

**USELESS ARITHMETIC?  
LESSONS LEARNT FROM AQUATIC BIOGEOCHEMICAL MODELING**

**George B. Arhonditsis**

Department of Physical & Environmental Sciences, University of Toronto  
Toronto, Ontario, Canada, M1C 1A4

\* Corresponding author

e-mail: [georgea@utsc.utoronto.ca](mailto:georgea@utsc.utoronto.ca); Tel.: +1 416 208 4858; Fax: +1 416 287 7279.

*“I stress that the problem was not mathematics per se but the place of idolatry we have given it. And it is idolatry. Like any priesthood, it has developed its own language, rituals and mystical signs to maintain its status and to keep a befuddled congregation subservient, convinced that criticism is blasphemy...Most frightening at all, our complacent acceptance of this approach shows that mathematics has become a substitute of science. It has become a defense against an appropriate humility, and a barrier to the acquisition of knowledge and understanding of our ocean environments...When used improperly, mathematics becomes a reason to accept absurdity”*

James O’Malley, former member of the New England Fisheries Management Council and Executive Director of the East Coast Fisheries Federation, as quoted in the *“Useless Arithmetic: Why Environmental Scientists Can’t Predict the Future”* by O.H. Pilkey & L. Pilkey-Jarvis.

### **Useless arithmetic or useful scientific tool?**

Since the final quarter of the twentieth century, mathematical modeling has been extensively used in environmental science as: (i) a technique for elucidating causal mechanisms, complex interrelationships, direct and indirect paths in ecological structures, (ii) a substitute for performing experiments that are technologically or economically unattainable by other means, and (iii) a heuristic tool for developing conceptual models, testable hypotheses, and -ultimately- theories. Models have widespread use in environmental decision-making and management, where they assist with the examination of *“what-if?”* scenarios representing alternative management schemes. In this context, models have offered an indispensable methodological tool for assessing system response to different environmental stress conditions and for communicating potential repercussions/preferred options to managers and politicians who must make decisions but lack scientific expertise. Examples of models

used in environmental science as decision support tools for the management of natural resources include those for air pollution chemistry, relating air pollution to adverse health effects, hydrology, surface and ground water quality, population dynamics (e.g., fisheries models), land use planning, climate change on global, regional and local scales, and system adaptation. In all these disciplines, the appeal of models mainly stems from their ability to synthesize among different types of information reflecting our best understanding of the ecosystem functioning, to identify the key individual relationships and feedback loops from an inconceivably complex array of intertwined environmental processes, and to probe their relative role on system behavior using a range of model application domains (Spear, 1997).

Despite the considerable progress accomplished over the last 3-4 decades, the credibility of models for forming the basis of public policy decisions has been severely criticized in the literature. With over 650 citations, the Oreskes et al. (1994) paper stands out as one of the classical critiques of the veracity of their scientific methodology, advocating the -somewhat provocative- viewpoint that the validation of models which deal with open systems is inherently impossible. If we go beyond the controversy arising from the technical/philosophical meaning of validation (Rykiel, 1996), this statement essentially highlights the important notion that model outputs should be viewed through the prism of the underlying assumptions and that model acceptance in two or more settings is not evidence for general model applicability, but rather the start of a perpetual race for confirmation. The greater the number of cases where the model is tested and confirmed, the higher the likelihood that its structure and conceptualizations are not fundamentally defective. More recently, Pilkey and Pilkey-Jarvis (2007) disputed the trustworthiness of many models for addressing vexing aspects of the environmental stewardship. Using several examples, their book demonstrated how recklessly developed models have provided shaky foundation for environmental management policies, thereby leading to faulty decisions with –oftentimes- devastating effects on the environment, e.g., poisoned mining sites, misleading

predictions of sea level rise rates, uncontrolled nuclear wastes, unrealistic cost projections of artificial beaches, and erroneous estimations of fish population yields. The authors attributed the documented inadequacy for addressing societally important issues to the fact that mathematical modeling has advanced without the healthy dose of criticism required to obtain good science; partly because the “impenetrable” nature of mathematics has prohibited sober views from the user community, which in turn “has allowed modelers to carry their trade far beyond the limits of reality” (Pilkey and Pilkey-Jarvis, 2007).

Regardless of how one responds to the critical views of mathematical modeling as a scientific endeavor, it cannot be denied that the arbitrary selection of higher –and often unattainable- threshold values for environmental variables (quality goals/standards) as a hedge against unknown forecast errors, risky model-based management decisions, and unanticipated system responses are often experienced in current management practice. Modelers are far from being in a position to claim that the current generation of ecological models has the capacity to address increasingly complex issues on the management of natural resources (Clark et al., 2001). Our experience indicates that the forecasting of natural system behavior is extremely difficult, as ecosystem dynamics are driven by foreseeable environmental processes which are often confounded with self-organized, complex adaptive behaviors that are difficult to be predicted (Stow et al., 2003). Nonetheless, despite the compelling reasons for identifying the idiosyncrasies and knowledge gaps of the natural environment, for differentiating between predictable and unpredictable patterns, and for critically evaluating model outputs, modelers frequently overstate the power of models making unfounded statements and misleading projections. Furthermore, the modeling community has not reached a consensus yet about the methodological protocol needed to develop mathematical models or about the appropriate standards, test procedures, and specific metrics required for objectively assessing model performance (Rykiel, 1996). Contrary to other

disciplines, there is still considerable confusion about the theoretical, operational, and data components pertaining to mathematical modeling, although several critical voices have appeared in the literature pinpointing the controversy amongst model developers and the resource managers who use them about how to develop, evaluate and interpret numerical earth science models (Levins, 1966; Holling, 1978; Beck, 1987; Franks, 1995; Rykiel, 1996).

What is the capacity of the current models to simulate the dynamics of environmental systems? How carefully do modelers develop their models? Which model features primarily determine our decision to utilize a specific model? How rigorously do we assess what a model can or cannot predict? The purpose of this paper is to answer some of these questions by reviewing the state of aquatic biogeochemical modeling; a research tool that has been extensively used for understanding and quantitatively describing aquatic ecosystems. Mechanistic aquatic biogeochemical models have formed the scientific basis for environmental management decisions by providing a predictive link between management actions and ecosystem response; they have provided an important tool for elucidating the interactions between climate variability and plankton communities, and thus for addressing questions regarding the pace and impacts of climate change. The sizable number of aquatic ecosystem modeling studies which successfully passed the scrutiny of the peer-review process along with the experience acquired from addressing a breadth of management problems can objectively reveal the systematic biases, methodological inconsistencies, and common misconceptions characterizing the modeling practice in environmental science. My arguments are that (i) models are not always developed in a consistent manner, clearly stated purpose, and predetermined acceptable model performance level, (ii) the potential “customers” select models without properly assessing their technical value, (iii) the development of novel methods for rigorously assessing the uncertainty underlying model predictions should be a top priority of the modeling community, and (iv) the model complexity should be

commensurate to the available knowledge from the system; we need to adopt techniques that allow selecting parsimonious models over unjustifiably complex or oversimplistic modeling constructs.

### **How effectively do we model aquatic ecosystem dynamics?**

It is about 40 years ago when Chen (1970) introduced a general model structure for addressing a broad class of water quality problems. This modeling framework essentially proposed a general set of equations for describing key physical, chemical, and biological processes with site-specific parameters, initial conditions, and forcing functions which then were used to reproduce real-world dynamics, to gain insights into the ecosystem functioning, and to project future system response under significantly different external conditions (e.g., nutrient enrichment, climate change). The philosophy and the basic set of equations proposed in these early models still remain the core of the current generation of mechanistic aquatic biogeochemical models, although advances in scientific understanding and improvements in methods of numerical analysis have brought significant progress with regards to the accuracy and sophistication. Reckhow and Chapra (1999) interpreted the fact that all the recent improvements in water quality modeling have built and evolved upon the foundation provided by early studies from the mid-70s, as evidence of the strength of the original modeling propositions (O'Connor et al., 1975; Thomann et al., 1975; Donigian and Crawford, 1976). On the other hand, Arhonditsis et al. (2006) argued that the absence of novel ideas and creativity may also be a pathological symptom of the field of aquatic ecosystem modeling inviting one to ask what it would take to prime the pump for significant breakthroughs to come along.

In an attempt to evaluate the current state of aquatic biogeochemical modeling, Arhonditsis and Brett (2004) presented a meta-analysis of 153 modeling studies published in the peer-reviewed literature between 1990 and 2002. Despite the heterogeneity of the modeling studies examined with respect to

model complexity, type of ecosystem modeled, spatial and temporal scales, and model development objectives, this study unveiled statistically significant trends of the model performance. The first striking feature of this analysis was the absence of systematic goodness-of-fit assessment of the original models, i.e., plots in which simulated values were visually compared with observed data were only presented for 16.8% of the model endpoints, and even less (1.3%) were the cases in which thorough statistical examination of the model fit was reported. In the cases in which measures-of-fit or comparison plots were presented, Arhonditsis and Brett (2004) independently assessed state-variable performance as expressed by the relative error ( $RE = \Sigma |\text{observed values} - \text{simulated values}| / \Sigma \text{observed values}$ ) and the coefficient of determination ( $r^2$ ) (Fig. 1). It was found that temperature and dissolved oxygen had the lowest  $RE$  (median < 10%) and the highest  $r^2$  values (the respective medians were 0.93 and 0.70). The typical limiting nutrient forms ( $NO_3$ ,  $NH_4$ ,  $PO_4$  and  $Si$ ) in freshwater and oceanic ecosystems along with the phytoplankton biomass had intermediate fit, with median  $r^2$  values varying from 0.40 to 0.60 and the median  $RE$  lying around the 40% level. Zooplankton dynamics were characterized by the highest  $RE$  (70%) and the widest range of  $r^2$  (interquartile range ~0.8) and  $RE$  (interquartile range ~85%) values. Similarly, bacteria were also poorly modeled (median  $r^2$  value <0.06), indicating that the performance of existing mechanistic aquatic biogeochemical models declines as we move from physical-chemical to biological components of planktonic systems. On a positive note, it was found that these results were obtained without the introduction of a major “calibration bias”, i.e., in the process of maximizing the fit for a specific state variable (usually phytoplankton biomass), the modelers do not seem to compromise on the fit for other state variables (such as limiting nutrient concentrations or herbivorous zooplankton biomass).

Arhonditsis and Brett (2004) also assessed the effects of model complexity (expressed as the number of state variables), spatial dimension (from zero- to three-dimensional models), simulation

period (from days to decades), and ecosystem-type on model performance. The study reported a positive correlation between the number of state variables and the *RE* values for the different model outputs ( $r = 0.219, p < 0.001$ ). This (counterintuitive) positive trend was even stronger when considering the *RE* values for phytoplankton ( $r = 0.248, p = 0.003$ ) and zooplankton ( $r = 0.626, p < 0.001$ ) biomass suggesting that more complex models usually result in slightly poorer model performance. Similarly, a (very weak) positive correlation was found between the duration of the simulation period and the state variable *RE* values ( $r = 0.098, p = 0.022$ ) indicating that longer simulations are also increasing model misfit. Marginally significant correlations also exist between the spatial complexity of the models and their (*RE* values) performance trends ( $r = 0.104, p = 0.015$ ). Finally, model error did not vary depending on the type of ecosystem modeled (lakes/reservoirs, embayments/lagoons, coastal ecosystems/estuaries and oceanic systems). These results provide overwhelming evidence that ambitious effort to increase the level of ecological information mathematically represented by the models, to increase spatial complexity and to use longer simulation periods, has not lead to a systematic improvement in model performance.

### **How carefully do we develop our models?**

After four decades of experience, there are many excellent presentations in several modeling textbooks of what “rational model development” is (e.g., Chapra, 1997; Jorgensen and Bendoricchio, 2001). The modeling literature emphatically argues the need for methodological consistency of the way models are being developed, highlighting the importance of several critical steps, such as sensitivity analysis, formal model calibration, and rigorous model validation. Nonetheless, Arhonditsis and Brett (2004) showed a surprising absence of a systematic methodological protocol widely followed from the modelers. In particular, only 27.5% of the models published from 1990 to 2002 reported results of sensitivity analysis, while 27.4% of the studies solely tested the influence of certain model structures

(e.g., alternative formulations) or parameters without providing any quantitative measures of model response to the perturbations induced to the input vector. On the other hand, 45.1% of the modeling studies did not provide any insights into the model behavior by evaluating the sensitivity of the model outputs to parameters, forcing functions, or state-variable submodels. These results are quite disappointing because this step is essential for indicating the accuracy required for the forcing function data as well as for identifying the parts of the model that need to be estimated with greater precision. However, modelers do not seem to appreciate the instrumental role of sensitivity analysis for selecting the optimal structure and complexity in the model development process.

Modelers are commonly confronted with what is called an inverse problem: there is sufficient information on the levels and the variability of the state (or dependent) variables, but little is known about the values of the model parameters. The procedure by which the modeler adjusts the model parameters to find the best agreement between modeled and observed data is called calibration. Model calibration can be carried out by trial and error or by using optimization techniques. The optimization methods are designed to search the parameter space for combinations of parameters which provide the best fit through minimization of cost or objective functions (i.e., functions that measure the discrepancy between observed data and model outputs). Thus, the latter strategy ensures that the calibration parameter set is optimal and that a significant lack of fit is due to the inadequacy of the model structure and not due to poor parameter choice (Chapra and Canale, 1998). Despite the conceptual and practical advantages of model optimization, Arhonditsis and Brett (2004) found that only a small proportion (8.5%) of the aquatic biogeochemical modeling studies base their calibration results on optimization algorithms, whereas the vast majority (91.5%) of the modelers adopted the traditional manual calibration. For some inexplicable reason, aquatic ecosystem modelers seem reluctant to embrace optimization techniques and to include them in their repertoire when developing models.

Modeling textbooks also emphasize that the calibration of a model (or “model training” phase) does not provide any information with regards to its predictive power, but merely examines the ability of a specific model structure to match a single dataset (Chapra, 1997). It is recommended that the calibration should always be followed by the predictive evaluation; a procedure whereby the modeler tests the model against an independent set of data, which ideally should be significantly different from the one used during the calibration phase. This phase is also referred to as model “validation”, although this term may be inappropriate for models that deal with open systems and numerous sources of uncertainty (Oreskes et al., 1994). Arhonditsis and Brett (2004) compiled information for three types of validation: (i) predictive validation defined as an evaluation of model performance against data acquired from the real system after model calibration, (ii) model transferability to different systems (performance of a specific model structure to different regions or ecosystem types), and (iii) structural validation defined as the assessment of the realistic reproduction of the operational characteristics, causal relationships, and relative magnitudes of various components of the system by the model (biological rates, derived quantities). These validation procedures were carried out (or at least explicitly reported) in 47.1% of the aquatic ecosystem modeling studies, while the remaining 52.9% were not predictively or structurally validated.

Generally, aquatic ecosystem modelers do not seem to consistently apply conventional methodological steps during the development of their models. The large majority of the published studies in the field over the last decade did not properly assess model sensitivity to the input vectors; aquatic ecosystem modelers are still reluctant to embrace optimization techniques during model calibration, and assess the ability of their models to support predictions in the extrapolation domain. Thus, the establishment of a systematic methodological protocol for model development, which is widely accepted by the aquatic biogeochemical modeling community, should be a top priority. The

modelers should understand that the methodological consistency is an analogue to the way a chemical analyst strives to attain clean laboratory conditions, excellent standardization curve, and faithful application of the analytical protocol.

### **Which factors determine the impact of a modeling study?**

Citation rates and impact factors are increasingly recognized as convenient tools for assessing the importance and utility of scientific research; ideally, high quality papers should motivate future research and should be used as source of information by subsequent studies in the field (Adams, 2002; Leimu and Koricheva, 2005). Building upon the results presented in the Arhonditsis and Brett (2004) study, Arhonditsis et al. (2006) attempted a second quantitative assessment of the current state of aquatic biogeochemical modeling by focusing on the factors that determine their citation frequency. The main objectives of this analysis were: (i) the evaluation of how has the modeling community received the 153 aquatic biogeochemical models published from 1990 to 2002; and (ii) the identification of the characteristics of a model that are more attractive to the potential users and may influence the frequency of its use and subsequent citation.

The articles citing aquatic biogeochemical modeling studies were classified in 60 different disciplines. Several of these disciplines (e.g., astronomy, computer science, software engineering, plant sciences, genetics and heredity) did not have close association with aquatic ecosystem models, which may be evidence that this field can produce scientific knowledge (e.g., methodological advancements for system analysis, ecological questions addressed) with broader application to different subject areas. Oceanography was found to be the most popular thematic area of the articles that cite mechanistic aquatic biogeochemical models and more than 27% (approximately 1500 counts) of the total citations were related to this research topic (Fig. 2a). The second most popular subject category was marine and

freshwater biology (21.75%) followed by geosciences (10.90%) and ecology (10.33%). On the other hand, disciplines more closely associated with environmental management, e.g., environmental sciences (8.21%), environmental engineering (2.37%) and water resources (1.81%), received a relatively low proportion of the total citations.

The same study also examined the association between citation rates and several characteristics of the published modeling studies: ecosystem-type, methodological consistency, and model performance. Using as a criterion the type of the ecosystem modeled, the published modeling studies were classified in six categories, i.e., “*Coastal area-Estuary*”, “*Mesocosm*”, “*Bay-Lagoon-Harbour*”, “*Lake-Reservoir*”, “*Ocean-Sea*”, and “*River*”. Ocean modeling studies have received significantly higher citations ( $F=7.87$ ,  $df=5$ ,  $p<0.001$ ) among the various ecosystem-types (Fig. 2b). Regarding the model complexity (expressed as the number of state variables) as a model feature that attracts citations, it was found that the citation rates of the individual articles were not significantly correlated with the corresponding model complexity ( $r=0.111$ ,  $p=0.183$ ), although there was an increasing citation trend for models with over 40 state variables (Fig. 3a). Model performance was also not considered as a criterion for citing modeling papers; e.g., citation rates and model performance for the “key” state variable phytoplankton were not significantly correlated ( $r=0.163$ ,  $p=0.07$ ; Fig. 3b). It was also examined whether the methodological consistency of the published modeling studies is a factor that determines their citation rates. The citation counts did not differ significantly among studies that presented (thorough/partial) sensitivity analysis or not ( $F=1.16$ ,  $df=2$ ,  $p=0.316$ ). The citation patterns of the modeling studies were not affected by whether or not the modelers reported assessment of the goodness-of-fit ( $F=0.05$ ,  $df=1$ ,  $p=0.943$ ; Fig. 3c), and by whether or not the original modeling studies presented structural or predictive validation ( $F=2.03$ ,  $df=1$ ,  $p=0.156$ ).

Overall, these results show that the citations of the aquatic biogeochemical modeling studies are determined neither by the model complexity nor by the reported performance and methodological consistency. The type of the ecosystem being modeled is the most influential factor shaping the citation patterns of the modeling papers. Ocean modeling studies receive considerable attention and overwhelmingly dominate the total citation counts. Arhonditsis et al. (2006) attributed the high number of citations of oceanic modeling to two basic reasons. First, many of the oceanic numerical models have been developed to offer insights into the oceanic response to climate change and also to illuminate the interplay between plankton dynamics and atmospheric CO<sub>2</sub> levels via several feedback mechanisms, e.g., “biological pump”, calcification (Hays et al., 2005). Thus, oceanic modeling produces knowledge that is appealing to a broader audience and stimulates research spanning a wide range of tightly intertwined disciplines. Second, oceanic modeling appears to be a more methodologically-coherent and vibrant area of research, oceanic modelers are keener to embrace technical advances for controlling prediction error or for addressing problems of underdetermination, e.g., to implement assimilation schemes, to test new ecological theories, and to include specific plankton functional types and multiple element cycles (Doney, 1999; Denman, 2003; Anderson, 2005).

Model application for addressing environmental management issues on a local scale faces challenges as a scientific tool. The number of studies from lakes, reservoirs, coastal embayments, estuaries, and harbors combined was approximately equal to the number of oceanic applications, whereas the majority of these -local character- modeling studies have received fairly low citations. Arhonditsis et al. (2006) highlighted as a main reason for the low citation rates the failure to engage novelty and creativity with solutions to management problems which inevitably results in unattractive modeling products that cannot export knowledge to other disciplines. Reckhow and Chapra (1999) emphasized that the conceptual weaknesses, methodological omissions, lack of quantification of the

residual variability and parameter uncertainty in predictions are more critical when addressing practical management problems. Evidently, modelers are reluctant to borrow experiences and new ideas from other disciplines or similar character modeling studies when addressing environmental management issues; for example, data assimilation techniques, formulations that consider new ecological theories (e.g., stoichiometric nutrient recycling theory), and novel calibration methods are relatively rare (Errico, 1997; Kennedy and Hagan, 2001; Arhonditsis and Brett, 2005a,b). Clearly, some of the current modeling practices have to change, as the need for robust modeling tools to assist with the restoration of impaired water bodies is more pressing now than ever before; e.g., the costly implementation of total maximum daily loads for pollutants during the next 10-15 years has raised the bar for innovative model developments that can accommodate rigorous error analysis (Borsuk et al., 2002).

### **Distinguishing between what we can and what we cannot learn from a model**

Uncertainty analysis of mathematical models has received considerable attention in aquatic ecosystem research, and there have been several attempts to rigorously address issues pertaining to model structure and input error (Beck, 1987; Reichert and Omlin, 1997). Model structure error is mainly associated with (i) the selection of the appropriate state variables for reproducing ecosystem behavior, (ii) the selection of the suitable equations among a variety of mathematical formulations for describing the ecological processes, e.g., Monod and Variable-Internal Stores (VIS) models for simulating the phytoplankton uptake of nutrients from the water column and their utilization/storage in phytoplankton cells, and (iii) the fact that our models are based on relationships which are derived individually in controlled laboratory environments but may not collectively yield an accurate picture of the real world dynamics (Reichert and Omlin, 1997). Model input error mainly stems from the uncertainty underlying the values of model parameters, initial conditions and forcing functions as well as the realization that all

models are drastic simplifications of reality that approximate the actual processes, i.e., essentially, all parameters are effective (e.g., spatially and temporally averaged) values unlikely to be represented by fixed constants.

Model practitioners also encounter the problem that several distinct choices of model inputs lead to the same model outputs, i.e., many sets of parameters fit the data about equally well. This non-uniqueness of the model solutions is known in the modeling literature as equifinality (Beven, 1993). The main reason for the equifinality (poor identifiability) problem is that the causal mechanisms/hypotheses used for understanding how the system works internally is of substantially higher order of what can be externally observed (Beck, 1987). As a result, our ability to set quantitative (or even qualitative) constraints as to what is realistic/behavioral simulation of an ecological structure along with the learning capacity of a model is significantly reduced. Having a unique determination of model structure (and associated parameter values) that realistically reflects the natural system dynamics is particularly important when our goal is to make predictions for future hypothesized states. For example, when an eutrophication model does not operate with realistic ecological structure (e.g., relative/absolute magnitudes of biological rates and transport processes), even if the fit between model outputs and observational data is satisfactory, its credibility to provide predictions about how the system will respond under significantly different external nutrient loading conditions is very limited. Finally, another problem that is seldom explicitly acknowledged is that the conventional calibration schemes, i.e., mere adjustment of model parameters until the discrepancy between model outputs and observed data is minimized, may provide the best fit to the dataset available at the moment, but it is specific to the given dataset at hand. As new data become available, the model should be recalibrated and in the common calibration practice there is no way of considering previous results. In this sense, we do not update

previous knowledge about model input parameters, but rather we make the models dataset-specific (Arhonditsis et al., 2007a).

In recognition of the uncertainty and equifinality problems, it is suggested that the model calibration practice should change from seeking a single “optimal” value for each model parameter, to seeking a distribution of parameter sets that all meet a pre-defined fitting criterion (Stow et al., 2007; Arhonditsis et al., 2007a). These acceptable parameter sets may then provide the basis for estimating model prediction error associated with the model parameters. Model uncertainty analysis can be viewed as an attempt to formulate the joint probability distribution of model inputs, reflecting the amount of knowledge available for model parameters, initial conditions, forcing functions, and model structure, and then to update our knowledge on the form of this distribution after considering the observed data from the system. In this regard, Bayes’ Theorem provides a convenient means to combine existing information (prior) with current observations (likelihood) for projecting future ecosystem response (posterior). Hence, the Bayesian techniques are more informative than the conventional model calibration practices, and can be used to refine our knowledge of model input parameters as well as to obtain predictions along with uncertainty bounds for output variables (Arhonditsis et al., 2007a). The work of Hornberger and Spear (1981) was the first “Bayesian-like” effort to identify plausible parameter sets for large multi-parameter environmental models. The proposed method, called regionalized (or generalized) sensitivity analysis (RSA), was a Monte Carlo sampling approach to assess model parameter sensitivity. RSA was a conceptually simple way to use limited information to bound model parameter distributions. Given a particular model and an environmental system being modeled, the modeler first defines the plausible range of certain key model endpoints as the “behavior”. For example, based on existing information from the system the total phosphorus concentration lies within the range of 10-30  $\mu\text{g/L}$ . The modeler then samples from (usually) uniform distributions of each of the model

parameters; all the parameter sets that result in total phosphorus predictions within the “behavioral” range are termed “behavior generating” and become part of the model parameter distribution.

The assessment of the uncertainty characterizing the multidimensional parameter spaces of mathematical models involves two critical steps: i) selection of the sampling scheme for generating input vectors which then are evaluated with regards to the model performance, and ii) selection of the likelihood measure to quantify model misfit. The selection of the scheme for generating input vectors addresses the sampling efficiency of the approach, e.g., Random sampling, Latin hypercube, Markov chain Monte Carlo (MCMC). Many Bayesian or non-Bayesian uncertainty analysis applications (e.g., Generalized Likelihood Uncertainty Estimation, Bayesian Monte Carlo) have been combined with sampling algorithms which draw samples uniformly and independently from the prior parameter space. These sampling strategies often insufficiently cover regions of high model likelihood; especially, when the joint prior parameter distribution is very wide or the parameters are highly correlated (Qian et al., 2003). To address this problem, several studies advocate the use of MCMC sampling schemes specifically designed to sample directly from the posterior distribution and to converge to the higher model likelihood regions (Gelman et al., 1995; Arhonditsis et al., 2007a; Stow et al., 2007). On the other hand, the selection of the likelihood measures entails conceptual dilemmas involving the selection of generalized (e.g., Root Mean Square Error, U-uncertainty, Reliability Index, Modeling Efficiency) or purely probabilistic (e.g., Normal, Lognormal or Poisson error) likelihood functions that can significantly change the results (Beven, 2001). In typical uncertainty analysis illustrations, the likelihood measure is broadly specified as any measure of goodness-of-fit that can be used to compare observed data with model predictions, e.g., sum of squared errors, fuzzy measures or even qualitative measures for model evaluation (Beven, 2001). From a statistical inference standpoint, the lack of formal representation of the model error has been criticized for providing biased parameter estimates when not

taking into account the correct model error structure (Thiemann et al., 2001). It has also been argued that unless the likelihood measure corresponds to a well-defined probability distribution that directly connects the data with model input parameters and output state variables, the uncertainty analysis results do not have a clear Bayesian interpretation (Engeland and Gottschalk, 2002; Hong et al., 2005).

Several recent water quality modeling studies have attempted to demonstrate how the Bayesian inference techniques combined with MCMC sampling schemes can improve model forecasts and management actions over space and time (Malve et al., 2007; Arhonditsis et al., 2007a; 2008a,b). Arhonditsis et al. (2007a) introduced a Bayesian calibration scheme using intermediate complexity mathematical models (4-8 state variables) and statistical formulations that explicitly accommodate measurement error, parameter uncertainty, and model structure imperfection. Namely, the Bayesian configuration of the model was based on statistical formulations that assumed (i) a “perfect” model structure along with additive (or multiplicative) measurement error; and (ii) a simulator that imperfectly represents the dynamics of the natural system. The former formulation postulates that the model misfit is solely caused by the error associated with the data, whereas the latter one also considers errors in the model structure, e.g., missing key ecological processes, misspecified forcing functions, erroneous formulations. It should also be noted that, aside from the analytical/sampling error, the term measurement error also reflects the notion that the observational data are just a “snapshot” of the real system, an instantaneous record of few components from numerous complex and interactive processes that depending on the sampling network used, the ecosystem modeled and the questions addressed, can form an objective criterion for evaluating model performance (Fagerstrom, 1987). The characterization of the uncertainty underlying the model parameters prior to model calibration (prior parameter distributions) was based on field observations from the lake, laboratory studies, literature information, and expert judgment using the protocol presented by the Steinberg et al. (1997) study. The Bayesian

calibration framework was then used to quantify the information the data contain about model inputs, to offer insights into the covariance structure among parameter estimates, and to obtain predictions along with credible intervals for model outputs (Arhonditsis et al., 2007a; 2008a, b).

Some of the technical advances from the Bayesian calibration methodology are as follows:

i) *Identification Problem*: By incorporating prior information on the model parameters, the Bayesian inference techniques offer an effective strategy to alleviate the identification problem. Prior knowledge of the magnitudes of ecological processes can be converted into probability distributions that reflect the relative plausibility of different values of the respective model parameters, which then can be included into the “prior-likelihood-posterior” update cycles. As a result, the use of additional information (along with the calibration dataset) reduces the disparity between what ideally we want to learn (internal description of the system) and what can realistically be observed, which is the primary reason for the poor model identifiability (Beck, 1987). In this regard, Arhonditsis et al. (2007a) showed that the assumption of a perfect model structure usually results in narrow-shaped parameter distributions, whereas the statistical formulation representing the imperfect simulator tends to provide flatter posteriors (Fig. 4). Although the latter finding indicates that the inclusion of the model discrepancy error terms reduces the information gained regarding the values of the calibration vector, the low posterior variances of the former approach do not necessarily depict the amount of knowledge with regards to the parameter values when considering prior literature information and available data from the system modeled. This result may stem from an overconditioning of the parameter estimates owing to an overestimation of the information content of the observations, which then can limit the potential applicability of the model in the extrapolation domain (Arhonditsis et al., 2008a)

ii) *Realistic uncertainty estimates of the ecological forecasts*: For the purpose of prediction, the Bayesian approach generates a posterior predictive distribution that represents the current estimate of the

value of the response variable, taking into account both the uncertainty about the parameters and the uncertainty that remains when the parameters are known (Gelman et al., 1995). Therefore, the estimates of uncertainty of Bayesian model predictions convey significant information in regards to model credibility. In particular, the statistical formulations that explicitly consider the discrepancy between mathematical model and environmental system have been repeatedly shown to improve the model performance, i.e., the median predictions along with the 95% credible intervals delineate zones that accurately describe the observed data (Fig. 5). Thus, Arhonditsis et al. (2008a) pointed out that the development of statistical formulations explicitly recognizing the lack of perfect simulators of natural system dynamics is a promising prospect for the Bayesian calibration framework, but future research should seek for possible modifications to accommodate complex mathematical models.

In a follow up study, Zhang and Arhonditsis (2008a) used a complex aquatic biogeochemical model that simulates multiple elemental cycles, multiple functional phytoplankton (diatoms, green algae and cyanobacteria) and zooplankton (copepods and cladocerans) groups to illustrate how the Bayesian calibration framework can be used for assessing the exceedance frequency and confidence of compliance of different water quality criteria (Fig. 6). A recent analysis also demonstrates how the incorporation of updated models into Bayesian hierarchical frameworks enables the transfer of information across systems, thereby allowing the effective modeling of lakes with limited information, i.e., problems of insufficient local data can be overcome by “borrowing strength” from well-studied sites on the basis of distributions that connect systems in space (Zhang and Arhonditsis, 2008b). This outcome is highly relevant to conservation practices of regions with high number of water resources for which complete data could never be practically gathered (Fig. 7). Finally, aside from the probabilistic assessment of the water quality conditions, another benefit of the Bayesian parameter estimation is the alignment with the policy practice of adaptive management, i.e., an iterative implementation strategy

that is recommended to address the often-substantial uncertainty associated with water quality model forecasts, and to avoid the implementation of inefficient and flawed management plans (Walters, 1986). Adaptive implementation or “learning while doing” supports initial model forecasts of management schemes with post-implementation monitoring, i.e., the initial model prediction serves as the Bayesian prior, the post-implementation monitoring data serve as the sample information (the likelihood), and the resulting posterior probability (the integration of monitoring and modeling) provides the basis for revised management actions (Qian and Reckhow, 2007).

### **Complex mathematical models: An emerging imperative or “shiny mathematical castles on grey biological sand”?**

As knowledge regarding the complex components of environmental systems continues to grow, there is a demand for increasing the articulation levels of our mathematical models. Generally, the premise for constructing complex models is to mirror the complexity of natural systems and consider ecological processes that can become important in future states driven by significantly different conditions (Reichert and Omlin, 1997). Modelers essentially believe that if they can include all the significant processes in the mathematical equations, then the model will closely mimic the real system and thus will have increased predictive ability under a wide range of environmental conditions. However, there are several important reasons to be very careful when we decide to opt for more complex models. First, the basic elements of large process-oriented models are simple equations adopted as useful first approximations of isolated behaviors in controlled laboratory experiments. Although this practice was convenient and necessary to make complex models manageable, it would seem surprising that simple disparate equations should collectively yield accurate information about ecosystem behavior. Indeed, Arhonditsis et al. (2006) showed that the typical assumptions made to characterize mechanisms

in mathematical models negate the premise that “*if the modelers use correct process descriptions then the models can effectively reproduce natural system dynamics*”. This is simply not a reasonable expectation! Beyond that, we seem to be locked into a space/time scale that has become the de facto modeling standard, yet is beyond our ability to correctly capture in the mathematics and is also incompatible with available data for parameter estimation (Arhonditsis et al., 2006). Furthermore, the complex models with their many degrees of freedom (tunable parameters) can in principle provide good fit to any data set, but it should not be neglected that the increasing complexity also reduces our ability to properly constrain the model parameters from observations, i.e., the number of parameters that must be specified from the data is approximately proportional to the square of the number of model compartments (Denman, 2003). Thus, to overcome the overparameterization problem, the ultimate challenge for any model training exercise is to find a parameter vector that performs equally well against both calibration and validation datasets, thereby effectively connecting current system dynamics with conditions representing chosen scenarios.

In the context of aquatic biogeochemical modeling, there is an increasing pressure to explicitly treat multiple biogeochemical cycles, to increase the functional diversity of biotic communities, and to refine the mathematical description of the higher trophic levels (Anderson, 2005; Arhonditsis and Brett, 2005a; Fennel, 2008). In particular, there are views in the literature suggesting the inclusion of multiple nutrients along with the finer representation of plankton communities, as necessary model augmentations for disentangling critical aspects of aquatic ecosystem dynamics, e.g., species populations are more sensitive to external perturbations (nutrient enrichment, episodic meteorological events), and key biogeochemical processes are tightly linked to specific plankton functional groups (Cottingham and Carpenter, 1998; Flynn, 2005; Arhonditsis et al., 2007b). Nonetheless, the derivation of distinct functional groups from fairly heterogeneous planktonic assemblages poses challenging

problems. Because of the still poorly understood ecology, we do not have robust group-specific parameterizations that can support predictions in a wide array of spatiotemporal domains (Anderson, 2005). For example, preliminary efforts to incorporate plankton functional types into global biogeochemical models were based on speculative parameterization and –not surprisingly- resulted in unreliable predictions (Anderson, 2005). Likewise, Zhao et al. (2008a) showed that the reproduction of seasonal succession plankton patterns in freshwater ecosystems is quite fragile and only occurs in a fairly narrow window of the model parameter space. The latter study also pondered if it is “reasonable to expect single-valued data set-specific parameter estimates of artificially defined biotic entities to be extrapolated over wider geographical regions?” Furthermore, recent attempts to integrate biogeochemistry with fish production underscore the uncertainty arising from the mismatch between the operating time scales of planktonic processes and fish life cycles as well as the need to consolidate the mechanistic description and parameterization of several critical processes, such as the reproduction and mortality of the adult stages (Fennel, 2008). Despite the repeated efforts to increase model complexity, we still haven’t gone beyond the phase of identifying the unforeseeable ramifications and the challenges that we need to confront so as to improve the predictive power of our models (Anderson, 2006).

Another topic that has received considerable attention is the mathematical representation of the biochemical heterogeneity at the primary producer–herbivore interface to illuminate the patterns of nutrient and energy flow transferred through the food web (Andersen, 1997; Arhonditsis and Brett, 2005a; Mulder and Bowden, 2007, Zhao et al., 2008b). One of the most debatable issues in the aquatic ecology is the relative role of the deficiencies in essential elements (e.g., carbon, nitrogen, phosphorus) and biochemical compounds (e.g., essential fatty acids) on zooplankton growth and reproduction. There are studies underscoring the critical role of the discrepancy between the prey and predator elemental somatic ratios on food web structure, e.g., mineral phosphorus-limitation hypothesis (Sterner and

Schulz, 1998; Elser and Urabe, 1999). Others suggest that the algal taxonomic differences in food quality due to differences in their fatty acid (FA), protein, amino acid content, and digestion resistance determine the trophic transfer efficiency in aquatic pelagic food webs, e.g., FA limitation hypothesis (Muller-Navarra et al., 2004). In this regard, considerable insights into the potential implications of the ecological stoichiometry have been gained by a series of homeostatic consumer models that examine the effects of P-deficient food on zooplankton growth rate as well as on consumer-driven P recycling (e.g., Loladze et al., 2000; Andersen et al., 2004). On the other hand, Perhar and Arhonditsis (2008) highlighted the significant gap in the literature of predictive frameworks for the FA limitation, i.e., modeling studies that explicitly consider the constraints on zooplankton growth stemming from the biochemical heterogeneity of the lake seston. Recently, Zhao et al. (2008b) pinpointed two major outstanding challenges of the multi-elemental lower trophic level models: (i) assessment of the role of the non-limiting element recycling (e.g., partitioning between dissolved and particulate phase of the excess carbon and nitrogen) on different ecological processes, such as sedimentation, nitrification/denitrification, and sediment diagenesis rates (Elser and Foster, 1998; Arhonditsis and Brett, 2005b); (ii) the relaxation of the assumption of strict elemental homeostasis and the impact of zooplankton adaptive stoichiometry on ecosystem functioning (Mulder and Bowden, 2007; Ferrão-Filho et al., 2007). Yet, to elucidate these unknown aspects of the food-web dynamics, we cannot merely rely on modeling exercises but rather we need comprehensive data sets that can uniquely constrain some of the pertinent pathways (Flynn, 2006).

Reviewing the recent literature, one can find several interesting recommendations with regards to the future aquatic ecosystem modeling. For example, the series of “Horizons” articles hosted in the Journal of Plankton Research has offered many fresh ideas and pointers for progress (Anderson, 2005; Flynn, 2006; Le Quéré, 2006). Their proposition to open the dialogue between biologists and

mathematicians is certainly one of the ways forward. I also find very intriguing the advocated focus on the relative role of the interconnections among the different biotic subunits on system dynamics. Flynn (2006) emphatically argues that “it is becoming increasingly clear that there are a whole host of interactions between members of plankton that the vast majority of models do not even hint at.” The question arising is are we ready to mathematically depict or even to frame data collection efforts in this direction? Until we can give a positive answer to this question, I believe that the gradual incorporation of complexity, where possible and relevant, is the most prudent strategy. But any such model development should be tightly coupled with rigorous assessment of the underlying uncertainty and the Bayesian inference can be an invaluable ally in this frontier. Of equal importance is the establishment of a systematic methodological protocol for aquatic biogeochemical model development along with performance criteria widely accepted by the modeling community. Even if the journals cannot enforce the submission of all the material required to reconstruct the mathematical models (Flynn, 2005), they can still demand the modeling studies to meet certain methodological and performance criteria. Finally, I couldn't agree more with Anderson's (2006) standpoint that prediction is not everything. We should not be afraid from complex models even if their structure reduces the predictive ability. Complex models offer excellent heuristic tools that allow insights into the direct, indirect, and synergistic effects of the numerous ecological mechanisms forming the foundation of system behaviour. They are an absolutely worthwhile scientific activity!

### **ACKNOWLEDGMENTS**

Funding for this paper was provided by the National Sciences and Engineering Research Council of Canada (NSERC, Discovery Grants), and the Connaught Committee (University of Toronto, Matching Grants 2006–2007). I wish to thank the members of the Ecological Modeling lab (University of

Toronto) Tony Zhang, Gurbir Perhar, Jingyang Zhao, Maryam Ramin, Vincent Cheng, Tony Law, Saman Reshadi, and Dr Ekram Azim for their insightful comments on an earlier version of the manuscript.

## REFERENCES

- Adams, D. 2002. The counting house. *Nature* 415: 726-729.
- Andersen, T. 1997. Pelagic Nutrient Cycles: Herbivores as Sources and Sinks. Springer-Verlag, New York.
- Andersen, T., Elser, J. J., and Hessen, D.O. 2004. Stoichiometry and population dynamics. *Ecol. Lett.* 7: 884-900.
- Anderson, T.R. 2005. Plankton functional type modelling: running before we can walk? *J. Plankton Res.* 27: 1073-1081.
- Ardenson, T.R. 2006. Confronting complexity: reply to Le Quere and Flynn. *J. Plankton Res* 28: 877-878.
- Arhonditsis, G.B., and Brett, M.T. 2004. Evaluation of the current state of mechanistic aquatic biogeochemical modeling. *Mar. Ecol. Prog. Ser.* 271: 13-26.
- Arhonditsis, G.B., and Brett, M.T. 2005a. Eutrophication model for Lake Washington (USA) Part I - model description and sensitivity analysis. *Ecol. Model.* 187: 140-178.
- Arhonditsis, G.B., and Brett, M.T. 2005b. Eutrophication model for Lake Washington (USA) Part II - model calibration and system dynamics analysis. *Ecol. Model.* 187: 179-200.
- Arhonditsis, G.B., Adams-VanHarn, B.A., Nielsen, L., Stow, C.A., and Reckhow, K.H. 2006. Evaluation of the current state of mechanistic aquatic biogeochemical modeling: citation analysis and future perspectives. *Environ. Sci. Technol.* 40: 6547-6554.

- Arhonditsis, G.B., Qian, S.S., Stow, C.A., Lamon, E.C., and Reckhow, K.H. 2007a. Eutrophication risk assessment using Bayesian calibration of process-based models: Application to a mesotrophic lake. *Ecol. Model.* 208: 215-229.
- Arhonditsis, G.B., Stow, C.A., Paerl, H.W., Valdes, L.M., Steinberg, L.J., and Reckhow, K.H. 2007b. Delineation of the role of nutrient dynamics and hydrologic forcing on phytoplankton patterns along a freshwater-marine continuum. *Ecol. Model.* 208: 230-246.
- Arhonditsis, G.B., Perhar, G., Zhang, W., Massos, E., Shi, M., and Das, A. 2008a. Addressing equifinality and uncertainty in eutrophication models. *Water Resour. Res.* 44: W01420.
- Arhonditsis, G.B., Papantou, D., Zhang, W., Perhar, G., Massos, E., and Shi, M. 2008b. Bayesian calibration of mechanistic aquatic biogeochemical models and benefits for environmental management. *J. Mar. Syst.* doi:10.1016/j.jmarsys.2007.07.004.
- Beck, M.B. 1987. Water-quality modeling—a review of the analysis of uncertainty. *Water Resour Res* 23: 1393–1442.
- Beven, K. 1993. Prophecy, reality and uncertainty in distributed hydrological modeling. *Adv. Water Resour.* 16: 41-51.
- Beven, K.J. 2001. Rainfall-Runoff Modeling: The Primer. John Wiley & Sons Ltd, West Sussex.
- Borsuk, M.E., Stow, C.A., Reckhow, K.H. 2002. Predicting the frequency of water quality standard violations: A probabilistic approach for TMDL development. *Environ. Sci. Technol.* 36: 2109-2115.
- Chapra, S. C. 1997. Surface Water-Quality Modeling. McGraw-Hill, New York.
- Chapra, S.C., and Canale, R.P. 1998. Numerical Methods for Engineers. McGraw-Hill, New York.
- Chen, C.W. 1970. Concepts and utilities of ecological models. *J. San. Engr. Div. ASCE* 96: 1085-1086.

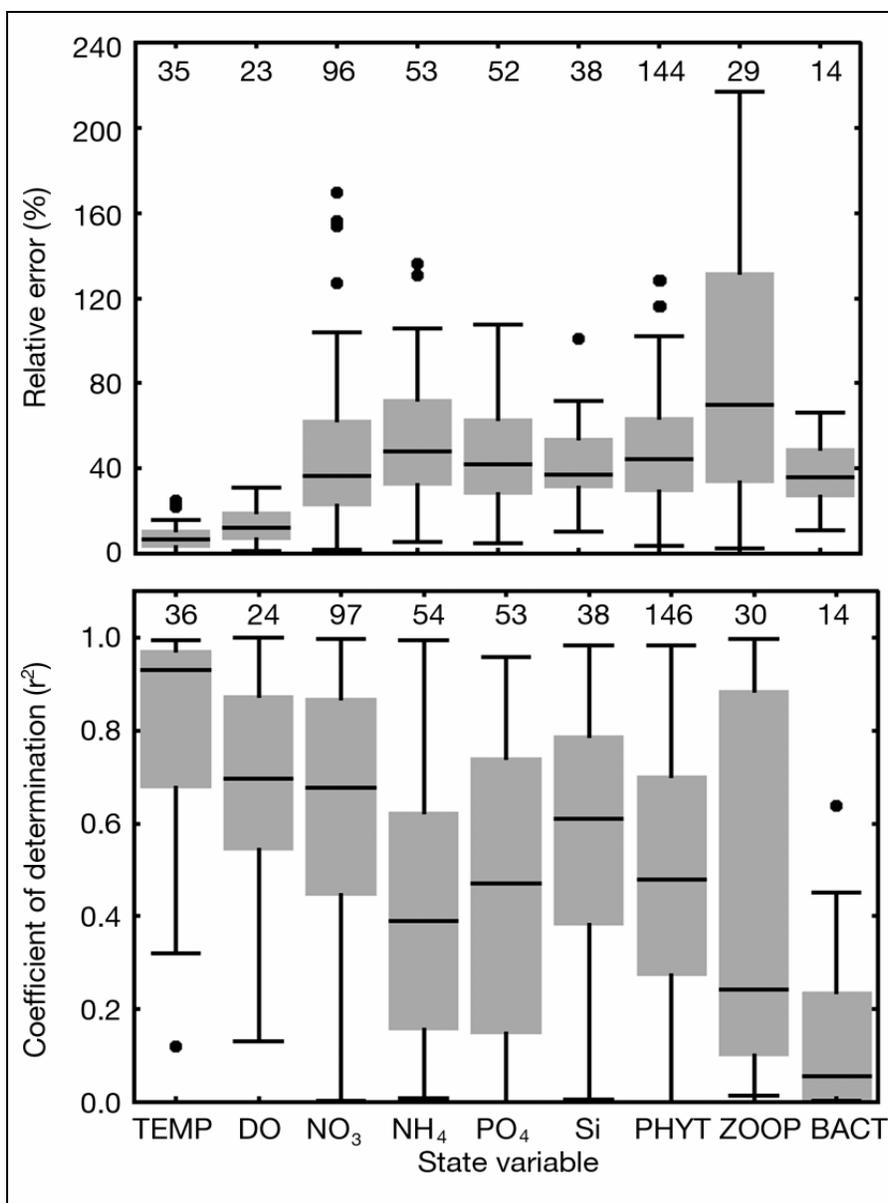
- Clark, J.S., Carpenter, S.R., Barber, M., Collins, S., Dobson, A., Foley, J.A., Lodge, D.M., Pascual, M., Pielke, R., Pizer, W., Pringle, C., Reid, W.V., Rose, K.A., Sala, O., Schlesinger, W.H., Wall, D.H., Wear, D. 2001. Ecological forecasts: An emerging imperative. *Science* 293: 657-660.
- Cottingham, K.L., and Carpenter, S.R., 1998. Population, community, and ecosystem variates as ecological indicators: phytoplankton responses to whole-lake enrichment. *Ecol. Appl.* 8: 508–530.
- Denman, K.L. 2003. Modelling planktonic ecosystems: parameterizing complexity. *Prog. Oceanogr.* 57: 429-452.
- Doney, S.C. 1999. Major challenges confronting marine biogeochemical modeling. *Global Biogeochem. Cy.* 13: 705-714.
- Donigian, A.S. Jr., Crawford, N.H. 1976. Modeling Pesticides and Nutrients on Agricultural Lands EPA 600/2-7-76-043., Environmental Research Laboratory, Athens, GA.
- Elser, J. J., and Foster, D. K. 1998. N:P stoichiometry of sedimentation in lakes of the Canadian shield: relationships with seston and zooplankton elemental composition. *Ecoscience* 5: 56–63.
- Elser, J. J., and Urabe, J. 1999. The stoichiometry of consumer-driven nutrient recycling: Theory, observations, and consequences. *Ecology* 80: 735-751.
- Engeland, K., and Gottschalk, L. 2002. Bayesian estimation of parameters in a regional hydrological model. *Hydrol. Earth Syst. Sc.* 6: 883-898.
- Errico, R.M. 1997. What is an adjoint model? *B. Am. Meteorol. Soc.* 78: 2577-2591.
- Fagerstrom, T. 1987. On theory, data and mathematics in ecology. *Oikos* 50: 258–261.
- Fennel, W. 2008. Towards bridging biogeochemical and fish-production models. *J Mar Syst.* 71: 171-194.

- Ferrão-Filho, A. D. S., Tessier, A. J., and DeMott, W. R. 2007. Sensitivity of herbivorous zooplankton to phosphorus-deficient diets: Testing stoichiometric theory and the growth rate hypothesis. *Limnol. Oceanogr.* 52: 407-415.
- Flynn, K.J. 2005. Castles built on sand: dysfunctionality in plankton models and the inadequacy of dialogue between biologists and modellers. *J. Plankton Res.* 27: 1205–1210.
- Flynn, K.J. 2006. Reply to Horizons Article ‘Plankton functional type modelling: running before we can walk’ Anderson (2005): II. Putting trophic functionality into plankton functional types. *J. Plankton Res.* 28: 873–875.
- Franks, P.J.S. 1995. Coupled physical-biological models in oceanography. *Rev Geophys* 33: 1177–1187.
- Gelman, A., Carlin, J.B., Stern, H.S., and Rubin, D.B. 1995. Bayesian Data Analysis. Chapman and Hall, New York.
- Hays, G.C., Richardson, A.J., and Robinson, C. 2005. Climate change and marine plankton. *Trends Ecol. Evol.* 20: 337-344.
- Holling, C.S. 1978. Adaptive Environmental Assessment and Management. John Wiley & Sons, New York.
- Hong, B.G., Strawderman, R.L., Swaney, D.P., and Weinstein, D.A. 2005. Bayesian estimation of input parameters of a nitrogen cycle model applied to a forested reference watershed, Hubbard Brook Watershed Six. *Water Resour. Res.* 41, W03007.
- Hornberger, G.M., and Spear, R.C. 1981. An approach to the preliminary analysis of environmental systems. *J Environ Manage* 12: 7-18.
- Jorgensen, S.E., and Bendoricchio, G. 2001. Fundamentals of Ecological Modelling. Elsevier, Amsterdam.

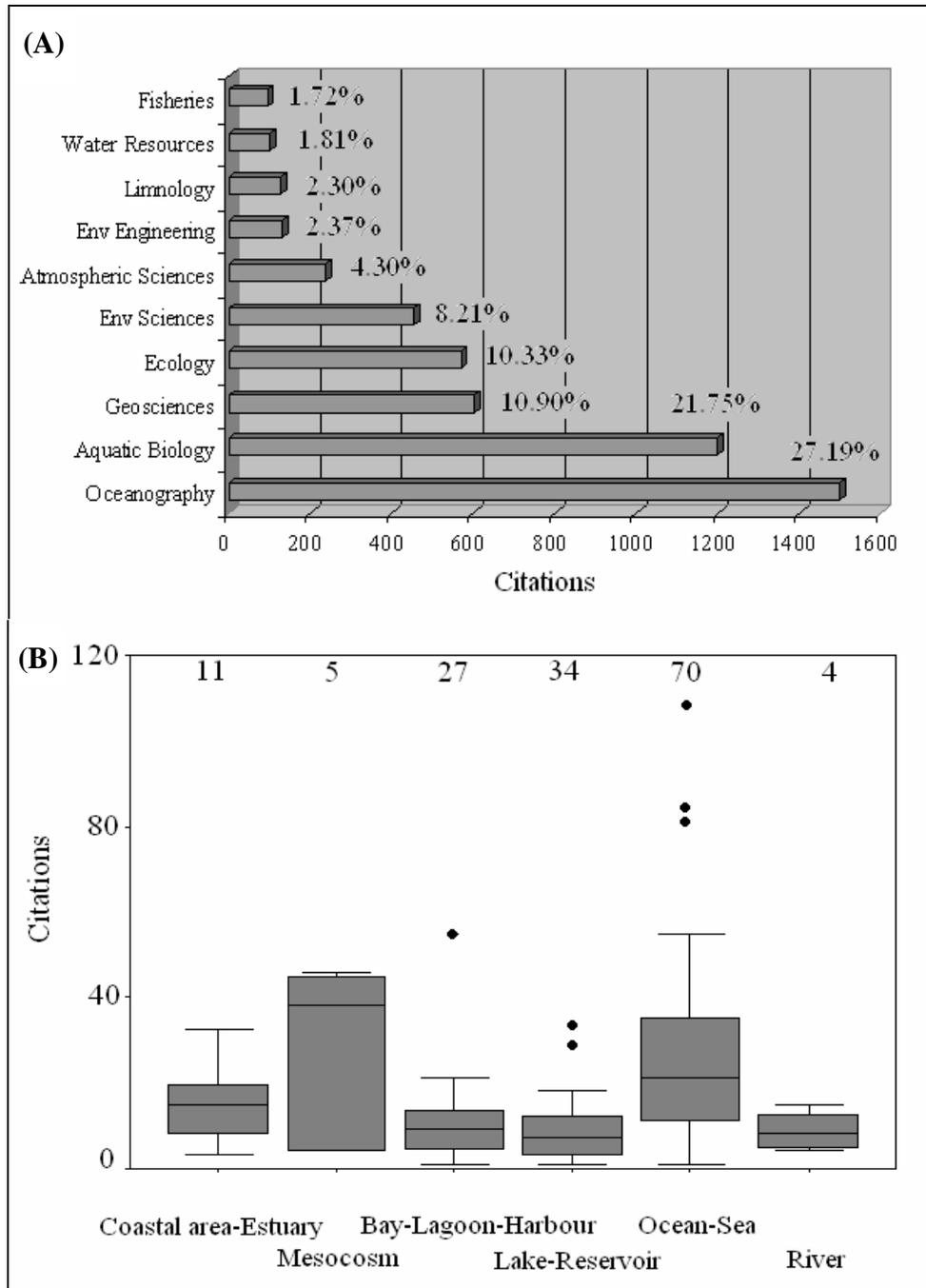
- Kennedy, C.M., and O'Hagan, A. 2001. Bayesian calibration of computer models. *J. R. Statist. Soc. B*, 63: 425-464.
- Le Quéré, C. 2006. Reply to Horizons Article 'Plankton functional type modelling: running before we can walk' Anderson (2005): I. Abrupt changes in marine ecosystems? *J Plankton Res* 28: 871-872.
- Leimu, R. and Koricheva, J. 2005. What determines the citation frequency of ecological papers? *Trends Ecol. Evol.* 20: 28-32.
- Levins, R. 1996. The strategy of model building in population biology. *Am. Sci.* 54: 421-431.
- Loladze, I., Kuang, Y., and Elser, J. J. 2000. Stoichiometry in producer-grazer systems: Linking energy flow with element cycling. *Bulletin of Mathematical Biology* 62: 1137-1162.
- Malve, O., Laine, M., Haario, H., Kirkkala, T., and Sarvala, J. 2007. Bayesian modelling of algal mass occurrences - using adaptive MCMC methods with a lake water quality model. *Environ. Modell. Softw.* 22: 966-977.
- Müller-Navarra, D. C., Brett, M. T., Park, S., Chandra, S., Ballantyne, A. P., Zorita, E., and Goldman, C. R. 2004. Unsaturated fatty acid content in seston and tropho-dynamic coupling in lakes. *Nature* 427: 69-72.
- Mulder, K., and Bowden, W.B. 2007. Organismal stoichiometry and the adaptive advantage of variable nutrient use and production efficiency in *Daphnia*. *Ecol. Model.* 202: 427-440.
- O'Connor, D.J., Di Toro, D.M., Thomann, R.V. 1975. Phytoplankton models and eutrophication problems. In: Russell, C.S. (Ed.), *Ecological Modeling in a Resource Management Framework*. Resources for the Future, Washington, DC.
- Oreskes, N, Shrader-Frechette, K, and Belitz, K. 1994. Verification, validation, and confirmation of numerical models in the earth-sciences. *Science* 263: 641-646.

- Perhar, G, and Arhonditsis, G.B. 2008. The effects of seston food quality on planktonic food web patterns. Submitted manuscript.
- Pilkey, O.H., and Pilkey-Jarvis, L. 2007. Useless Arithmetic: Why Environmental Scientists Can't Predict the Future. Columbia University Press, New York.
- Qian, S.S., Stow, C.A., and Borsuk, M.E. 2003. On Monte Carlo methods for Bayesian inference. *Ecol. Model.* 159: 269-277.
- Qian, S.S., and Reckhow, K.H. 2007. Combining model results and monitoring data for water quality assessment. *Environ. Sci. Technol.* 41: 5008-5013.
- Reckhow, K.H., and Chapra, S.C. 1999. Modeling excessive nutrient loading in the environment. *Environ Pollut* 100: 197–207.
- Reichert, P., and Omlin, M. 1997. On the usefulness of overparameterized ecological models. *Ecol. Model.* 95, 289-299.
- Rykiel, E.J. 1996. Testing ecological models: the meaning of validation. *Ecol. Model.* 90: 229–244.
- Spear, R.C. 1997. Large simulation models: calibration, uniqueness and goodness of fit. *Environ Modell Softw* 12: 219-228.
- Steinberg, L.J., Reckhow, K.H., and Wolpert, R.L. 1997. Characterization of parameters in mechanistic models: A case study of a PCB fate and transport model. *Ecol. Model.* 97: 35-46.
- Sterner, R. W., and Schulz, K. L. 1998. Zooplankton nutrition: Recent progress and a reality check. *Aquat. Ecol.* 32: 261-279.
- Stow, C.A., Roessler, C., Borsuk, M.E., Bowen, J.D., and Reckhow, K.H. 2003. Comparison of estuarine water quality models for total maximum daily load development in Neuse River Estuary. *J Water Res. Pl.-ASCE* 129: 307-314.

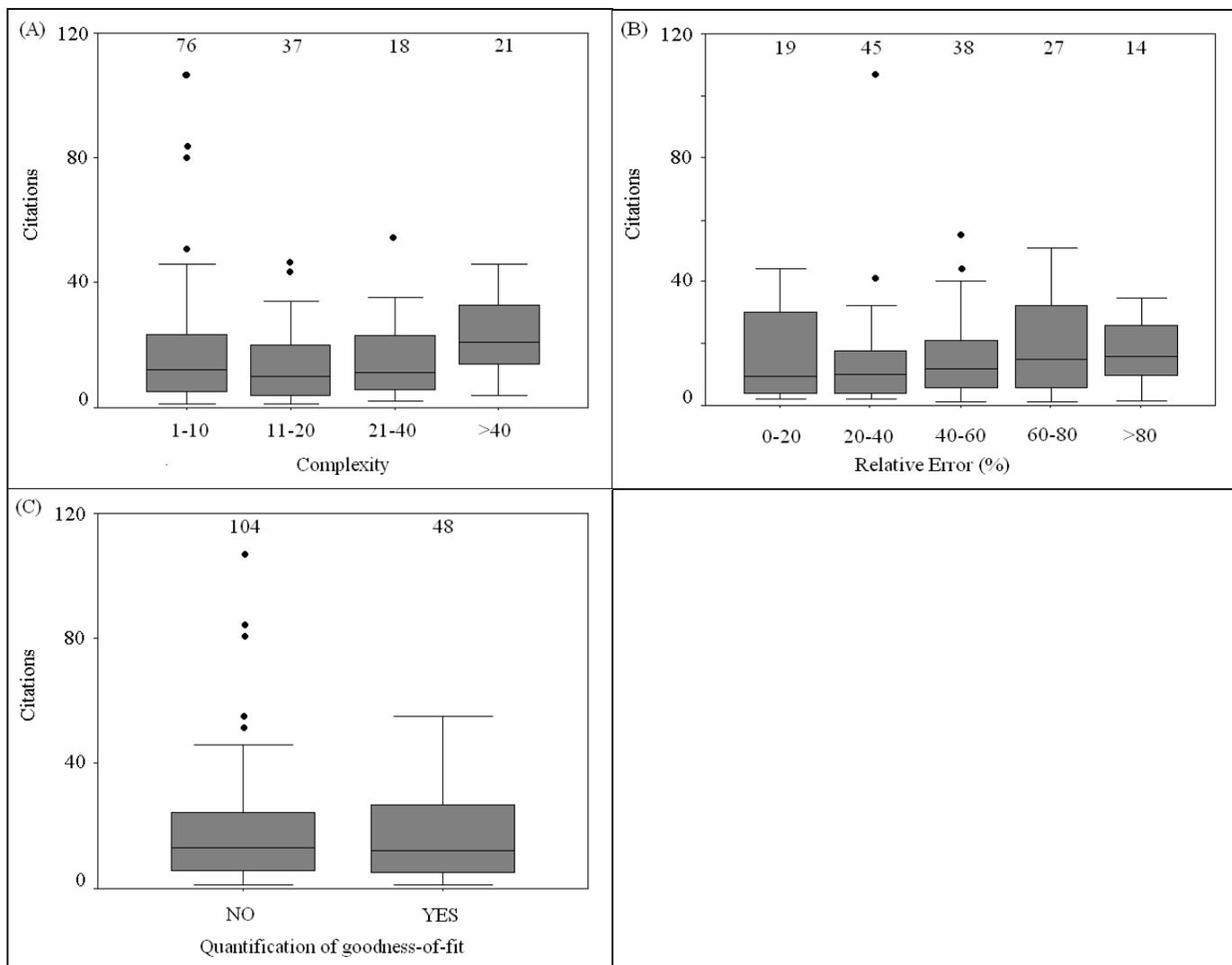
- Stow, C.A., Reckhow, K.H., Qian, S.S., Lamon, E.C., Arhonditsis, G.B., Borsuk, M.E., and Seo, D. 2007. Approaches to evaluate water quality model parameter uncertainty for adaptive tmdl implementation. *J. Am. Water Resour. As.* 43: 1499-1507.
- Thiemann, M., Trosset, M., Gupta, H., and Sorooshian, S. 2001. Bayesian recursive parameter estimation for hydrologic models. *Water Resour. Res.* 37: 2521– 2535.
- Thomann, R.V., Di Toro, D.M., Winfield, R.P., O'Connor, D.J. 1975. Mathematical modeling of phytoplankton in Lake Ontario. 1. Model development and verification. U.S. Environmental Protection Agency, EPA/600/3-75-005, Corvallis, OR 97330.
- Walters, C.J. 1986. *Adaptive Management of Renewable Resources*. McMillan, New York.
- Zhang, W., and Arhonditsis, G.B. 2008a. Predicting the frequency of water quality standard violations using Bayesian calibration of eutrophication models. *J Great Lakes Res.* In press.
- Zhang, W., and Arhonditsis, G.B. 2008b. A Bayesian hierarchical framework for calibrating aquatic biogeochemical models. Submitted manuscript.
- Zhao, J., Ramin, M., Cheng, V., and Arhonditsis, G.B. 2008a. Plankton community patterns across a trophic gradient: The role of zooplankton functional groups. *Ecol. Model.* 213: 417-436.
- Zhao, J., Ramin, M., Cheng, V., and Arhonditsis, G.B. 2008b. Competition patterns among phytoplankton functional groups: How useful are the complex mathematical models? *Acta Oecol.* 33: 324-344.



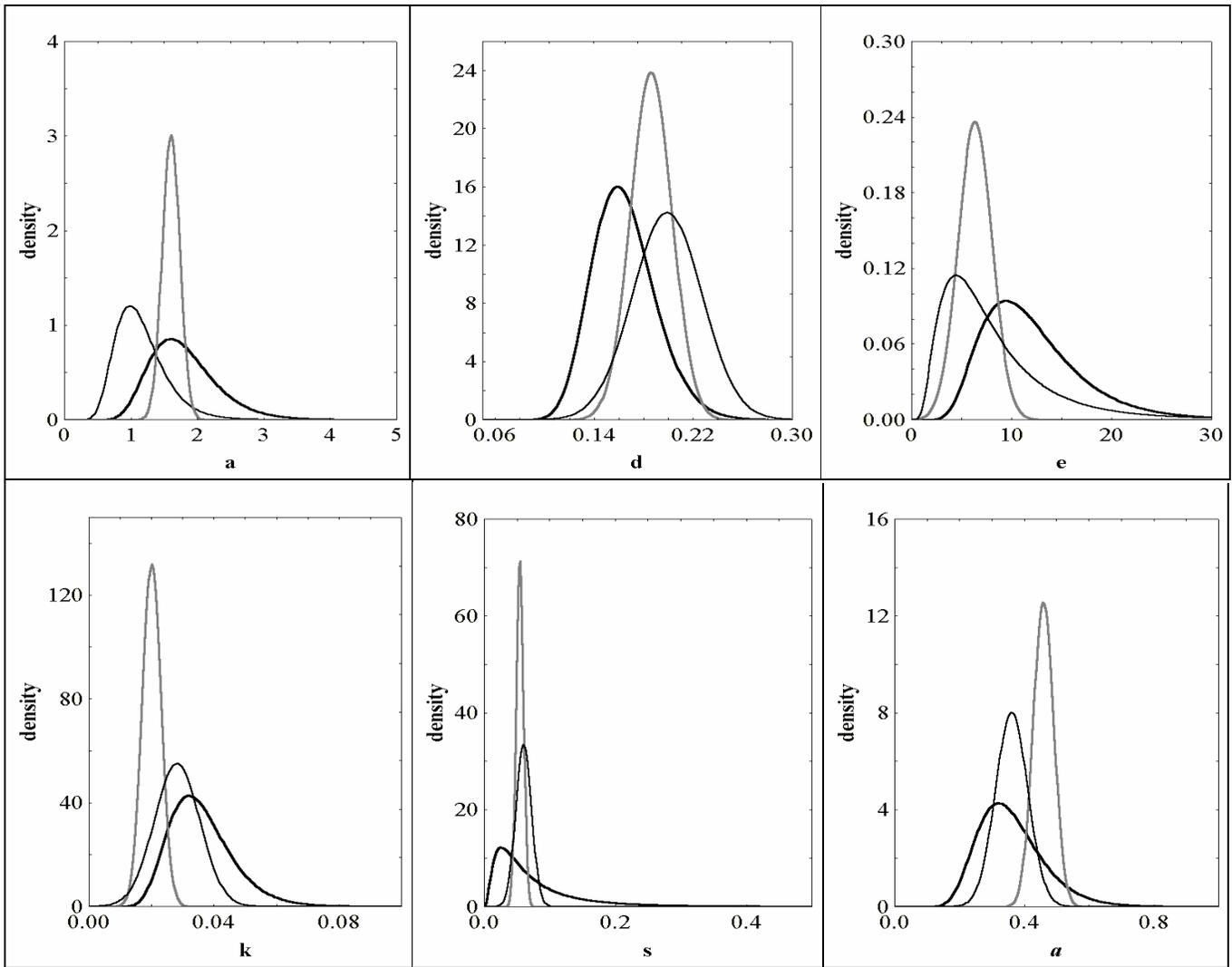
**Figure 1:** Performance of aquatic biogeochemical models for the study period 1990 to 2002. Relative error (%) and coefficient of determination ( $r^2$ ) values for temperature (TEMP), dissolved oxygen (DO), nitrate ( $\text{NO}_3$ ), ammonium ( $\text{NH}_4$ ), phosphate ( $\text{PO}_4$ ), silicate (Si), phytoplankton (PHYT), zooplankton (ZOOP) and bacteria (BACT). Numbers of studies for each state variable are indicated at the top of the corresponding box-plots. [Adapted from Arhonditsis and Brett, 2004. Evaluation of the current state of mechanistic aquatic biogeochemical modeling. Marine Ecology Progress Series, 271, 13–26.]



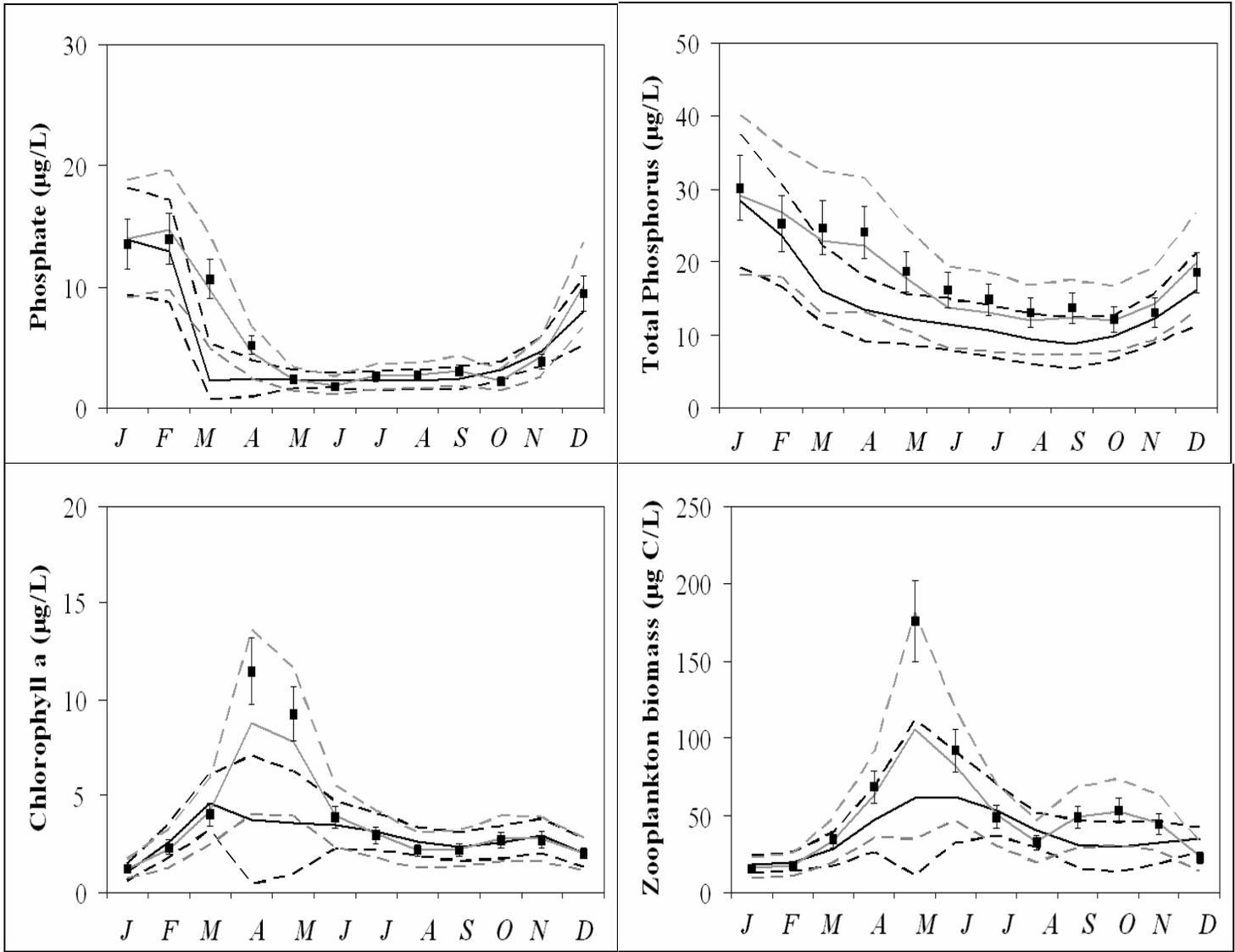
**Figure 2:** (A) Frequency histogram of the scientific classification (subject category) of the papers that cite mechanistic aquatic biogeochemical modeling papers. (B) Citation frequency for different types of modeled ecosystem. [Adapted from Arhonditsis et al., 2006. Evaluation of the current state of mechanistic aquatic biogeochemical modeling: Citation analysis and future perspectives. *Environmental Science & Technology*, 40, 6547–6554.]



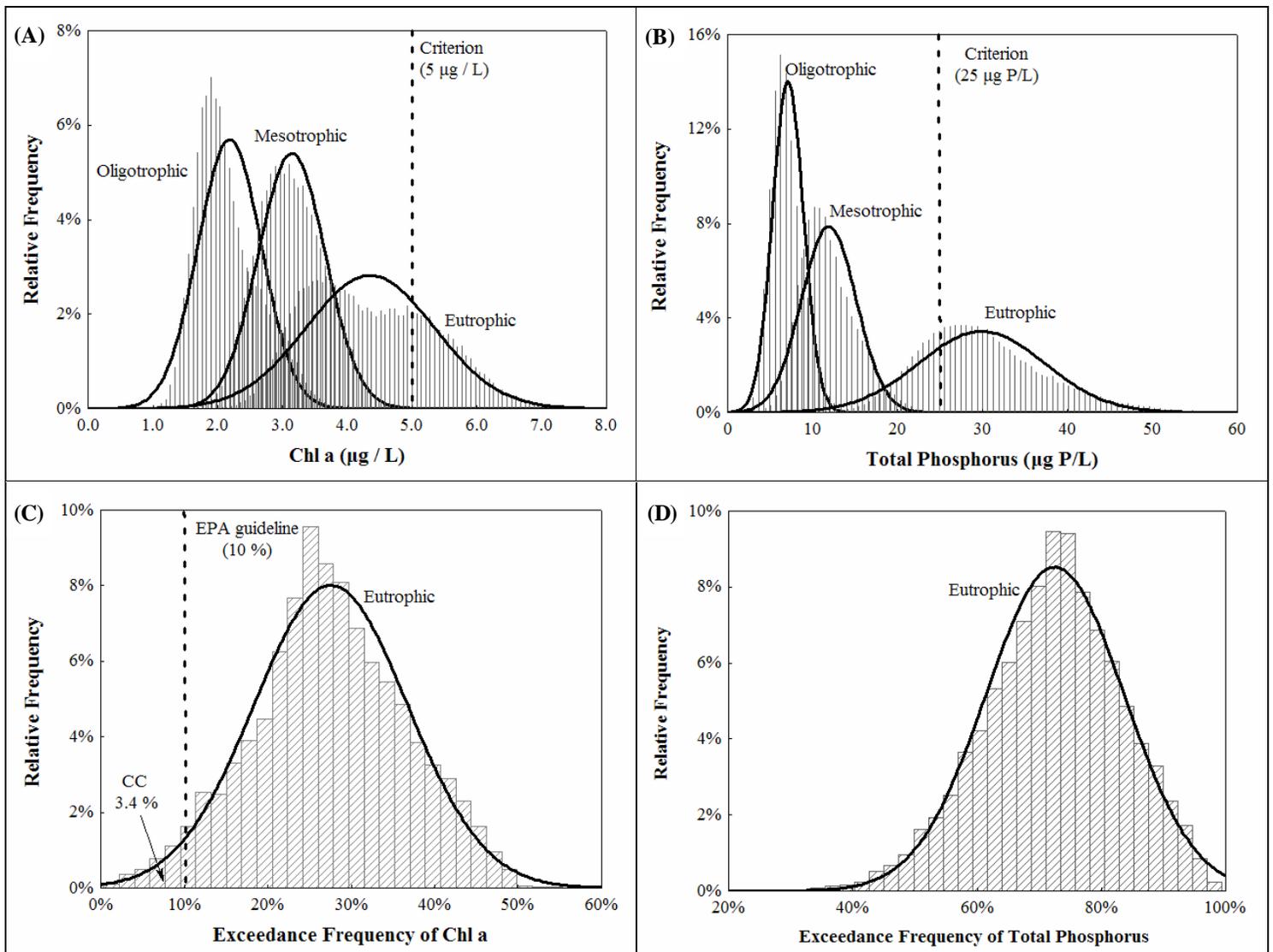
**Figure 3:** Citation frequency for different levels of (A) model complexity (number of state variables), (B) model performance for phytoplankton, and (C) reported assessment of the goodness-of-fit in the original modeling study. [Adapted from Arhonditsis et al., 2006. Evaluation of the current state of mechanistic aquatic biogeochemical modeling: Citation analysis and future perspectives. *Environmental Science & Technology*, 40, 6547–6554.]



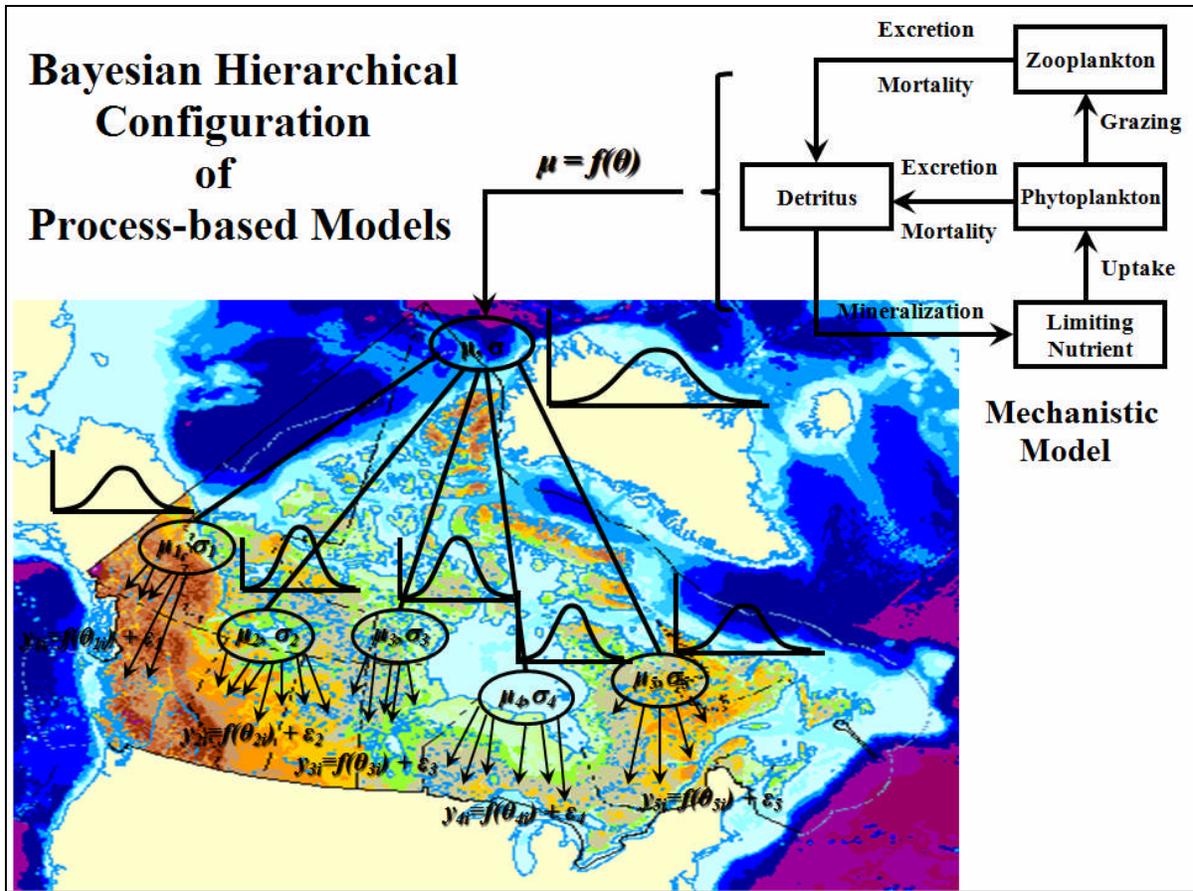
**Figure 4:** Prior (thick black lines) and posterior parameter distributions of a limiting nutrient-phytoplankton-zooplankton-detritus model: gray and thin black lines correspond to statistical formulations that assume perfect and imperfect model structure, respectively. [Adapted from Arhonditsis et al., 2007a. Eutrophication risk assessment using Bayesian calibration of process-based models. *Ecological Modelling*, 208, 215–229.]



**Figure 5:** Comparison between the observed data and posterior predictive monthly distributions for phosphate, total phosphorus, chlorophyll *a* and total zooplankton abundance. Black and gray lines correspond to statistical formulations that assume perfect and imperfect model structure, respectively. Dashed lines correspond to the 2.5 and 97.5% credible intervals. [Adapted from Arhonditsis et al., 2007a. Eutrophication risk assessment using Bayesian calibration of process-based models. *Ecological Modelling*, 208, 215–229.]



**Figure 6:** (A-B) Predictive distributions for water quality variables of management interest (chlorophyll *a*, total phosphorus). The dashed lines correspond to the numerical criteria used to determine the frequency of violations under different trophic conditions. (C-D) The exceedance frequency of the different water quality standards (chlorophyll *a*: 5  $\mu\text{g/L}$ , total phosphorus: 25  $\mu\text{g/L}$ ). In these distributions, the area below the 10% cutoff point is termed the confidence of compliance (CC), and represents the probability that the true exceedance frequency is below the 10% EPA guideline.



**Figure 7:** Structure of the Bayesian hierarchical framework: The problem of parameter estimation using cross-system data is viewed as a hierarchy. At the bottom of the hierarchy are the process-based models for individual waterbodies  $i$ ,  $\mu_{vi} = f(\theta_{vi})$ . In the next level, the spatial heterogeneity is accommodated by introducing  $v$  ( $=5$ ) “regional” distributions; i.e., depending on the geographical location, the model parameters  $\theta_{vi}$  are drawn from one of these local populations. Similarly, in the upper stage, the local population parameters  $\mu_v$  and  $\sigma_v$  are specified probabilistically in terms of global population parameters or hyper-parameters; for example, a global mean  $\mu$  and variance  $\sigma$  that correspond to the wider Canadian area. The observed data  $y_{vi}$  are used to estimate the model parameters  $\theta_{vi}$ , the “regional” population parameters  $\mu_v$ ,  $\sigma_v$  and the hyperparameters  $\mu$ ,  $\sigma$ .