Grimm, V. (2023):
Ecology needs to overcome siloed modelling
Trends Ecol. Evol. **38** (12), 1122 - 1124

https://doi.org/10.1016/j.tree.2023.09.011

Ecology needs to overcome siloed modelling

Keywords: ecological modelling; culture; lingua franca; fit-for-purpose; validation

Volker Grimm[1,2,3]

[1]Helmholtz Centre for Environmental Research - UFZ, Department of Ecological Modelling, Permoserstr. 15, 04318 Leipzig, Germany. volker.grimm@ufz.de, ORCID 0000-0002-3221-9512.

[2]University of Potsdam, Department of Plant Ecology and Nature Conservation, Am Mühlenberg 3, 14476, Potsdam-Golm, Germany

[3]German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Puschstr. 4, 04103 Leipzig, Germany

Abstract

Bahlburg *et al.* [1] re-implemented eight growth models of Antarctic krill and show that their predictions are all over the place. The authors discuss the reasons for this and how more coherence in modelling could be achieved through systematic model comparison and integration. For this, we need a common language.

Hidden assumptions

Models are simplified representations, designed for a certain purpose. The purpose or question we are asking determines how much we simplify. For example, some forest models do not even include trees, while others start from the stomata of single leaves. This is common and good practice. But does this mean that if the purpose is the same, models developed independently of each other are more or less the same in structure and hence predictions? This is not the case.

For example 15 models, all addressing the causes of cod (*Gadus morhua)* population dynamics in the Baltic Sea, are structurally very different ([2], Fig. 1). As a result, the explanations derived from these models are different or even contradictory. How can we still find robust answers to the original research question and thereby advance ecology?

Obviously, there are more determinants of model structure and simplifying assumptions than the models‘ purpose. Modellers talk to different experts and use different data for calibration, but also prefer different model types, approaches to analyse and use the models, or simply have different world views. Some prefer more abstract models, while others appreciate more realism, e.g. by including physiology and behaviour.

All these factors strongly influence how modellers ultimately design their models. However, we do not have a culture of reflecting on and communicating these hidden assumptions and how they constrain what we can learn from our models in the first place. If we want to make models more useful for solving the current ecological and environmental challenges, we cannot continue to develop models in isolation. This is clearly demonstrated and visualised by a recent comparison of krill models by Bahlburg *et al*. [1].

Comparing krill models

Antarctic krill (*Euphausia superba*) is a key species of the Southern Ocean and is of paramount importance for carbon sequestration and ecosystem functioning [3]. Predicting how krill populations will be affected by climate change and commercial fishing is essential for policy and management. As a result, many models were developed to predict the growth of individual krill in relation to their environment. The growth predicted by these models is then used as a proxy for abundance or habitat quality.

There are a large number of such growth models, some based on data from laboratory experiments and others on field observations. Bahlburg *et al*. [1] were surprised by the large differences between these models. They systematically compared eight krill growth models, three of them being empirical and five mechanistic. They then re-implemented these eight models and ran them for the same temperature and resource scenario for the Southern Ocean.

Although all the models tested were designed to predict krill growth, there was little agreement in their projections. While each of the models certainly has its merits, and all have been carefully developed and tested, the results from Bahlburg *et al.* leave us with the question of how we should use them to support real-world management decisions. To answer this question, the models need to be compared.

A common format for comparing models

A first and necessary step to move forward is to document the different models in a common format. Bahlburg *et al*. list, for each model, the number and type of the krill’s life stages considered (between 1 and 3) and the number and type of environmental drivers considered (up to 4 out of 7). The eight models represented seven different combinations of life stages and environmental drivers. The models were then further compared in terms of the methodological and spatial origin(s) of their calibration data, and the shape of the functional relationship between growth, temperature and chlorophyll-a concentration. Finally, the authors tested how the growth trajectories predicted by the models varied over the growing season.

This systematic comparison of the models, their functions, predictions, and data sources allow Bahlburg *et al.* to conclude that the main reason for the different spatiotemporal predictions are systematic differences in the model calibration data. Most importantly, the data had different means of generation (remote sensing, *in situ*, biogeochemical simulation models), came from different regions and spanned over different time periods.

Although a mechanistic understanding of krill growth is challenging due to the complex krill life cycle, this comparison provides a starting point for exploring and testing different assumptions in a more systematic way. Bahlburg *et al.* also suggest that for large-scale predictions, either a single model should be recalibrated for different regions, or several models developed for different regions should be combined.

A lingua franca for coherence and testing

Comparing models in a common format is extremely useful in hindsight, but much work could have been saved if the models had been described in the same language when first published, written in a way that is independent of model type, complexity or programming language. The design of new models would have been supported by a comprehensive and accessible overview of the state of the art, leading to more informed, coherent and explicit decisions about what to include in the model. It would have encouraged modellers to justify their choices more thoroughly, in particular why they excluded certain factors that were included by others.

Because of the great potential of a common, non-technical language or format for describing both empirical and mechanistic models, standard formats have been suggested (ODMAP protocol, [4]; ODD protocol,[5]). Accordingly, the recently established Open Modelling

Foundation (OMF[1]) is currently trying to develop, test, and establish such a language. Using such a language for describing models, and making model code accessible, as Bahlburg *et al.* do with their re-implementations, allows modellers to learn from each other, and test models and submodels in more systematic and comprehensive ways. For example, alternative growth models could be compared in terms of their ability to reproduce not only one data set, e.g. a size distribution, but entire sets of observed patterns („pattern-oriented modelling“, [6,7]).

A new culture of modelling

We still have a culture where it is considered more original and valuable to develop new models than to start from existing ones. As a result, it is difficult to see and quantify the limitations of our own models, because in a given context, our models may just look good. Only comparisons like the one presented here [1,2] or elsewhere [8] allow us to place our models in a broader context, to identify limitations and their reasons, and to try to integrate the findings from individual models. Replication and comparison of existing models should be more common and highly valued [9]. Ultimately, we need a culture that encourages learning from each other within and across disciplines, and for this we need better ways of communicating and testing our models. A common format for describing models would be extremely helpful for "anti-siloing" modelling in ecology and elsewhere.

References

1. Bahlburg, D. *et al.* (2023) An intercomparison of models predicting growth of Antarctic krill (*Euphausia superba*): The importance of recognizing model specificity. *PloS One* 18, e0286036
2. Banitz, T. *et al.* (2022) Model-derived causal explanations are inherently constrained by hidden assumptions and context: The example of Baltic cod dynamics. *Environ. Modell. Softw.* 156, 105489
3. Cavan, E. *et al.* (2019) The importance of Antarctic krill in biogeochemical cycles. *Nat. Commun.* 10, 4742
4. Zurell, D. *et al.* (2020) A standard protocol for reporting species distribution models. *Ecography* 43, 1261-1277
5. Grimm, V. *et al.* (2020) The ODD protocol for describing agent-based and other simulation models: A second update to improve clarity, replication, and structural realism. *JASSS* 23, 10.18564/jasss.4259

---

[1] https://www.openmodelingfoundation.org/

6. Grimm, V. and Railsback, S.F. (2012) Pattern-oriented modelling: a 'multiscope' for predictive systems ecology. *Phil. Trans. R. Soc. Lond. B* 367, 298-310. http://dx.doi.org/10.1098/rstb.2011.0180

7. Gallagher, C.A. *et al.* (2021) From theory to practice in pattern-oriented modelling: identifying and using empirical patterns in predictive models. *Biol. Rev.* 96, 1868-1888.

8. Alexander, P. *et al.* (2017) Assessing uncertainties in land cover projections. *Glob. Change Biol.* 23, 767-781

9. Thiele, J.C. and Grimm, V. (2015) Replicating and breaking models: good for you and good for ecology. *Oikos* 124, 691-696. http://dx.doi.org/10.1111/oik.02170

Figure

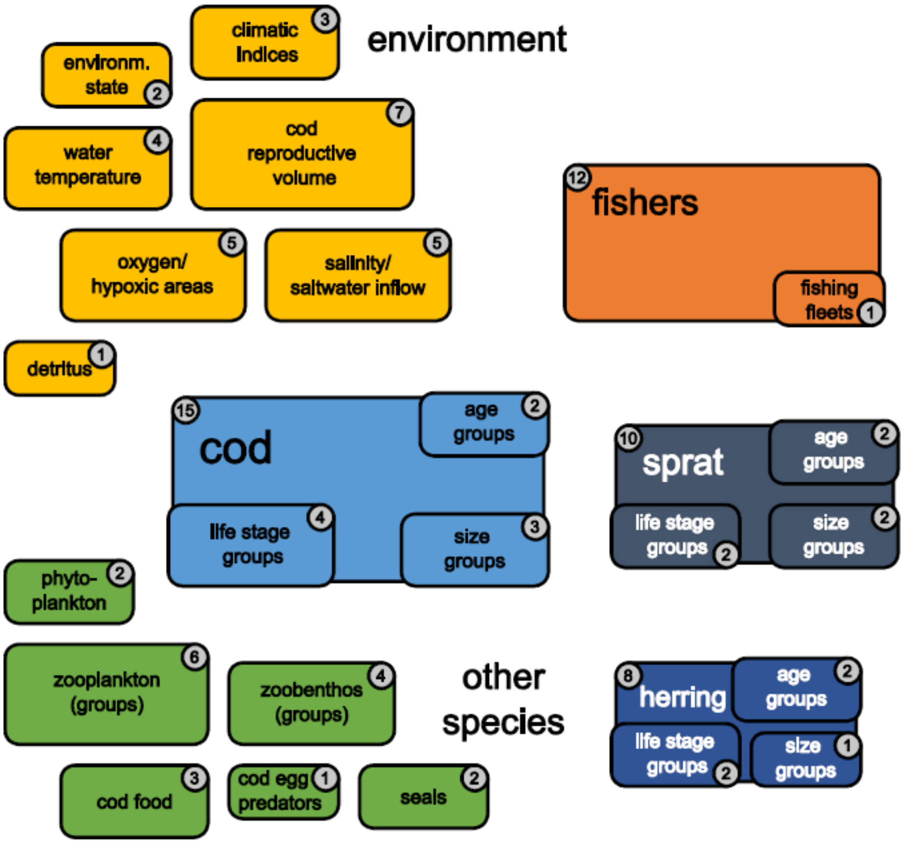

Figure 1. Entities represented in 15 models addressing the dynamics of the Baltic cod (*Gadus morhua*). Box sizes and numbers indicate how many models include this entity. The smaller boxes within larger ones show the representation of that entity via age, size or life stage groups, and the distinction of fishers into multiple fleets (from Banitz *et al*. [2]).