in taxonomische Gruppen, (ii) Unterteilung der taxonomischen Gruppen in Untergruppen durch Gruppierung (clustering) von Spezies mit ähnlichem Vorkommen innerhalb jedes Jahres (durch Clusteranalyse der Daten), (iii) Herausfinden der Eigenschaften die für das Erscheinungsmuster der Untergruppen verantwortlich sind, mit Hilfe von biologischem Vorwissen, (iv) Zusammenfassen von Untergruppen mit ähnlichen Eigenschaften und Vorkommen auch über taxonomische Grenzen hinweg. Dieses Vorgehen führte zu Vorschlägen für die Einteilung des Phytoplanktons in funktionelle Gruppen auf verschiedenen Aggregationsniveaus.

(3.) In einem weiteren Schritt wurde einer der Vorschläge für die Einteilung des Phytoplanktons in funktionelle Gruppen aus Schritt 2 in das Seemodell aus Schritt 1 implementiert und getestet, indem die Modellergebnisse mit Daten verglichen wurden. Um die Rechenzeit zu verkürzen, wurde die räumliche Auflösung des Modells für diesen Schritt vereinfacht. Es wurden vier funktionelle Gruppen implementiert, die dem höchsten Aggregationsniveau aus Schritt 2 entsprechen. Die Gruppen unterscheiden sich in den Parameterwerten für die spezifische Wachstumsrate ($k_{gro,ALG}$), die Halbsättigungs-Lichtintensität ($K_{I,ALG}$), die Fressbarkeit ($k_{gro,ZOO}$), die Sedimentationsgeschwindigkeit (v_{sed}) und die Halbsättigungs-Konzentration für Phosphat ($K_{P,ALG}$). Nach der Kalibrierung konnten die typischen Vorkommensmuster (occurrence pattern) der funktionellen Gruppen qualitativ vom Modell wiedergegeben werden. Es bleiben aber quantitative und systematische Abweichungen zwischen Modellergebnissen und Daten. Die Unterteilung des Phytoplanktons in funktionelle Gruppen führt dazu, dass die Simulation für diese funktionellen Gruppen viel sensitiver auf Parameteränderungen reagiert als die Simulation für das gesamte Phytoplankton im aggregierten Modell. Die Sensitivität des Gesamtphytoplanktons auf Parameteränderungen bleibt jedoch etwa gleich. Das deutet auf eine schlechte Vorhersagbarkeit des Auftretens der funktionellen Gruppen in diesen Seen hin. Die Vorhersagbarkeit der gesamten Phytoplankton-Biomasse scheint gleich zu bleiben, wenn funktionelle Gruppen modelliert werden. Die starke Phosphatlimitierung während des Sommers könnte ein Grund für diese Beobachtung sein.

Allgemeine Schlussfolgerungen: Sowohl das einfache Phytoplanktonmodell als auch das Modell mit funktionellen Gruppen zeigt nach der Kalibrierung mit Daten aus drei verschiedenen Seen eine gute Übertragbarkeit. Mit beiden Modellen konnten gute Simulationsergebnisse erreicht werden, wobei nur wenige Parameter verschiedene Werte für die verschiedenen Seen haben. Die hohe Sensitivität der Modellergebnisse auf die Parameterwerte führt aber zu einer hohen Unsicherheit der Modellvorhersagen. Bei der Datenanalyse zeigte sich eine sehr hohe Variabilität im Auftreten der verschiedenen Spezies in den drei Seen. Das macht das Finden von funktionellen Gruppen schwierig. Es wurde eine Technik vorgeschlagen, die die Clusteranalyse basierend auf typischen Mustern im jährlichen Auftreten („occurrence patterns“), mit biologischem Wissen kombiniert, um funktionelle Gruppen zu identifizieren. Die Berücksichtigung von funktionellen Gruppen im Modell führte zu einer guten Übereinstimmung der Simulationsergebnisse mit dem typischen Auftreten, das in den Messungen beobachtet werden kann, aber sie erhöhte die Sensitivität des Modells auf Parameterwerte. Die Modellierung der funktionellen Gruppen verbessert die Ergebnisse für das gesamte Phytoplankton, aber sie reduziert die Identifizierbarkeit des Modells erheblich, wenn keine Daten für die Phytoplanktongruppen vorhanden sind.