



75 years since Monod: It is time to increase the complexity of our predictive ecosystem models (opinion)

Ferdi L. Hellweger

Civil and Environmental Engineering Department, Northeastern University, Boston, MA, United States

ARTICLE INFO

Article history:

Received 17 August 2016
Received in revised form 5 December 2016
Accepted 5 December 2016
Available online 5 January 2017

Keywords:

Ecosystem model
Biogeochemical model
Eutrophication model
Phytoplankton model
Water quality model
Complexity
MIMO

ABSTRACT

The complexity of our mechanistic ecosystem (biogeochemical, eutrophication, phytoplankton, water quality) models has not changed substantially since they were developed in the 1970s. Consequently, there is now a large disconnect with contemporary problems (e.g., toxin production), knowledge of biological and ecological processes (e.g., intracellular mechanisms, sediment bed overwintering) and environmental observational technologies (e.g., metatranscriptomics). This limits the utility of models for making predictions and supporting management. There are several reasons against increasing complexity, including (a) number of required assumptions, (b) risk of overfitting, (c) higher uncertainty, (d) missing knowledge, (e) lack of observations for calibration and validation and (f) difficulty of developing, running, analyzing and communicating the model. Here I review those arguments and conclude that, for mechanistic, predictive ecosystem modeling, they either do not apply, are not a significant problem in practice or can readily be solved by providing more resources to modelers. Further, a review of these issues leads to the conclusion that more complexity generally increases the predictive skill of a model, because more information is used to constrain it. This can be formulated as a new rule: more in, more out (MIMO). MIMO suggests that more complex models make better predictions, but this should not be adopted as a universal modeling strategy, because in practice, the difficulty associated with developing, understanding and communicating complex models has to be considered. However, those are problems readily solved by more resources and I argue that more funding needs to be made available to develop complex ecosystem models.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

In 1942, the French microbiologist Jacques Monod published an equation relating the growth rate of a microbial population to the concentration of the limiting nutrient. Ever since, this (or the equivalent Michaelis-Menten equation) has been the method of choice for simulating resource-limited growth in ecological models (Franks, 2009; Hellweger, 2015). There have been some advances, such as using the internal (vs. external) nutrient level (Droop, 1968) or resolving some intracellular speciation (Flynn et al., 1997), but those have generally not been adopted into operational models. This stagnation of our models is a stark contrast to the progress made in the understanding of biological and environmental systems, where today's knowledge must be orders of magnitudes larger than what it was 75 years ago. There is also a disconnect with modern problems, like trace metal transformation (e.g., mercury methylation), proliferation of antibiotic resistance or toxin production. This limits the utility of models for research and management.

Lake Erie exemplifies the present situation. In the 1960s and 1970s the lake was highly eutrophic, which prompted large-scale and expensive P loading reductions (i.e., wastewater treatment plants, P detergent ban) resulting in an initial significant improvement in water quality. However, since the mid-1990s the lake is again eutrophic and now plagued annually by toxic *Microcystis* blooms, which in 2014 led to the shutdown of the drinking water supply of the City of Toledo (Bullerjahn et al., 2016; Steffen et al., 2014a). Over the years, Lake Erie has served as a testbed for water quality modeling and many excellent models have been and continue to be developed for this system (Di Toro et al., 1987; Verhamme et al., 2016). Current models can predict the size of the annual bloom as a function of P loading, but they do not make predictions of many important features, such as speciation (e.g., diatoms vs. *Microcystis*) or toxin production, do not include many important biological and ecological processes (e.g., overwintering in the sediment bed, Kutovaya et al. (2012)) that may affect how the system responds to change, and do not produce output that can be compared to modern observations (e.g., transcript levels, Steffen et al. (2015)). Consequently their value to understanding and managing this system is limited.

E-mail address: ferdi@coe.neu.edu

Why are we not building more complex models? The level of complexity is a constant debate in ecological modeling, whether it is in the context of using models to develop general theory or to make prediction (Anderson, 2005, 2010; Arhonditsis and Brett, 2004; Evans et al., 2013a; Evans et al., 2013b; Flynn, 2005; Franks, 2009; Fulton et al., 2003; Hood et al., 2007; John and Flynn, 2000; Shimoda and Arhonditsis, 2016; Topping et al., 2015). Here I will focus on the use of models for prediction (i.e., tactical modeling, Evans et al. (2013b)), in the sense of forecasting how the concentration of cyanobacteria toxins may change when nutrient input is reduced, or how the carbon flux from the surface to the deep ocean may change under a warmer climate. In this context, the main arguments for keeping models simple can be summarized as follows:

- (a) Number of required assumptions. Occam's Razor states that among competing hypotheses (here models), the one with the fewest assumptions should be selected. Therefore, unless the complex models can clearly out-perform the simple ones, they should not be used.
- (b) Risk of overfitting. Complex models have more degrees of freedom and this can result in overfitting, where the model is calibrated to noise in the observations, rather than the underlying functional relationship. This decreases the predictive skill of the model.
- (c) Higher uncertainty. Complex models are underdetermined or overparameterized and often can produce the same results using different parameterizations or model formulations. The associated uncertainty propagates to the predictions and increases their uncertainty.
- (d) Missing knowledge. The understanding of processes, especially biological, is limited and insufficient to support detailed mechanistic simulation.
- (e) Lack of observations for calibration and validation. Complex models produce more output, and the environmental observations needed to calibrate and validate them are not available.
- (f) Difficulty of developing, running, analyzing and communicating the model. Complex models are harder to develop, compute, understand and communicate.

Here I review and challenge these arguments in the context of mechanistic, predictive ecosystem modeling, by providing counterarguments and making use of hypothetical and real-world examples. Some of these arguments have been criticized previously by others using the same or similar counterarguments (see references below). My aim here is not to present only novel arguments (although to my knowledge some are new), but to counter the case against increasing complexity in a comprehensive manner.

The purpose of this paper is to critique the state of our predictive ecosystem modeling. I want to state upfront that this criticism is not directed at the modelers *per se*. For the most part they are doing excellent work given the limited resources available. Rather, as I argue below, the problem is one of resources made available to the modelers and that is the responsibility of the entire environmental science and engineering community

I find that the arguments against more complex models either do not apply, are not a practical problem or can readily be solved by providing more resources to modelers. Further, for most issues I conclude that the value of a model increases with complexity, because more information is used in the development or calibration. This leads to an alternate approach to evaluating the complexity of models, where the usefulness of a model, in the context of prediction, is a function of the amount of information that is fed into it. This idea is summarized as "more in, more out (MIMO)"

2. Number of required assumptions

2.1. Argument

Occam's Razor, also known as the law of parsimony, states that among competing hypotheses, the one with the fewest assumptions should be selected. In the context of modeling, this is often interpreted as support for fewer processes, components and parameters (i.e., simpler models) (Elliott and Thackeray, 2004; Flynn, 2005; Hodges and Rudnick, 2004; Shimoda and Arhonditsis, 2016).

2.2. Counterargument: any model makes assumptions about every part of the system

The basic problem with applying Occam's Razor to mechanistic ecosystem models in this manner is that the complexity of a model is not related to the number of assumptions made by the model. The assumptions depend on the system being modeled and any model makes assumptions about every part of the system. Simple models tend to assume that various components are constant and/or mechanisms are not important, whereas complex models explicitly include them. Rather, by adding more complexity we make more of the assumptions explicitly (vs. implicitly when mechanisms are ignored), which allows us to use more information to make the model better.

2.3. Example phytoplankton diversity

Phytoplankton models now routinely include multiple species or functional types (e.g., N-fixers), but the diversity in our models is typically still far below that of their real-world counterparts. Does increasing the number of species increase the number of assumptions made by a model? Consider two models of cyanobacteria in a lake. The simple model lumps all cyanobacteria species into one state variable and the complex models breaks them up into N-fixers and non-N-fixers. For this example, let's consider ammonia and nitrate as limiting nutrient and assume the growth rate is simulated using a Monod function that requires the specification of a maximum growth rate (μ_{max}) and a half-saturation constant for ammonia and nitrate (K_{m,NH_4} , K_{m,NO_3}). The simple model requires specification of only three parameters, whereas the complex model needs six. At the surface, this may suggest the complex model makes more assumptions. However, the simple one-component model also makes assumptions about μ_{max} , K_{m,NH_4} and K_{m,NO_3} for N-fixers and non-N-fixers. It assumes they are all the same, but this is also an assumption. So the number of species or parameters does not translate into the number of assumptions made. Then the question becomes: Is assuming that all groups have the same nutrient limitation parameters better than assuming they are different? There is no reason to believe this, which suggests the complex model is not any worse than the simple model. However, the complex model can be informed by more knowledge. There is evidence that non-N-fixing cyanobacteria are better competitors for ammonia and worse competitors for nitrate (Blomqvist et al., 1994), so when multiple species are included, this information can be fed into the model. Unless we know absolutely nothing about the properties of the added species, including them in the model will open a gate to use more information and increase the value of the model. How far beyond broad functional groups can this be taken? Modern observational technology is increasingly able to characterize microdiversity. A recent study performed single-cell genomics on over 1,000 cells of the marine cyanobacteria *Prochlorococcus* collected near Bermuda, and found substantial differences. On average 1–5% of their genomes was different and much of these differences

Download English Version:

<https://daneshyari.com/en/article/5742347>

Download Persian Version:

<https://daneshyari.com/article/5742347>

[Daneshyari.com](https://daneshyari.com)