Forum

Editor's Choice OIKOS

Oikos 126: 1071–1077, 2017
doi: 10.1111/oik.03885

Subject Editor: Gerlinde De Deyn. Editor-in-Chief: Dustin Marshall. Accepted 19 March 2017

On the missing link in ecology: improving communication between modellers and experimentalists

Jan Heuschele, Mikael T. Ekvall, Patrizio Mariani and Christian Lindemann

J. Heuschele (http://orcid.org/0000-0001-8203-0542) (janheuschele@gmail.com) and M. T. Ekvall, Dept of Biology, Ecology Building, Lund Univ., Lund, Sweden. JH also at: Dept of Biosciences, Univ. of Oslo, PO Box 1066, Blindern, NO-0316 Oslo, Norway. MTE also at: Chemical Centre, Dept of Biochemistry and Structural Biology, Lund Univ., Lund, Sweden. – P. Mariani, Centre for Ocean Life, National Inst. of Aquatic Resources, Technical Univ. of Denmark, Charlottenlund Slot, Charlottenlund, Denmark. – C. Lindemann (http://orcid.org/0000-0003-3684-2480), Dept of Biology, Univ. of Bergen & Hjort Centre for Marine Ecosystem Dynamics, Bergen, Norway.

Collaboration between modellers and experimentalists is essential in ecological research, however, different obstacles linking both camps often hinder scientific progress. In this commentary, we discuss several issues of the current state of affairs in this research loop. Backed by an online survey amongst fellow ecologists, modellers and experimentalists alike, we identify two major areas that need to be mended. Firstly, differences in language and jargon lead to a lack of exchange of ideas and to unrealistic mutual expectations. And secondly, constraint data sharing, accessibility and quality limit the usage of empirical data and thereby the impact of ecological studies. We discuss ways to advance collaboration; how to improve communication and the design of experiments; and the sharing of data. We hope to start a much-needed conversation between modellers and experimentalists, to further future research collaboration and to increase the impact of single ecological studies alike.

Synthesis

A fruitful exchange between experimentalists and modellers is essential for progress in ecology. We tested this using a bibliometric analysis and by studying the background of highly cited ecologists. We found that studies and researchers that combine both experimental and modelling approaches gain more citations.

Unfortunately, many obstacles often hinder a successful collaboration between the fields. Conducting a survey among ecologists we were able to pinpoint two major problematic areas: jargon differences and limited data exchange. We discuss ways and ideas on how to overcome these hurdles and how to improve the collaboration in ecology, hoping to start a much-needed discussion in the field.

The collaboration between modellers and experimentalists is essential for progress in biology. In this commentary, we want to point to several obstacles in the research loop linking both camps. We think that through improving communication and understanding across those boundaries progress in ecological research could easily be facilitated. The initial idea for this commentary came from the lingering feeling that modelling articles are often written in a language foreign to most experimentalists and that experimental results and methods are often described in cryptic ways and thus hard to use for modelling purposes. To not limit ourselves to our own narrow point of view, we conducted an online survey, which confirmed that other scientists, modellers and experimentalists alike, share our impression of the current state of affairs. To illustrate the benefits of collaboration and integration of research approaches we also conducted a bibliometric analysis of ecology articles and explored the methodological fields of highly cited ecologists.

This essay is by no means a specific criticism of the work of modellers or experimentalists. Rather it is meant as an encouraging reminder to consider using different experimental approaches, to improve data documentation and to ease the description of model and experimental papers. Ultimately we would like to promote collaborations between modellers and experimentalists in ecology by identifying present obstacles.

As ecologists we aim to describe and predict the patterns we observe in nature, with the ultimate goal to get a better understanding of nature, its components and their interplay. We use field and experimental data as input to new and existing numerical models to describe the system (Flynn 2005, Allen et al. 2007). These models can then be used to falsify hypotheses and produce new testable ones about the driving mechanisms of the system. Other models are used to simulate scenarios and to predict future states. Similarly, some models are used to simulate specific systems where available data is limited, or when multiple feedbacks and non-linear interactions make data interpretation difficult. While not all models need explicit input data, the great majority of models use 'real world data' in one way or the other. Model output in turn can stimulate testable hypotheses and thereby serve as a seedling for new experiments. Thus, a loop between modellers and experimentalists is necessary to advance our understanding of the biological processes and entire ecosystems (Fig. 1). Few biologists are able to learn, conduct and excel in both modelling and experimental studies, or even to communicate comprehensively in between those two fields. There are multiple reasons for this including, different traditions in practicing research (Allen et al. 2010), different use of language, and a lack of understanding or scepticism regarding the technique (Flynn 2005), but also merely due to time constraints in learning new techniques. An exception to this lack of understanding is the use of standard statistical methods, which are part of the study curriculum and are applied by both modellers and experimentalist. Given this division of labour, communication between modellers and experimentalists is crucial for the scientific process (Connor and Simberloff 1986, Brewer and Smith 2011, Vincx et al. unpubl.) and has been suggested as a stepping-stone towards improving model-based management applications (Dyble et al. 2008). Yet, we think that the feedback and the collaboration between both 'camps' are far from being perfect and could be improved substantially.

Definitions

Before we start, let us first get some definitions out of the way. The terms used to describe professional fields such as experimentalist or modeller can have different meaning to the different scientific communities and a good amount of miscommunication in between these methodological fields stems from different use of language (Flynn 2005, Allen et al. 2010). We define an *experimentalist* as someone describing the system and testing hypothesis by means of manipulating a system or an organism in the laboratory or in the field. Approaches include, but are not limited to, studies under conditions close to the natural environment (in situ experiments), semi enclosed systems such as mesocosms, as well as more abstract experimentation, where laboratory studies test biological rates from individuals all the way down to the level of genes and metabolic pathways (in vivo or in vitro).

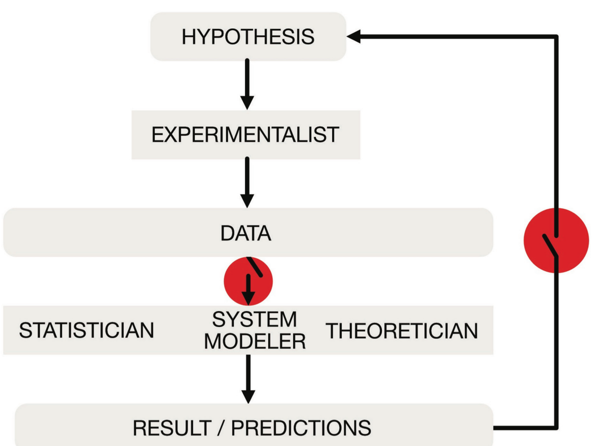

Figure 1. Illustration of the research loop involving both experimentalists and modellers. Red areas indicate the main obstacles we identified.

We define a *modeller* as someone who describes a system and tests hypothesis by constructing a theoretical framework or quantitative numerical models (in silico). She mostly uses mathematical tools to test questions or simulate specific biological systems. There are different levels of abstraction and scales that can reach from describing metabolic pathways to earth-system models. A theoretician derives and tests general ideas from first principles and does also not necessarily rely on empirical data. However, also he or she will draw inspiration from published data of experimental and observational studies.

In contrast, simulation models are often designed as a predictive tool, and require experimental data for parameter estimates and field data for boundary conditions and to validate the model. Statistical modellers analyse empirical relationships of data. This can involve simple statistics or can evolve into more complex models, which may serve the same purpose as simulation models. We chose to exclude the use of common (inferential) statistics from the definition of a being a modeller, as they are used by experimentalist and modeller alike and serve to support findings gained through a specific method, rather than being the method itself.

Methods

The survey

The online survey was accessible from 28 August 2014 to 12 September 2014. An invitation was send out per email, Researchgate and Twitter. We asked both camps (modellers and experimentalists) about data usage, research inspiration and their opinion about the collaboration between the fields. The questionnaire consisted of a general section with three questions and two sections (each with nine questions) aimed at either modellers or experimentalists. All questions and answers are available in the Supplementary material Appendix 2. In total 92 biologists participated (Supplementary material Appendix 1 Fig. A1–A2). Given that most were ecologists and that modelling as well as experimental approaches differ between biological fields, we further reduced the dataset by including only ecologists. Of those, 23 had a modelling background, while 42 described themselves as experimentalists. In addition to answering multiple choice and check box questions, most participants used the open comment section to describe their own experiences and impressions.

Bibliometric analysis

To test whether there was a difference in the citation impact, collaborative index and the number of authors between experimental, modelling and combined approaches, we searched the ISI web of knowledge (Core collection) for all ecology related papers published in 14 leading journals from 1990 to 2010 (Supplementary material Appendix 1). This yielded a total of 7415 articles. From these, we took a random and unbiased subset by choosing every fifth article (1484 articles), and then classified each article using information in the title, abstract and keywords into three main categories: modelling article (including studies with statistical models of non-manipulative field observations), experimental article, and articles that used both experiments and modelling. Subsetting the dataset was necessary as the classifying all articles was unfeasable. We then conducted a bibliometric analysis using the R-package 'bibliometrix' (Aria and Cuccurullo 2016), with which we calculated the summary statistics of citations and authors for each group as outlined in Table 1 (more details can be found in the Supplementary material).

Background analysis of highly cited researchers

To examine the background of the top cited ecologists we identified ecologists from Clarivate analytics (formerly Thomson Reuters) list of 'Highly cited researchers 2016' (<http://hcr.stateofinnovation.thomsonreuters.com>) in the category Environment/Ecology (n = 113) and sorted them into three groups: 1) researchers that only apply modelling approaches; 2) researchers that only apply experimental approaches in the laboratory or in the field; and 3) researchers that apply both modelling and experimental approaches. We further discriminated between researchers that collect their own observational data from non-manipulative descriptive studies (e.g. monitoring) and those that do not use observations or obtain them from other sources (e.g. databases).

Data deposition

Data available from the Dryad Digital Repository: <http://dx.doi.org/10.5061/dryad.nk35d> (Heuschele et al. 2017).

Results

Bibliometric analysis

Our bibliometric analysis revealed that the methodological approach influences the number of citations of the article (df = 2, F = 3.54, p = 0.029). Studies combining experimental and modelling approaches, and pure modelling articles have on average a higher citation count compared to experimental articles, but not when compared to each other. The number of citations is further influenced by publication year (df = 1, F = 165.55, $p < 2 \times 10^{-16}$), probably due to the fact that older articles could accumulate more citations over the years. The number of co-authors per publication is not different between the groups (Kruskal–Wallis $\chi^2 = 2.17$, df = 2, p-value = 0.34). However, the collaborative index of mixed approach articles is higher compared to both purely modelling and experimental articles (Table 1).

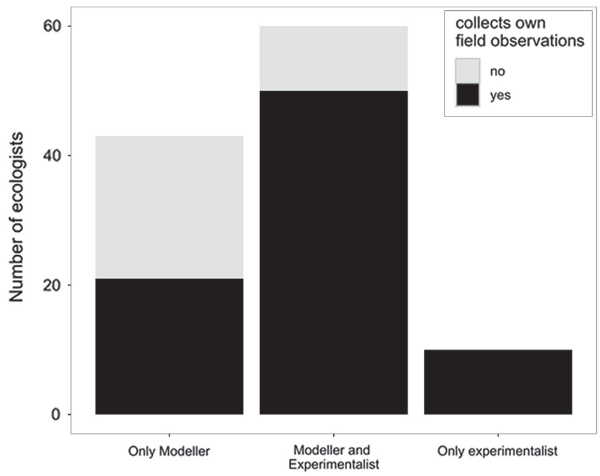

Figure 2. Number of highly cited ecologists that are either modellers, experimentalists or uses both approaches. Field observations are measurements from non-manipulative descriptive studies.

Background analysis of highly cited researchers

The analysis of the highly cited researcher shows that ~53% (n = 60) of all highly cited ecologists work with both, experimental and modelling approaches and most of them also collecting their own (descriptive) field data. Only about 9% (n = 10) of all highly cited ecologists do not use models of some kind and all of those collect their observational data, whereas ~38% (n = 43) of highly cited ecologists use models, but only around half of these collect their own data (Fig. 2).

Questionnaire

The results show that we are not alone in perceiving a problem between the fields. Only 30% of the modellers and 7% of the experimentalist think that there is no problem in the field. All others mentioned that they see some issues that have room for improvements. Most participants pointed to the same problematic areas: *missing knowledge or assumptions about the other approach*, *communication* and *data related*

Table 1. The results from the bibliometric analysis of 1484 articles from 14 leading journals that were grouped into different scientific approaches based on their abstract, title and keywords.

	Complete search for "Ecology"	Every 5th article	Modelling article	Experimental article	Article with experimental and modelling
No. of articles	7415	1484	901	404	33
Sources	14	14	14	13	11
Average citations per article	55.93	55.27	55.69	48.32	73.53
Authors	14861	3946	2555	1114	94
Articles per author	0.499	0.37	0.356	0.363	0.351
Authors per article	2	2.66	2.81	2.76	2.85
Co-authors per article	3.06	3.08	3.13	2.99	2.88
Collaboration index	2.3	3.03	3.16	3.07	3.58

issues. The questionnaire revealed that modelling papers make up only 18% of the studied scientific literature for most questionnaire participants that identified themselves as experimentalists (Supplementary material Appendix 1 Fig. A3). The majority of experimentalists stated that especially the methodology of modelling papers is difficult to understand (Fig. 3A, Supplementary material Appendix 1 Fig. A4). They thus draw inspiration for new research ideas mainly from their own or other experimental research (Fig. 3B). In contrast, the biggest inspiration for modellers comes from discussions with colleagues, and to equal extent from reviews, modelling and experimental studies.

In the questionnaire, modellers answered that access to ecological data is of medium difficulty (Supplementary material Appendix 1 Fig. A5). While modellers seem to use all kinds of input data from linear functions, to categorical data and summary statistics, most would prefer raw data (Fig. 3C). In general, experimentalists seem to be happy to share their data (Tenopir et al. 2011, Supplementary material Appendix 1 Fig. A7), consensus in our questionnaire is, however, that this process should be made easy and that they should also be acknowledged for providing data (Fig. 3D, Klump et al. 2006). While some experimentalists share their data in electronic appendix at the journals, fewer make use of public databases (Supplementary material Appendix 1 Fig. A7). An overview over the databases used by modellers in our survey can be found in the Supplementary material Appendix 1 Fig. A6. The questionnaire further revealed that some experimentalists are concerned that modellers use their data inappropriately, for example without realizing or acknowledging the limitations of the data. More information about the participants and their answers can be found in the Supplementary material Appendix 3.

Discussion

Communication and the exchange of ideas and data is crucial for a fruitful progress in biology. However, looking at the results of the questionnaire and our own impression the actual exchange between modellers and experimentalists seems to be limited. In the following we identify several areas that contribute to this suboptimal situation and also present possible remedies.

Language

Differences in language and jargon in the different fields are likely a big part of the problem (Flynn 2005). Especially the lack of in depth mathematical knowledge of experimentalists hampers the information exchange and communication. Many modelling papers are written in mathematical jargon, and often require a deep understanding of the equations

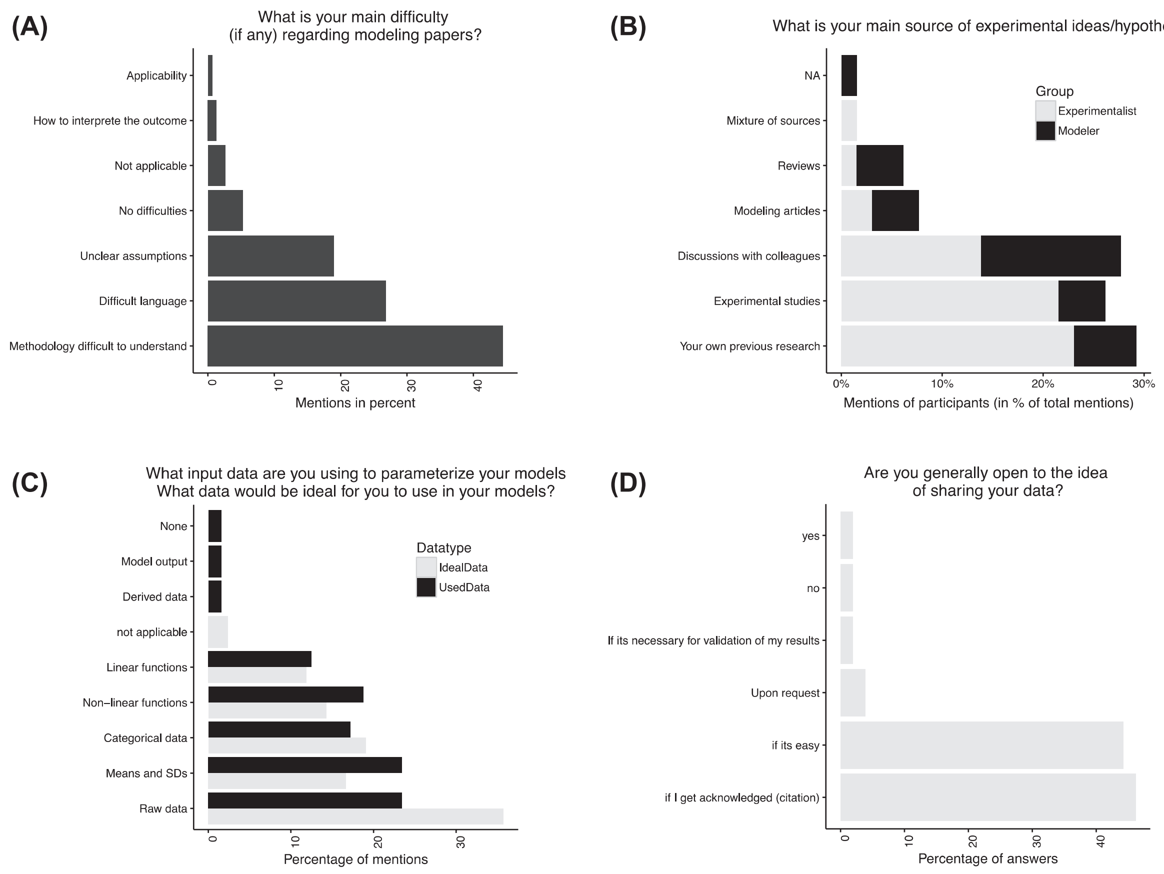

Figure 3. (A) The main difficulties of experimentalists with modelling papers. (B) The main source of experimental ideas/hypotheses for experimentalists and modellers. (C) Ideal and actually used input data of modellers. (D) Answers to the question if experimentalists are open to the idea of sharing the data.

presenting the model. This probably originates from the fact that modellers traditionally have a background in physics and mathematics and later get interested in problems in biology. While this might be the most concise and precise way to present your research, it can alienate experimentalists with limited mathematical knowledge. Likewise, the specific jargon and focus of work used by experimentalists can be hard to understand by modellers. This shortcoming does not only include the technicalities of the experimental setup itself, but also the lack of description of environmental conditions, which can be highly relevant for a proper interpretation of the results, not only with regard to modelling but also in the light of future insights. New research might for example reveal a previously disregarded factor (e.g. state of the animals, specific light condition) that influenced the results. Here an exhaustive reporting of experimental factors, in addition to those immediately relevant to the experiment, can improve the applicability of the data to models (Allen et al. 2010) and could explain inconsistencies or patterns in future meta analyses.

Mutual expectations and understanding

Whereas the previous points directly relate to the use of language or a technical understanding of the other field, other issues are more related to the perception of the limitations and focus of the specific approaches. The reduction of explanatory factors in modelling work is much more obvious than in experimental studies. A model might for example assume that 'zooplankton is a community of grazers', while in reality the group also includes some carnivorous and piscivorous species. This can lead to the perception that many models are generally unrealistic and useless. Yet a reduction of parameters is often necessary to test ideas and to understand a specific question. Especially theoreticians, who aim to understand very general mechanisms and patterns, tackle questions that are hard or impossible to test without having very narrow assumptions. While these might not explicitly fit to any specific study system, they are relevant to improve the general understanding. A similar reduction of explanatory factors is done by experimentalists. A 'community of grazers' in a mesocosm experiment will also likely miss out on some carnivorous and piscivorous species. These limitations can be addressed using results from other empirical studies, while such small simplifications in models can gradually affect model outcomes with growing number of iterations (error-propagation) and when increasing model complexity and scale.

In the case of large simulation models, such as ecosystem, end-to-end, or earth-system models, the model itself often includes numerous parameters. While these models will always deliver results, they are hard to interpret and prone to errors due to the heavy parameterization (Franks 2009). They also often include tuning or fudge parameters, which are used to fit model output and validation data. These factors can be useful despite the lack of a mechanistic foundation. To experimentalists, however, they can appear as an easy shortcut to fit models without adding to the understanding of the system, and thus are hard to sell to experimentalists. And again, when looking a bit closer we can find that experimentalists also might prescript conditions that are not found in nature and therefore cannot reflect a mechanistic reasoning from an ecological point of view. For example, individuals are extremely unlikely to find themselves swimming in test tubes, while in the nature. However, trusting that the behaviour of individuals under those laboratory conditions resembles those in nature, results are scientifically valuable and contain important insides. Hence, for both approaches making assumptions and simplifications is necessary to untangle the complexity of nature.

Given these shortcomings we can think of several ways and areas to improve the communication and collaboration between modellers and experimentalists, which we will present below.

Improving the communication

A better communication of model ideas and structure would improve the integration of modelling and experimental work. Modelling articles could include for example information boxes that highlight key assumptions in 'plain text', and thus facilitate the 'non-computational' ecologists' access to the paper. The end of articles could also include follow-up questions, limitations of the model, point towards missing input data and new testable hypotheses. While explicitly pointing out the limitations of your study sounds like a harmful behaviour in the current publishing landscape where limitations are often hidden deep in the text, it would improve the usefulness of the study. As a side effect, it could increase the number of cross citations between the two fields.

Teaching

In the long run, improving mathematical (and computational) teaching in ecology would allow all fields to gain a better understanding of mathematical methods in ecology, in other words increasing mathematical and computational literacy in the field. This aspect is a key qualification in future research and could be done by simultaneously teaching both approaches in ecology courses in an integrated way. Courses could for example start out from basic ideas and questions, then develop a conceptual model and end with the implementation of a quantitative model. The necessary model assumptions and choices could then be tested with and informed by simple laboratory experiments. In addition, input data for model parameterization should be generated by experiments. If 'real' experiments are unfeasible one could also think of replacing them with simulation experiments. A good topic for such an approach is for example population dynamics. A predator prey community could be modelled using predation rates supplied by data from simple experiments that determine movement patterns, handling times and capture rates. The degree of guidance with model formulation would depend then on the amount of time allocated to the course. A recent overview over active learning approaches in mathematical biology is given by Waldrop et al. (2015).

Improved mathematical education would also lead to a new generation of ecologists which can excel and combine both modelling and experimental approaches themselves.

Active collaboration between fields

Conferences and journals

In the survey both modellers and experimentalists asked for more communication platforms. Suggestion included that these could take the form of joint conferences and symposia that highlight integrative studies, and focus on the advantages and challenges of these approaches. In the same spirit dedicated workshops bringing together modellers and experimentalists should be organised to tackle real ecological problems.

Another way to communicate would be to form journals that publish exclusively collaborative studies, although most available ecology journals already welcome them. We think such initiatives would advance the field of ecology, especially since many modellers get inspired from conversations with other colleagues (Fig. 3A) and hopefully experimentalist will get too.

Data sharing, usage and quality

Data is the fuel of science; without it no theories could be verified or tested. Most models rely on data in one way or the other. Even theoretical ecologists get inspired by the available observations and experimental data, despite of the fact that these are not always used directly in their models.

Most of the data in ecological papers is provided as summary statistics, with limited use as input for models. The same is true for statistical tests, where non-significant results are often described by simple p-values. However, old analyses can contain useful information especially for meta-analyses and overcome publication bias of positive results. But instead they live a secret existence in dusty drawers. Luckily journals and the EU nowadays increasingly request the raw data and better summary statistics. Given that we cannot know about coming analytic methods, the only future-proof way is to provide the raw data, which is also preferred by over 30 percent of the modellers (Fig. 2C). We think that dedicated data repositories like Dryad and Pangaea are preferable over adding data in the electronic appendix of each specific article. This is for example well established in molecular biology (e.g. GenBank, EMBL). Databases have the advantage that they can be indexed, provide a DOI, and can also be linked to a centralized directory. This makes the search for data sets more efficient, and includes the possibility to directly extract data from figures. Another advantage is that data papers are fully citable, and thus allows experimentalists and field ecologists to add to crucial metrics (h-factor, citation count) used in hiring procedures and grant evaluation. This is especially important as gathering data is often tedious and long-winded, and involves many working hours. The results of our questionnaire suggest that a significant proportion of scientist might be unaware of sharing possibilities in databases (Supplementary material Appendix 1 Fig. A7). Issues involving the topic of data archiving in ecology and its requirements are not new and have been covered in more detail by Whitlock (2011). Besides access to raw data, the standardization in reporting of data seems nowadays less than ideal. Data standardization has been discussed and implemented for example in systems biology (Brazma et al. 2001, SBML Language) and genetics (FASTA-Format), but ecology has not had the same discussion.

The questionnaire also revealed that some experimentalists are concerned that modellers use their data inappropriately, for example without realizing or acknowledging the limitations of the data. This could be partly caused by the aforementioned fact that ideal input data is hard to find for modellers and they therefore often have to work with what is available. Modellers also often have a non-biological background, and are not aware of biological constraints and pathways. Hence a more accurate description of the experimental set-up and its limitations could reduce this perceived risk. Another way forward to overcome the inappropriate use and lack of usable data could be including modellers in the experimental design phase. While this is already pointed out during basic statistical courses, it is far from being the standard in practice. And with regard to theoreticians it is in our experience even done less. Such collaboration would improve results, save time and increase the 'productivity' of ecological studies. Furthermore, it would help to improve the availability of the right kind of data for modellers. It could also provide experimentalists with a better understanding of the possibilities and limitations of models, and would allow them to adjust the experimental design to optimize experimental feasibility and data quality. The success of such an approach is apparent when we look at the citation rates of researchers and groups that combine both approaches. In the best-case modellers and experimentalists are under one roof or even within the same group, however, especially in smaller universities or research institutes, this might not be always the case. Therefore, good data sharing practices and comprehensive model descriptions are even more important to actively contribute to the latest developments in ecology.

Summary

In ecological research, a lack of communication and collaboration between modellers and experimentalists hinders scientific progress. Here we pointed towards the main obstacles and laid out possible remedies, from improving modelling articles, over improving mathematical literacy of students, to the creation of communication platforms, changing the design of experiments and the ways of sharing of data. We hope that this essay initiates a discussion between modellers and experimentalists, and ultimately leads to more collaboration in ecology.

Acknowledgements – We wish to thank Prof. L.-A. Hansson for discussions on the topic and his comments on an earlier version of the manuscript. We also want to thank Heidi Sjursen Konestabo and Kirsten Borse Haraldsen for discussions about the bibliometric analysis.

References

Allen, J. I. et al. 2007. Quantifying uncertainty in high-resolution coupled hydrodynamic-ecosystem models. – J. Mar. Systems 64: 3–14.

Allen, J. I. et al. 2010. Marine ecosystem models for earth systems applications: the MarQUEST experience. – J. Mar. Systems 81: 19–33.
Aria, M. and Cuccurullo, C. 2016. bibliometrix: a R tool for comprehensive bibliometric analysis of scientific literature. – Scientometrics 1: 1–17.
Brazma, A. et al. 2001. Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. – Nat. Genet. 29: 365–371.
Brewer, C. A. and D. Smith 2011. Vision and change in undergraduate biology education: a call to action. – Am. Assoc. Advancement Sci., Washington, DC.
Connor, E. F. and Simberloff, D. 1986. Competition, scientific method and null models in ecology: because field experiments are difficult to perform, ecologists often rely on evidence that is nonexperimental and that therefore needs to be rigorously evaluated. – Am. Sci. 74: 155–162.
Dyble, J. et al. 2008. Environmental controls, oceanography and population dynamics of pathogens and harmful algal blooms: connecting sources to human exposure. – Environ. Health 7.2: S5.
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.
Franks, P. J. S. 2009. Planktonic ecosystem models: perplexing parameterizations and a failure to fail. – J. Plankton Res. 31: 1299–1306.
Heuschele, J. et al. 2017. Data from: On the missing link in ecology: improving communication between modellers and experimentalists. – Dryad Digital Repository, <http://dx.doi.org/10.5061/dryad.nk35d>.
Klump, J. et al. 2006. Data publication in the open access initiative. – Data Sci. J. 5: 79–83.
Tenopir, C. et al. 2011. Data sharing by scientists: practices and perceptions. – PloS One 6: e21101.
Waldrop, L. D. et al. 2015. Using active learning to teach concepts and methods in quantitative biology. – Integr. Compar. Biol. 55: 933–948.
Whitlock, M. C. 2011. Data archiving in ecology and evolution: best practices. – Trends Ecol. Evol. 26: 61–65.

Supplementary material (available online as Appendix oik-03885 at <www.oikosjournal.org/appendix/oik-03885>). Appendix 1–3.