Environmental Modelling & Software 100 (2018) 74–81

Contents lists available at ScienceDirect

Environmental Modelling & Software

journal homepage: www.elsevier.com/locate/envsoft

Towards evidence-based parameter values and priors for aquatic ecosystem modelling

Barbara J. Robson [a, *], George B. Arhonditsis [b], Mark E. Baird [c], Jerome Brebion [a, d], Kyle F. Edwards [e], Leonie Geoffroy [a, f], Marie-Pier Hébert [g], Virginie van Dongen-Vogels [h], Emlyn M. Jones [c], Carla Kruk [i], Mathieu Mongin [c], Yuko Shimoda [b], Jennifer H. Skerratt [c], Stacey M. Trevathan-Tackett [h], Karen Wild-Allen [c], Xiangzhen Kong [j], Andy Steven [c]

[a] CSIRO Land and Water, GPO Box 1700, Canberra, ACT 2601, Australia
[b] University of Toronto, Canada
[c] CSIRO Oceans and Atmosphere, Hobart and Brisbane, Australia
[d] Ecole Nationale Supérieure d'Ingénieurs de Limoges, France
[e] University of Hawaii, USA
[f] Institut Polytechnique La Salle Beauvais, France
[g] McGill University, Montréal, QC, Canada
[h] Climate Change Cluster, University of Technology Sydney, Australia
[i] Limnology, Faculty of Sciences, CURE-Rocha, University of the Republic, Uruguay
[j] MOE Laboratory for Earth Surface Processes, College of Urban and Environmental Sciences, Peking University, Beijing 100871, PR China

ARTICLE INFO

Article history:
Received 9 June 2017
Received in revised form
1 October 2017
Accepted 8 November 2017

Keywords:
Parameterisation
Biogeochemical rates
Biological traits
Aquatic ecosystems
Parameter priors
Seagrass
Phytoplankton
Zooplankton
Remineralisation

ABSTRACT

Mechanistic models rely on specification of parameters representing biophysical traits and process rates such as phytoplankton, zooplankton and seagrass growth and respiration rates, organism sizes, stoichiometry, light, temperature and nutrient responses, nutrient-specific excretion rates and detrital stoichiometry and decay rates. Choosing suitable values for these parameters is difficult. Current practise is problematic. This paper presents a resource designed to facilitate an evidence-based approach to parameterisation of aquatic ecosystem models. An online tool is provided which collates relevant, published biological trait and biogeochemical rate observations from many sources and allows users to explore, filter and convert these data in a consistent, reproducible way, to find parameter values and calculate probability distributions. Using this information within a traditional or Bayesian paradigm should provide improved understanding of the uncertainty and predictive capacity of aquatic ecosystem models and provide insight into current sources of structural error in models.

Software and data availability

Software and data described in this paper are available from the Parameter Library Exploration Interface at http://shiny.csiro.au/CDM/parameterlibrary/latest/

1. Introduction

Aquatic ecosystem modelling has a long history of use to shed light on ecosystem function (e.g. Scheffer et al., 2001) and provide input to policy and management decisions to improve water quality, limit the impact of pollutants and toxicants, specify fish catches, inform management of catchments and riparian vegetation, and control eutrophication (e.g. Wang et al., 2012).

Although there are several approaches to simulation of environmental systems (Robson, 2014), process-based models, including complex mechanistic models, have been enduringly popular. These models aim to mimic the roles of individual processes through mathematical descriptions with realistic parameterisations. The premise is that mathematical representations of individual processes combine to collectively offer a faithful

---

* Corresponding author.
E-mail address: bjrobson@gmail.com (B.J. Robson).

https://doi.org/10.1016/j.envsoft.2017.11.018
1364-8152/

depiction of ecosystem dynamics. The appeal of this approach is that by simulating systems in terms of real, ideally measurable physical, geochemical and biological processes, mechanistic models offer the prospect of greater insight into system function than other types of models, as well as the promise of defensible extrapolation beyond historical conditions to predict future changes. The degree to which a mechanistic model can deliver on these promises depends not only on the accuracy of the conceptual understanding of the system represented by the equations used to build the model, but also on the accuracy of its parameterisation.

2. Current state of the art

Parameterisation of aquatic ecosystem models has historically been a difficult and labour-intensive task. Parameter values are typically required for a range of model components, defining traits and responses of phytoplankton, zooplankton and other aquatic biota, characteristics of sediments and detrital material, physical and chemical process rates, and, more recently, bacterial traits and processes. Although mathematical methods have long been available to automate parameter estimation and calibration, these have not been rapidly or universally adopted in this field due to computational constraints or data limitations (Shimoda and Arhonditsis, 2016).

Traditional approaches to setting parameter values have been either to combine (where available) local measurements of flux rates and component responses for the study site with literature values for parameters that have not been measured locally (e.g. Mao et al., 2015; Murray and Parslow, 1999; Robson and Hamilton, 2004); or to begin with the values that have been set during a previous application of the model or a similar model, and to calibrate from there to achieve an acceptable model-data match (e.g. Skerratt et al., 2013). Issues with these approaches are:

- It is often difficult to find relevant parameter values in the literature. While phytoplankton growth rates are commonly measured, there are relatively few available measurements relevant to some other biogeochemical processes (e.g. zooplankton mortality).
- It can be difficult to ascertain how relevant a particular literature value might be. For instance, is a published maximum growth rate for a *Microcystis aeruginosa* strain isolated from a temperate lake relevant to a simulation of *Microcystis aeruginosa* in a subtropical estuary?
- It can be difficult or time-consuming to convert from experimental results published in the literature to the specific parameters required, under the required conditions and in the required units.
- The process is usually poorly documented, so that the evidence from which a parameter value was derived cannot readily be traced, and the strength of that evidence cannot be assessed.
- Parameter uncertainty is poorly understood. While a sensitivity analysis might tell us the degree to which variation in the parameter value affects model results, we retain little information about the degree of confidence we have in the "correct" value of that parameter. If calibration of the model results in a shift away from an initially hypothesised parameter value, it is difficult to assess whether this shift is reasonable, or whether the calibration is compensating for an error in another parameter or a structural error in the model.

Many of these issues arise from the fact that ecosystem models are usually representing aggregate processes, but the ultimate determinants of these processes are the traits of individual organisms. Models may not represent how these traits change as the environment causes changes in community structure.

In practise, there is considerable inconsistency in the parameter values used to represent process rates in similar models applied across systems or even within the same system. A characteristic example can be found in the Lake Erie modelling work reviewed by Kim et al. (2014) and Shimoda and Arhonditsis (2016), where equivalent phytoplankton functional groups have been characterised as having low (<2 day^{-1}) or high (3 day^{-1}) maximum growth rates, depending on the model considered.

There are a range of reasons for this variability in parameter values. In the case of phytoplankton growth rates, for example, we may consider:

- The maximum growth rate of a phytoplankton group typically aims to characterise the average patterns of diverse assemblages of functionally similar phytoplankton species and strains. The actual composition of these assemblages will vary in time and space.
- Different ways of clustering species into functional, phylogenetic or morphological groups might result in different model parameters even with the same initial information.
- The maximum growth rate of a particular phytoplankton strain depends on its history and the conditions to which it is adapted.
- The concept of a phytoplankton growth rate is a simplification of several physiological processes, each of which may vary separately: these include carbon fixation (photosynthesis), chlorophyll production and cell division. Some of the variability in the "maximum growth rate" parameter will be due to actual variability in carbon to chlorophyll ratios.
- Different approaches to calibration of models and different goal metrics will yield different optimal parameter values.
- Some models are over-parameterised relative to the available data, so that no single optimal calibrated parameter set exists.
- Different models embody different conceptualisations of the system, and every model has some degree of structural error. Variations in assigned values may compensate for structural errors. For example, a low maximum phytoplankton growth rate may compensate for a missing loss term such as viral lysis, while a high maximum growth rate may compensate for a missing source term such as germination of akinetes.

Hence, some of the variability in parameter values is appropriate in terms of our physical and physiological understanding of the system, while some of the variability may be due to errors in the structure or calibration of models. Insofar as we implicitly acknowledge this when we develop a model, we treat parameters not strictly as measurable physiological properties, but rather as mathematical "properties of convenience." While treating parameters in this way may produce good model predictions in some circumstances, it reduces the connection between models and the processes they claim to represent, the predictive capability of the model, and our ability to anticipate the circumstances in which our models may fail (Kruk et al., 2011; Ramin et al., 2011).

3. Collation of parameter values

There have been several efforts to collate and tabulate parameter values from across the literature. Some of these have been broadly focused on the most commonly used parameters in aquatic ecosystem models (e.g. Bowie et al., 1985) or eco-chemistry (Jorgensen et al., 2000), while others have comprehensively reviewed the literature for a narrower range of parameters relating to a particular component or process of an aquatic ecosystem (e.g. phytoplankton: Edwards et al., 2015; Eppley, 1985; zooplankton: Hébert et al., 2016; phytoplankton: Hoogenhout and Amesz, 1965; seagrass: Lee et al., 2007; coral: Madin et al., 2016), in some cases synthesising this information to derive insight into allometric and ecological functional relationships (Droop et al., 1982; Edwards et al., 2012, 2015; Enriquez et al., 1993; Hansen et al., 1997; Hebert et al., 2016; Kruk et al., 2010; Litchman et al., 2007). These sources have been, and remain, a very valuable resource for modellers.

A natural progression from the disparate reporting of parameter values is to collate them into a probabilistic form. Collated process rates or parameter values from diverse studies may be summarised in the form of a probability distribution whereby some parametric distribution is fitted to the reported values. If a Bayesian nomenclature is adopted, these distributions are typically referred to as “prior probability” distributions (e.g. Arhonditsis et al., 2008; Jones et al., 2010; Zhang and Arhonditsis, 2009). Bruggeman (2011) considers phytoplankton traits in a probabilistic form, predicting unknown values from known values for related species and phylogenetic relationships between species.

In this paper, we build on this legacy by:

- Bringing together parameter values from a wide range of sources, including past review papers as well as individual peer-reviewed experimental and observational research papers.
- Introducing an online tool (http://shiny.csiro.au/CDM/parameterlibrary/latest/) to facilitate exploration and visualisation of these data, allowing the data to be downloaded, filtered, combined, and manipulated in various common ways.
- Presenting parameter data in a form that facilitates its use within a Bayesian framework, for instance, allowing the user to easily fit probability distributions and calculate key statistics such as the 5th, 50th and 95th percentiles.
- Introducing a community initiative to maintain and expand this database so that it will continue to grow and become more useful in a wider range of modelling contexts over time.
- Presenting some guidelines for a more evidence-based approach to parameter specification in aquatic ecosystem models.

While the database is not yet comprehensive, the current version comprises 4834 observations relating to traits and responses of marine and freshwater phytoplankton, zooplankton, detritus, seagrasses, coral and microphytobenthos. Such a data synthesis effort not only identifies some of the recurrent knowledge gaps in the literature but it is our hope that it will also encourage aquatic researchers to contribute to this collaborative project by disclosing and sharing measurements. As such, we have focused on several key parameters that are relevant to both marine and freshwater modelling, selecting parameters for inclusion on the basis that they are directly relevant to current modelling projects.

For each parameter, we have found measured values in the literature, drawing on previous reviews where appropriate, but also checking original sources where possible to minimise propagation of errors from reviews, and bringing in new data from more recent studies. We have included only parameter estimates derived from peer-reviewed experimental or observational studies, not parameter values derived from calibration of models. Although a review of (posterior) parameter values actually used in modelling would be an interesting topic for another study, introducing calibrated model values to this database might bias the distributions away from the empirical evidence base.

4. Online interface

We have implemented a preliminary online user interface to allow further exploration of the parameter library, which can be found at http://shiny.csiro.au/CDM/parameterlibrary/latest/. A screenshot is presented as Fig. 1. In addition to calculating and visualising probability distributions (gamma, normal, log-normal or uniform) fitted to each parameter, the interface allows users to filter datasets where relevant information exists in the database. In most cases, there is an option to restrict the dataset to include only data-points for which we have independently checked that the value given can be found in the original data source (in other cases, we have relied on a third party such as a review article).

The user can view or download a plot and statistical description of the parameter distribution as well as the relevant raw data, including the full reference and DOI of the source publication and the review article (if any) in which it was found and the name and contact details of the person who contributed each data point to the database.

Histograms show the actual distribution of observations. The number of bins used in each histogram depends on the total number of observational values included: wider bins are used when fewer measurements are available to provide a visual indication of the strength of evidence.

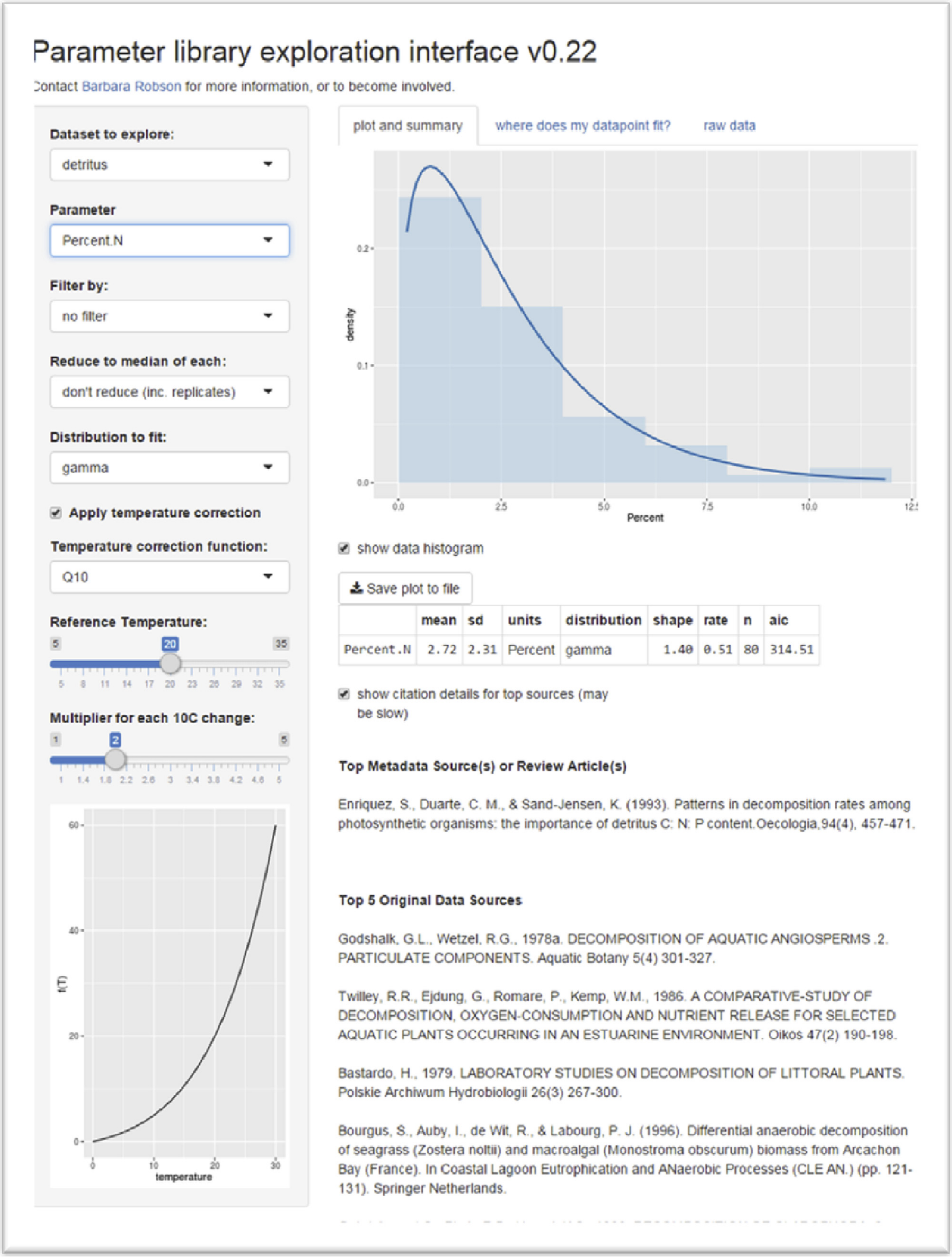
The online tool also provides an option, labelled “reduce to median of each:” that allows the user to consolidate replicated measurements to avoid over-weighting results for species (or other filter options) which are over-represented in the data-set. For instance, a user may choose to consolidate by species, in which case a single (median) value is used for each species for which there is more than one data-point.

The user can choose the assumptions that are made about the temperature response of biogeochemical rates, choosing between (a) no temperature correction, (b) a Q10 or Arrhenius response curve or (c) the response curve implemented in CAEDYM (Robson and Hamilton, 2004; after Griffin et al., 2001), which allows inhibition at higher temperatures.

The Arrhenius and Q10 response curves are equivalent. In the “Arrhenius” option, $K_T = K_{20}\nu^{\left(\frac{T}{T-20}\right)}$, where K_T is the value of a parameter at a specified temperature, T (°C), K_{20} is the value of that parameter at 20 °C and ν is a specified constant (Goldman and Carpenter, 1974). In the “Q10” option, (e.g. Volta et al., 2014), Q is specified as a multiplier applied to a parameter value for each 10 °C increase in temperature.

In its current (beta) form, the online database contains a relatively small selection of the data available in the literature, and in some cases, the initial sources are skewed towards particular contexts relevant to the current projects of the authors. With community contributions, the depth and breadth of parameter coverage will be improved over time. This may potentially include physical parameters such as particle sizes and roughness lengths and aggregate biogeochemical process rates such as sediment oxygen demand and oxygen transfer rates as well as the type of biological rates and traits considered so far.

In addition to making use of the available filter options, we encourage users of the online tool to explore the original data sources to get a feel for the conditions under which measurements were made and form their own judgement about the relevance of these conditions to their own studies. In some circumstances, it might be appropriate to recalculate prior distributions, giving more weight to measurements made using more reliable techniques, or in field versus laboratory conditions, or on the basis of other considerations.



Parameter library exploration interface v0.22

Contact Barbara Robson for more information, or to become involved.

Dataset to explore: detritus

Parameter: Percent.N

Filter by: no filter

Reduce to median of each: don't reduce (inc. replicates)

Distribution to fit: gamma

☑ Apply temperature correction

Temperature correction function: Q10

Reference Temperature: 20

Multiplier for each 10C change: 2

☑ show data histogram

	mean	sd	units	distribution	shape	rate	n	aic
Percent.N	2.72	2.31	Percent	gamma	1.40	0.51	80	314.51

☑ show citation details for top sources (may be slow)

Top Metadata Source(s) or Review Article(s)

Enriquez, S., Duarte, C. M., & Sand-Jensen, K. (1993). Patterns in decomposition rates among photosynthetic organisms: the importance of detritus C: N: P content.Oecologia,94(4), 457-471.

Top 5 Original Data Sources

Godshalk, G.L., Wetzel, R.G., 1978a. DECOMPOSITION OF AQUATIC ANGIOSPERMS .2. PARTICULATE COMPONENTS. Aquatic Botany 5(4) 301-327.

Twilley, R.R., Ejdung, G., Romare, P., Kemp, W.M., 1986. A COMPARATIVE-STUDY OF DECOMPOSITION, OXYGEN-CONSUMPTION AND NUTRIENT RELEASE FOR SELECTED AQUATIC PLANTS OCCURRING IN AN ESTUARINE ENVIRONMENT. Oikos 47(2) 190-198.

Bastardo, H., 1979. LABORATORY STUDIES ON DECOMPOSITION OF LITTORAL PLANTS. Polskie Archiwum Hydrobiologii 26(3) 267-300.

Bourgus, S., Auby, I., de Wit, R., & Labourg, P. J. (1996). Differential anaerobic decomposition of seagrass (Zostera noltii) and macroalgal (Monostroma obscurum) biomass from Arcachon Bay (France). In Coastal Lagoon Eutrophication and ANaerobic Processes (CLE AN.) (pp. 121-131). Springer Netherlands.

Fig. 1. Screen capture showing the preliminary parameter library exploration interface available at http://shiny.csiro.au/CDM/parameterlibrary/latest.

5. Summary of parameter data

Table 1 summarises the data included in the parameter library as at May 2017. We anticipate that we will add to this resource over time, drawing on community contributions.

5.1. Detritus

Our initial detritus dataset currently includes measured values for detrital composition (%C, %N and %P), C:N and C:P, remineralisation/decay rates, and decay rates of labile and refractory fractions where rates presented in observational papers were calculated for two fractions. A key resource was the work of Enriquez et al. (1993), who collated many of the pre-1993 data. Several filters are offered, including detritus source (aquatic or terrestrial material), plant type of source material (amphibious plant, freshwater angiosperm, grass, macroalgae, mangrove, phytoplankton, seagrass, sedge or shrub), genus, species, plant fraction (many categories including leaf, stem or roots), conditions (e.g. buried, high intertidal, low intertidal, subtidal, water), climate (e.g. coastal Mediterranean, Temperate, Tropical Wet and Dry), geographic location and type of water body (e.g. lake, river, wetland, bay, estuary, sea).

5.2. Zooplankton and other grazers

The database currently includes measured zooplankton growth rates and individual animal volumes and lengths, growth efficiency, ingestion, clearance and respiration rates, and nutrient-specific

Table 1
Overview of the number of data points and related sources available to date (*i.e.* May 2017) for each parameter collected for distinct aquatic ecological compartments.

Ecological compartment	Parameters	Total data points	Number of original sources
Coral	Uptake rate constant	54	1
Detritus	Percent N, P, C	174, 84, 116	55, 30, 36
	C to N, C to P ratios	162, 49	40, 18
	Decay rate (K)	224	63
	Decay rate of labile fraction (K1)	12	7
	Decay rate of refractory fraction (K2)	12	7
	Bacterial growth rate	4	1
Zooplankton	Body mass	830	52
	Body length	109	15
	Individual volumes	20	10
	Corporal content N, P, C	753, 661, 672	14, 28, 19
	Growth efficiency	40	34
	Growth rates (umax)	72	42
	Maximum specific growth rates	72	42
	Maximum ingestion rates	60	41
	Maximum clearance rates	60	41
	Excretion rates N, P	451, 154	12, 6
	Individual hourly O2 consumption rates	792	26
	Individual hourly respiration rates	78	1
	Individual Ash free dry weight	78	1
	Individual respiration rates	78	1
	Estimated Zm_Q	78	1
	Hourly respiration rates	78	1
	Specific Respiration rates	78	1
Phytoplankton	Cell volume	827	40
	Cell surface area	734	4
	Cell length mLD	711	3
	Maximum growth rate	1044	227
	Any growth rate	35	19
	Respiration rates	30	4
	Carbon per cell	17	13
	Growth rate for NH4, NO3, PO4 limited growth	6, 22, 49	5, 13, 28
	Max growth rate for NH4, NO3, PO4 limited growth	50, 51, 117	19, 29, 41
	Km for NH4, NO3, PO4 limited growth	20, 31, 79	6, 15, 27
	K for NH4, NO3, PO4 uptake	12, 80, 139	10, 22, 44
	Vmax for NH4, NO3, PO4 uptake	26, 78, 144	13, 22, 48
	Vmax to C for NH4, NO3, PO4 uptake	10, 6, 7	4, 5, 5
	Qmin for NH4, NO3, PO4 limited growth	19, 68, 162	15, 38, 76
	Qmin to C for NH4, NO3, PO4 limited growth	4, 11, 22	3, 9, 13
	Qmax for NH4, NO3, PO4 limited growth	14, 18, 61	3, 7, 27
	Qmax to C for NH4, NO3, PO4 limited growth	0, 2, 11	0, 2, 6
Phytoplankton PI	Alpha, Beta, Pmax	410, 234, 410	11 (22)
	AlphaB, BetaB, PmaxB	1672, 1060, 1671	11 (22)
	Errors for AlphaB, BetaB, PmaxB	638, 393, 638	11 (22)
	Ek, EkB	410, 515	6 (8)
	Intercept	1363	7 (15)
Microphyto-benthos	Growth rates	35	19
Seagrasses	Photosynthesis-irradiance curves (Pmax)	132	40
	Compensation irradiance (Ic)	99	29
	Saturation irradiance (Ik)	99	32
	Photosynthetic quantum efficiency (alpha)	105	31
	Optimal temperature for growth	37	27

excretion artes from a range of freshwater, marine and estuarine studies. Key resources included Hébert et al. (2016) for crustacean sizes and Gsell et al. (2016) for rotatorians. For zooplankton respiration rates, 78 measured values were extracted from Ikeda (1974), and the remainder of the respiration data (i.e. 714 observations) comes from the data compilation of Hébert et al. (2016). Other key review articles and data collations for zooplankton included (Hansen et al., 1991, 1997; Hebert et al., 2016).

In addition to zooplankton, the dataset currently includes parameter values to define coral uptake rates of 6 distinct particle types including heterotrophic bacteria, picophytoplankton (synechococcus, picoeukaryotes), nanoeukaryotes, microphytoplankton, and particulate organic nitrogen. These data were taken from values measured by Ribes et al. (2003).

The database does not yet include other animals such as benthic macroinvertebrates. This is an area for future expansion and we would welcome assistance.

5.3. Phytoplankton and benthic microalgae

For phytoplankton maximum growth rates, we draw upon several important synthesis papers and data papers on this topic, including (Edwards et al., 2015; Hoogenhout and Amesz, 1965; Kruk et al., 2010). These can be filtered by morphologically based functional group (as defined by Kruk et al., 2010), by type of water (fresh, euryhaline, estuarine, marine, hypersaline), source of strain (e.g. lake, river, estuary, ocean, sewage treatment lagoon, laboratory culture), phylogenetic domain, class, genus or species. We also have data for a range of parameters relating to phytoplankton nutrient uptake, nutrient use, and carbon content per cell (derived largely from Edwards et al., 2015).

For phytoplankton respiration, we have 30 data points from 4 literature sources.

Phytoplankton photosynthesis-irradiance (P vs. I) curve parameters have been measured across diverse oceanographic conditions (freshwater photosynthesis-irradiance data have not yet been collated). Data are taken so far from 22 distinct oceanographic voyages covering regions of the Southern Ocean, the Tasman Sea, Great Barrier Reef, Equatorial waters, and Australian coastal waters. The relatively high number of data points for each parameter is due to the multiple depths sampled for each studied location.

Microphytobenthos (benthic microalgae) parameter data are currently limited to a few sources, including growth rates for 31 species of microphytobenthos in 5 taxonomic groups.

5.4. Seagrasses

For seagrasses, the database currently contains parameters to describe photosynthesis-irradiance curves and parameters to describe optimal temperatures for seagrass growth. The majority of these are sourced from the review presented by Lee et al. (2007).

6. Making effective use of parameter prior distributions

The simplest approach to using the information presented via this online resource may be to select the subsample of data most relevant to the application at hand, plot a distribution, and calibrate within the range suggested by the 5th and 95th percentiles. After calibration, the final parameter set can be compared with the expected prior distributions. If more than a few calibrated parameter values fall close to the outer limits of the expected distribution, this may warrant further investigation, as it may be suggestive of a structural error in the model or a systematic error in the observational data (Fig. 2).

If multiple models applied to different systems consistently produce calibrated values outside or near the outer limits of the expected distributions, this may indicate either a common problem in model conceptualisation or a problem in the way that the relevant biophysical metric or rate has been measured or reported in the literature. If the problem is a common structural error in models, then a focus on improving this component of our models may yield substantial improvements in predictive performance. Conversely, if the problem lies in the biophysical rate data, it may indicate either that more measurements are required to properly characterise this parameter (perhaps using a wider variety of species from a wider variety of aquatic systems or systems more relevant to the contexts of model application) or that better or more consistent measurement protocols are needed to ensure that measurements are relevant to real-world conditions.

For example, Fig. 3 compares the distribution of observed maximum growth rates of freshwater diatoms with the distribution of calibrated values found in 72 published models from the literature (Shimoda and Arhonditsis, 2016). Although all 72 models use growth rates that are within the range of observed values, values at

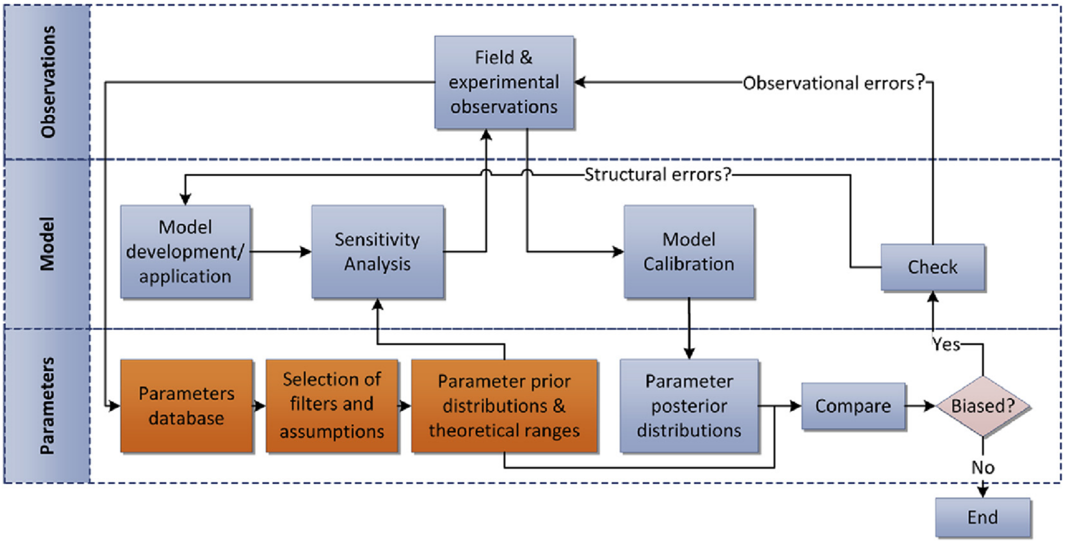


Fig. 2. Suggested use of the parameter database and information provided in this paper in a modelling process.

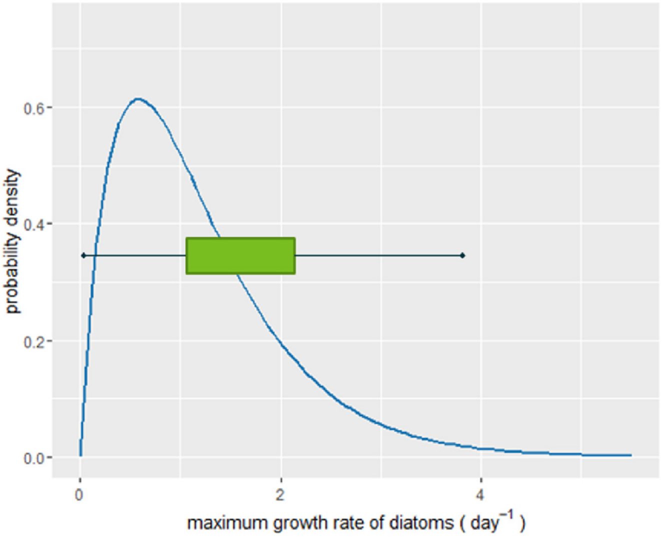


Fig. 3. Comparison between the distribution of maximum growth rates of diatoms from 143 observations in the parameter database, adjusted to 20 °C assuming a rate that doubles with every 10 °C increase in temperature (line) with maximum growth rates used for diatoms in 72 models from the published literature (Shimoda and Arhonditsis, 2016). The green bar shows the 25th to 75th percentile of the model values; the line extending from the bar shows the full range of values. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

either extreme are either too high or too low to be likely. Further, the preponderance of values used in models are above the median of observed rates. This suggest that either there is a systematic bias in measured rates, or a systematic bias in the models: for example, it may be that most of these models over-estimate a loss term (such as settling to sediments) or are missing an important source term for diatoms, so an elevated growth rate is compensating for this error. Both possibilities present opportunities for further investigation, which might lead to an improved next generation of models.

A more powerful approach is to use the parameter priors within a Bayesian modelling framework. The field of physical-statistical modelling (Berliner, 2003; Parslow et al., 2013) describes the combination of mechanistic models with statistical methods. The Bayesian Hierarchical Modelling (BHM) framework (see Cressie and Wikle, 2011, for a thorough introduction to the subject) can be used to account for multiple sources uncertainty using conditional probability distributions, to simultaneously provide estimates of parameter values and estimates of uncertainty associated with those values and with the resulting model predictions. According to Bayes law, the posterior distribution for a BHM is given by:

$$[X,\theta|Y] \sim [Y|X,\theta][X|\theta][\theta] \tag{1}$$

where X are the state variables of the model, θ are the parameters and Y are the observations. The square brackets, $[\theta]$ denotes the probability distribution of θ, and $[X|\theta]$ is the conditional probability of X given θ. The key elements of eqn. (1) are: $[\theta]$, the prior model, $[X|\theta]$, the process model and $[Y|X,\theta]$, the data model. In many cases, there is no analytical solution to the posterior, and therefore sampling-based methods such as Bayesian Monte Carlo Markov Chain (MCMC) analysis must be used (Andrieu et al., 2010; Jones et al., 2010; Murray et al., 2013; Saloranta et al., 2008).

The formal inclusion of prior knowledge of parameters within a BHM occurs via the prescription of $[\theta]$. In the absence of adequate information, the prior over $[\theta]$ may take a very uninformative form. In the most extreme case, $[\theta]$ could be sampled from a uniform distribution spanning $[-\infty,\infty]$. However, even subjective information using expert opinion has been shown to reduce uncertainty by placing finite bounds. From the results presented in this paper, there is substantial information available to place parametric distributions on some parameters, rendering them in many cases to be quite informative.

To date, there have been a handful of studies that have exploited the various components of the BHM framework within aquatic and marine biogeochemical modelling. Dowd et al. (2014) reviewed recent developments applications. With the exception of Parslow et al. (2013), very few studies have included a rigorous derivation of prior distributions. Parslow et al. (2013) derive a number of model-dependent prior distributions, however there is a need to extend this to model-independent parameterisations that cover a broad range of parameterisation and environments.

7. Future development

The beta version of the parameter database presented here and currently available online is a static resource, manually updated. It is our intention (subject to funding) to maintain and update this to provide a modern data service. This will include not only expansion of the database itself and improvement of the online user interface, but also improvement of the underlying data structure, consideration of issues of semantics, inter-operability, provenance and data delivery, and provision of an API that will allow third parties to access the data in novel ways. This will, for instance, allow modellers to embed calls to the database within automated calibration and data assimilation routines, or to develop tools to extract tailored datasets, apply specified transformations to the data, and automatically produce parameter files in the format needed by a specific model.

It is anticipated that a future version of the online interface will allow registered users to create accounts, save selected filters or transformations, share generated prior distributions by generating unique addresses to the results, and — most importantly — contribute by uploading their own new data points or datasets. In the meantime, we encourage anyone interested in contributing to contact the authors.

Future work will also include an analysis of the distribution of parameter values actually used in models and posterior distributions obtained by applying Bayesian hierarchical methods in combination with the prior distributions made available through the online tool. This should provide insight into potential structural issues in the current generation of models as well as priorities for additional observational measurements.

The current version of the database considers point values from various sources, but does not take into account uncertainty estimates given by some sources. This is not straightforward, due to the variability in the type of ranges given by primary sources (confidence intervals, standard errors, standard deviations and absolute ranges), but is something that should be addressed in future to minimise loss of information.

8. Conclusion

By presenting physiological trait and biogeochemical process rate information through an online tool that not only synthesises observations from multiple sources but also supports consistent and reproducible processing of these data (for instance, by applying a consistent temperature correction function to observed metabolic rates), we facilitate more evidence-based Bayesian parameterisation of aquatic systems models as well as improved model evaluation and development processes.

Acknowledgements

This work was funded in part through the eReefs project, a public-private collaboration between Australia's leading operational and scientific research agencies, government, and corporate Australia. JB and LG contributed to this work during internships at CSIRO Land and Water, completed as part of their studies. Thanks to Philip Balding for assisting with phytoplankton data entry.

Author contributions

BJR led the conceptualisation and development of the parameter library, planning and writing of the paper, and wrote the R code for the (currently beta) online resource. MEB, KWA, MM and AS contributed to the planning and design of the parameter library. BJR, GBA, EJ, XSK, MPH and MEB wrote the manuscript. XSK drafted a section on use of the parameter library in a modelling workflow and prepared Fig. 2. GBA provided the analysis behind Fig. 3. KFE collated data relating to phytoplankton nutrient utilisation. CK, JHS, JB and BJR collated data relating to phytoplankton growth, nutrient utilisation and cell sizes. MPH collated, analysed and formatted data related to zooplankton size and metabolic rates. JB and BJR collated other zooplankton data. STT and LG collated parameters relating to seagrass and detritus, including decay rates. VDV collated data relating to phytoplankton photosynthesis-irradiance curves, performed a QA/QC check of each variable related to these data, and prepared Table 1. YS collated and prepared metadata.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.envsoft.2017.11.018.

References

Andrieu, C., Doucet, A., Holenstein, R., 2010. Particle Markov chain Monte Carlo methods. J. R. Stat. Soc. Ser. B Stat. Methodol. 72, 269–342.

Arhonditsis, G.B., Papantou, D., Zhang, W.T., Perhar, G., Massos, E., Shi, M.L., 2008. Bayesian calibration of mechanistic aquatic biogeochemical models and benefits for environmental management. J. Mar. Syst. 73 (1–2), 8–30.

Berliner, L.M., 2003. Physical-statistical modeling in geophysics. J. Geophys. Res. Atmos. 108 (D24).

Bowie, G.L., Space Environment Laboratory (U.S.), Tetra Tech Inc, 1985. Rates, Constants, and Kinetics Formulations in Surface Water Quality Modeling, second ed. Environmental Research Laboratory, Office of Research and Development, U.S. Environmental Protection Agency, Athens, Ga.

Bruggeman, J., 2011. A phylogenetic approach to the estimation of phytoplankton traits. J. Phycol. 47 (1), 52–65.

Cressie, N.A.C., Wikle, C.K., 2011. Statistics for Spatio-temporal Data. Wiley, Hoboken, N.J.

Dowd, M., Jones, E., Parslow, J., 2014. A statistical overview and perspectives on data assimilation for marine biogeochemical models. Environmetrics 25 (4), 203–213.

Droop, M.R., Mickelson, M.J., Scott, J.M., Turner, M.F., 1982. Light and nutrient status of algal cells. J. Mar. Biol. Assoc. U. K. 62 (2), 403–434.

Edwards, K.F., Thomas, M.K., Klausmeier, C.A., Litchman, E., 2012. Allometric scaling and taxonomic variation in nutrient utilization traits and maximum growth rate of phytoplankton. Limnol. Oceanogr. 57 (2), 554–566.

Edwards, K.F., Thomas, M.K., Klausmeier, C.A., Litchman, E., 2015. Light and growth in marine phytoplankton: allometric, taxonomic, and environmental variation. Limnol. Oceanogr. 60 (2), 540–552.

Enriquez, S., Duarte, C.M., Sandjensen, K., 1993. Patterns in decomposition rates among photosynthetic organisms - the importance of detritus C-N-P content. Oecologia 94 (4), 457–471.

Eppley, R.W., 1985. Citation classic - temperature and phytoplankton growth in the sea. Curr. Contents/Agric. Biol. Environ. Sci. (37), 20.

Goldman, J.C., Carpenter, E.J., 1974. A kinetic approach to the effect of temperature on algal growth. Limnol. Oceanogr. 19 (5), 756–766.

Griffin, S.L., Herzfeld, M., Hamilton, D.P., 2001. Modelling the impact of zooplankton grazing on phytoplankton biomass during a dinoflagellate bloom in the Swan River Estuary, Western Australia. Ecol. Eng. 16, 373–394.

Gsell, A.S., Özkundakci, D., Hébert, M.-P., Adrian, R., 2016. Quantifying change in pelagic plankton network stability and topology based on empirical long-term data. Ecol. Indic. 65, 76–88. https://doi.org/10.1016/j.ecolind.2015.11.014.

Hansen, B., Hansen, P.J., Nielsen, T.G., 1991. Effects of large nongrazable particles on clearance and swimming behavior of zooplankton. J. Exp. Mar. Biol. Ecol. 152 (2), 257–269.

Hansen, P.J., Bjornsen, P.K., Hansen, B.W., 1997. Zooplankton grazing and growth: scaling within the 2-2,000-mu m body size range. Limnol. Oceanogr. 42 (4), 687–704.

Hébert, M.-P., Beisner, B.E., Maranger, R., 2016. A compilation of quantitative functional traits for marine and freshwater crustacean zooplankton. Ecology 97.

Hebert, M.P., Beisner, B.E., Maranger, R., 2016. A meta-analysis of zooplankton functional traits influencing ecosystem function. Ecology 97 (4), 1069–1080.

Hoogenhout, H., Amesz, J., 1965. Growth rates of photosynthetic microorganisms in laboratory cultures. Arch. Mikrobiol. 50 (1), 10–25.

Ikeda, T., 1974. Nutritional Ecology of Marine Zooplankton. Hokkaido University Faculty of Fisheries, pp. 1–97.

Jones, E., Parslow, J., Murray, L., 2010. A Bayesian approach to state and parameter estimation in a Phytoplankton-Zooplankton model. Aust. Meteorol. Oceanogr. J. 59, 7–15.

Jorgensen, S.E., Nielsen, S.N., Jorgensen, L.A., 2000. Ecotox Ecological Modelling and Ecotoxicology. Elsevier Science,, Amsterdam, p. 1 online resource.

Kim, D.K., Zhang, W.T., Watson, S., Arhonditsis, G.B., 2014. A commentary on the modelling of the causal linkages among nutrient loading, harmful algal blooms, and hypoxia patterns in Lake Erie. J. Gt. Lakes. Res. 40, 117–129.

Kruk, C., Huszar, V.L.M., Peeters, E.T.H.M., Bonilla, S., Costa, L., Lurling, M., Reynolds, C.S., Scheffer, M., 2010. A morphological classification capturing functional variation in phytoplankton. Freshw. Biol. 55 (3), 614–627.

Kruk, C., Peeters, E.T.H.M., Van Nes, E.H., Huszar, V.L.M., Costa, L.S., Scheffer, M., 2011. Phytoplankton community composition can be predicted best in terms of morphological groups. Limnol. Oceanogr. 56 (1), 110–118.

Lee, K.S., Park, S.R., Kim, Y.K., 2007. Effects of irradiance, temperature, and nutrients on growth dynamics of seagrasses: a review. J. Exp. Mar. Biol. Ecol. 350 (1–2), 144–175.

Litchman, E., Klausmeier, C.A., Schofield, O.M., Falkowski, P.G., 2007. The role of functional traits and trade-offs in structuring phytoplankton communities: scaling from cellular to ecosystem level. Ecol. Lett. 10 (12), 1170–1181.

Madin, J.S., Anderson, K.D., Andreasen, M.H., Bridge, T.C., Cairns, S.D., Connolly, S.R., Darling, E.S., Diaz, M., Falster, D.S., Franklin, E.C., Gates, R.D., 2016. The Coral Trait Database, a curated database of trait information for coral species from the global oceans. Sci. Data 3.

Mao, J.Q., Jiang, D.G., Dai, H.C., 2015. Spatial-temporal hydrodynamic and algal bloom modelling analysis of a reservoir tributary embayment. J. Hydro Environ. Res. 9 (2), 200–215.

Murray, A.G., Parslow, J.S., 1999. Modelling of nutrient impacts in Port Phillip Bay - a semi-enclosed marine Australian ecosystem. Mar. Freshw. Res. 50 (6), 597–611.

Murray, L.M., Jones, E.M., Parslow, J., 2013. On disturbance state-space models and the particle marginal Metropolis-Hastings sampler. SIAM/ASA J. Uncertain. Quantification 1 (1), 494–521.

Parslow, J., Cressie, N., Campbell, E.P., Jones, E., Murray, L., 2013. Bayesian learning and predictability in a stochastic nonlinear dynamical model. Ecol. Appl. 23 (4), 679–698.

Ramin, M., Stremilov, S., Labencki, T., Gudimov, A., Boyd, D., Arhonditsis, G.B., 2011. Integration of numerical modeling and Bayesian analysis for setting water quality criteria in Hamilton Harbour, Ontario, Canada. Environ. Model. Softw. 26 (4), 337–353.

Ribes, M., Coma, R., Atkinson, M.J., Kinzie Iii, R.A., 2003 Aug 7. Particle removal by coral reef communities: picoplankton is a major source of nitrogen. Mar. Ecol. Prog. Ser. 257, 13–23.

Robson, B.J., 2014. When do aquatic systems models provide useful predictions, what is changing, and what is next? Environ. Model. Softw. 61, 287–296.

Robson, B.J., Hamilton, D.P., 2004. Three-dimensional modelling of a Microcystis bloom event in the Swan river estuary, western Australia. Ecol. Model. 174 (1–2), 203–222.

Saloranta, T.M., Armitage, J.M., Haario, H., Naes, K., Cousins, I.T., Barton, D.N., 2008. Modeling the effects and uncertainties of contaminated sediment remediation scenarios in a Norwegian Fjord by Markov chain Monte Carlo simulation. Environ. Sci. Technol. 42 (1), 200–206.

Scheffer, M., Carpenter, S., Foley, J.A., Folke, C., Walker, B., 2001. Catastrophic shifts in ecosystems. Nature 413 (6856), 591–596.

Shimoda, Y., Arhonditsis, G.B., 2016. Phytoplankton functional type modelling: running before we can walk? A critical evaluation of the current state of knowledge. Ecol. Model. 320, 29–43.

Skerratt, J., Wild-Allen, K., Rizwi, F., Whitehead, J., Coughanowr, C., 2013. Use of a high resolution 3D fully coupled hydrodynamic, sediment and biogeochemical model to understand estuarine nutrient dynamics under various water quality scenarios. Ocean Coast. Manag. 83, 52–66.

Volta, C., Arndt, S., Savenije, H.H., Laruelle, G.G., Regnier, P., 2013 Nov. C-GEM (v 1.0): a new, cost-efficient biogeochemical model for estuaries and its application to a funnel-shaped system. Geosci. Model Dev. Discuss. 6, 5645–5709.

Wang, R., Dearing, J.A., Langdon, P.G., Zhang, E.L., Yang, X.D., Dakos, V., Scheffer, M., 2012. Flickering gives early warning signals of a critical transition to a eutrophic lake state. Nature 492 (7429), 419–422.

Zhang, W.T., Arhonditsis, G.B., 2009. A Bayesian hierarchical framework for calibrating aquatic biogeochemical models. Ecol. Model. 220 (18), 2142–2161.