Parameterization of aquatic ecosystem functioning and its natural variation: Hierarchical Bayesian modelling of plankton food web dynamics

Veera Norros [a,b,*], Marko Laine [c], Risto Lignell [a], Frede Thingstad [d]

[a] *Marine Research Centre, Finnish Environment Institute, Helsinki, Finland*
[b] *Metapopulation Research Centre, Department of Biosciences, University of Helsinki, Finland*
[c] *Finnish Meteorological Institute, Helsinki, Finland*
[d] *Department of Biology, University of Bergen, Bergen, Norway*

ABSTRACT

Methods for extracting empirically and theoretically sound parameter values are urgently needed in aquatic ecosystem modelling to describe key flows and their variation in the system. Here, we compare three Bayesian formulations for mechanistic model parameterization that differ in their assumptions about the variation in parameter values between various datasets: 1) global analysis - no variation, 2) separate analysis - independent variation and 3) hierarchical analysis - variation arising from a shared distribution defined by hyperparameters. We tested these methods, using computer-generated and empirical data, coupled with simplified and reasonably realistic plankton food web models, respectively. While all methods were adequate, the simulated example demonstrated that a well-designed hierarchical analysis can result in the most accurate and precise parameter estimates and predictions, due to its ability to combine information across datasets. However, our results also highlighted sensitivity to hyperparameter prior distributions as an important caveat of hierarchical analysis. In the more complex empirical example, hierarchical analysis was able to combine precise identification of parameter values with reasonably good predictive performance, although the ranking of the methods was less straightforward. We conclude that hierarchical Bayesian analysis is a promising tool for identifying key ecosystem-functioning parameters and their variation from empirical datasets.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

The level of detail used in contemporary plankton ecological models varies considerably, from classic, simple nutrient-phytoplankton-zooplankton (NPZ) models with no representation of the microbial food web to models with a comprehensive coverage of known relevant chemical and biological food web components and interactions (Anderson, 2005; Vichi et al., 2007). Complexity comes at a price; e.g. in the biogeochemical flux model of Vichi et al. (2007), the modeler is faced with parameterizing model equations describing five or six state variables (chemical constituents), including a total of app. 20 group-specific parameters for every phytoplankton functional group included in the model. The increasing use of complex models in the 2000s was accompanied by cautionary comments that questioned the sufficiency of the ecological knowledge and empirical data used to support them (Anderson, 2005; Flynn, 2006). On the other hand, climate change and

other environmental threats and challenges of the modern world will clearly not wait for the acquisition of perfect scientific knowledge. Thus, models that are realistic enough to allow extrapolation of possible future data with at least some level of confidence are in urgent demand (Le Quere, 2006).

One suggested approach for walking the line between models too complex to be reliable or understandable and those too simple to be useful has been to start with parsimonious models that are transparent and biologically still feasible, while including only essential aspects of the ecosystem. With parsimonious (but not too simplistic) mechanistic models surprisingly complicated, sometimes even counter-intuitive perturbation response patterns of enclosed natural plankton communities can be reproduced (e.g. Thingstad et al., 2010 and Lignell et al., 2013, and references therein). These models can then be made stepwise more complex with thorough exploration of the importance of the new properties introduced at each step (Anderson, 2005; Thingstad et al., 2010; Shimoda and Arhonditsis, 2016). Moreover, such models require rigorous assessment of the adequacy of their structure and parameterization. For example, Franks (2009) reported a difference of more than two orders of magnitude in the half-saturation constant (K) values for both

* Corresponding author at: Marine Research Centre, Finnish Environment Institute, Helsinki, Finland.
E-mail address: veera.norros@ymparisto.fi (V. Norros).

ammonium (NH_4^+) and nitrate (NO_3) uptake of phytoplankton from 10 recent well-cited plankton ecosystem models. This variation in the values adopted for parameter K is much larger than any realistic natural variation in the efficiency of the nutrient uptake it is related to (e.g. Aksnes and Egge, 1991; Lignell et al., 2013). A possible explanation for this wide variation is that even in cases of new nutrient inputs, the net change in phytoplankton standing stocks is determined by the balance between growth and loss, and similar net changes can be reproduced in models by combining rapid growth (low K value) with rapid loss (e.g. via grazing) and correspondingly low growth (high K) and loss rates.

Recently, Shimoda and Arhonditsis (2016) evaluated the development practices and performance of 124 freshwater and marine plankton functional group (also called plankton functional type, PFT) models used in the past 30 years. They concluded that while there have been some improvements since previous criticisms, aquatic modelling studies still do not consistently follow conventional procedures, such as sensitivity analyses and validation. Moreover, the models generally do not perform very well in predicting empirical data. The authors recommended more rigorous practices, including the handling of uncertainty, and called for improved dialogue between empiricists and modelers.

We argue that Bayesian analysis with Markov chain Monte Carlo (MCMC) methods offers attractive solutions to the challenges in aquatic ecosystem modelling, such as in obtaining the 'right' model parameters, improving predictive performance and handling uncertainty. Bayesian MCMC methods provide a powerful and highly flexible tool for studying ecosystem functioning and identifying key parameters by fitting models describing biological processes to empirical data (Haario et al., 2006; Malve et al., 2007; Lignell et al., 2013). For instance, Malve et al. (2007) used a dynamic plankton model parameterized by an adaptive MCMC algorithm (Haario et al., 2006) to study algal blooms and concluded that a Bayesian approach with explicit handling of uncertainty can enable management-relevant conclusions to be made despite the presence of noisy data. Bayesian methods also have the advantage of straightforward combination of data from different sources (Clark, 2005). In the parameterization of aquatic ecosystem models, this can entail using laboratory measurements or theoretically derived values of phytoplankton nutrient uptake affinities as priors for food web model parameters estimated from time-series data (Lignell et al., 2013).

One of the simplifications potentially constraining aquatic ecosystem model improvement is the assumption that the parameters describing the processes modelled are constant, while in reality the processes likely vary geographically and temporally, e.g. due to variation in the species composition of the system (Zhao et al., 2008; Shimoda and Arhonditsis, 2016). Apart from the modelling challenge the variation presents, knowledge of the natural variation in ecosystem functioning is crucial for the reliable detection of trends in the state of the environment and for designing indicators of ecosystem status that reflect these long-term changes. Furthermore, variability in functioning is one of the facets of ecosystem stability and thus an interesting and important measure of ecosystem status in itself.

One way to deal with the variability in these biological/biogeochemical processes is to assume that the same mechanistic model describing the processes applies everywhere, while the values of its parameters may vary between subsystems, such as geographic locations. The simplest way to achieve this is to parameterize the mechanistic process model independently for each subsystem. However, the empirical data available for each individual subsystem may not be sufficient to allow adequate identification of the parameter values. A more derived solution is presented by hierarchical models in which the underlying mechanistic model is embedded in a parameter model that describes the variation in parameter values between subsystems. Hierarchical Bayesian models are particularly useful for quantifying different sources of variation in the system (Clark, 2005), including spatial and temporal variation (Thorson et al., 2015; Ovaskainen et al., 2016b; Thorson et al., 2016). The hierarchical structure makes it possible to include all the empirical data of interest in a single analysis and allows the

information on one subsystem (e.g. an intensively sampled species or location) to be generalized to other subsystems (e.g. scantly sampled species or locations) through shared higher-level parameters. Due to this exchange of information between subsystems, hierarchical models have the advantage over meta-analysis of independent model components, especially in cases where the data are sparse and/or unbalanced (Ovaskainen and Soininen, 2011).

Lignell et al. (2013) formulated a parsimonious Northern Baltic-PFT (NB-PFT) model with parameterization based on sound plankton ecological theory (e.g. Thingstad et al., 2010; Lignell et al., 2013 and references therein) and validation on comprehensive mesocosm data. They applied Bayesian inference to deal with uncertainties in data, parameterizations and model structure. Here, we extend this Bayesian approach to allow hierarchical model structure: a mechanistic food web model embedded in a statistical parameter model (see Fig. 1 for the general model structure). As the parameter model, we simply assume that the parameter value for each subsystem is a sample from a shared lognormal distribution. The goal of the modelling exercise is to illustrate and critically assess the usefulness of hierarchical Bayesian analysis in identifying parameter values, quantifying their variation and providing useful predictions of empirical data. Using both generated test data and natural data from the mesocosm experiment in Lignell et al. (2013), we compared full hierarchical analysis with two alternative Bayesian approaches: (1) global analysis: fitting one global model, i.e. a mechanistic food web model with a single set of parameter values to all subsystems (as in Lignell et al., 2013) and (2) separate analysis: fitting the same mechanistic food web model independently to each subsystem, without assuming that the parameter values for each subsystem come from a shared distribution. In our examples, the various subsystems are individual experimental units, such as mesocosms. The different analysis methods can also be seen as hierarchical models with the same mechanistic process model, but contrasting parameter models; in the global and separate analyses, we simply assume a global shared value or a set of independent values, respectively (Fig. 1B). We evaluated two different aspects of the different analysis methods: first, their accuracy and precision in identifying the parameters describing the functioning of the system; second, their predictive performance. We also discuss the feasibility and computational challenges in these approaches.

2. Material and methods

This study consists of two parts. First, we formulated a simple three-compartment model for the planktonic food web (NPZ; Fig. 2A). We fitted the NPZ model to data generated by the same model to compare the ability of the hierarchical, global and separate Bayesian analyses (Fig. 1B) to estimate the parameter values and to predict the data in a situation where we have complete knowledge and control over the true structure and parameter values of the system. In the second part, we adopted the more complex, but still parsimonious and transparent (understandable) plankton food web model of Lignell et al. (2013) (NB-PFT; Fig. 2B), which successfully reproduced the responses of the enclosed plankton communities to the experimental nutrient perturbations (Lignell et al., 2013). We used the NB-PFT model and experimental data from Lignell et al. (2013) to compare the performance of the hierarchical, global and separate analyses in a natural situation. Both the generated and the empirical data describe a situation in which the mass of the various food web components (nutrients, phytoplankton, zooplankton or their various functional groups) in the enclosed experimental units is followed over time. The experimental setup of both examples is described in Table 1, and the key parameters of both food web models are listed and explained in Table 2.

2.1. The nutrient-phytoplankton-zooplankton food web model

We began by testing the parameter estimation and predictive performance of the different analysis methods with synthetic observations