Responses to reviewers

Data-based, synthesis-driven: setting the agenda for computational ecology

REVIEWER 1

The manuscript by Poisot et al. is an opinion/synthesis paper about "computational ecology". The authors explain why it is crucial to develop this "new" field and describe how this should be achieved. In particular, the authors emphasize rightly the need to better link ecological data and computer models.

Thought I find this manuscript interesting I have three types of general criticism:

Is all this very new? My understanding is that 20 years ago a large part of the scientific community working in ecology was somehow despising complex computer models, basically because these models were thought as too complicated to help understanding reality and because computer programs were deemed not enough reliable... However, I think that already for some years the community has been recognizing the value of computer models to increase ecological understanding and that some papers have already pointed out the value of these models.

> **RESPONSE:** We have revised the text significantly to (i) highlight the novelty of some approaches and (ii) stress the need to define "best practices" for the application of computational techniques in ecological research.

Many points are interesting in the manuscript but are not enough supported by detailed arguments or examples. For example, it is explained that mathematical models cannot take into account the complexity of ecological systems, e.g. stochasticity, but it is not fully explained why. Similarly, no successful example of computational ecology project is detailed enough to fully convince of the value of computational ecology. The authors also rise rightfully about important issues, e.g. the need to increase data sharing, but do not really give new solutions to these problems.

> **RESPONSE:** We have re-written large sections of text to make this clearer, notably by adding references to many recent studies that have benefited from these approaches. We do not believe that there is a single study that will result in a "eureka" moment about the usefulness of these techniques, and so we have instead decided to give a broad survey of the field.

I see all approaches in ecology as a continuum and the only way for me to develop ecology is to recognize the usefulness and shortcomings of all approaches and to use these approaches in a complementary way. I find that somehow the manuscript does not recognize enough that mathematical models in ecology have some virtue and that conversely computer models have some drawbacks. For example, theoretical mathematical model are hard to parameterize and their results are difficult to compare quantitatively to real data, but qualitative comparisons are also useful. Computer models are very flexible and very good at making precise quantitative predictions but these predictions require a huge amount of data for each new ecosystem / species for which predictions have to be made.

> **RESPONSE:** We agree with this comment, and we have revised the text at multiple place to give appropriate credit to mathematical models, and to highlight current possible issues with computational methods (notably surrounding error propagation).

REVIEWER 2

SUMMARY

Section 1 gives Species Distribution Models as a flagship example of computational ecology: having started as purely statistical, this field incorporates more and more theoretical concepts, and is implied (more than said, though that may qualify as common knowledge) to be rather heavily computational in the types of tools and practices employed to do so.

Section 2 attempts to better define the boundaries of computational ecology and its relationship to other branches of ecological research. Mainly, integration of data and models emerges as the determining factor; a second distinction of suggesting rather than documenting mechanisms is proposed, but its concrete implications are less clear.

> **RESPONSE:** We have vastly expanded this section, and give more references to empirical studies. Section 2 has, in fact, been almost entirely re-written.

Section 3 then lists recommendations toward ecological synthesis and collaboration, which largely fall into two bins:

1. data best practices, including metadata and common formats, reproducibility, data sharing and how to foster and credit it.

2. sensitivity training to quantitative skills, including fostering a healthy (and critical) relationship to code and software among empiricists, as well as more dialogue between the three poles of empirism, computation and theory.

More in-passing remarks concern the ability of computational ecology to provide tools relaying fundamental research toward its many applications (predictions for "action ecology" and stakeholders, etc). This paper starkly (sometimes wittily) states a number of problems, and emphasizes important directions in solving them (such as the problem of properly crediting and encouraging data producers, as their data increasingly must be reused beyond what they could hope to achieve on their own).

GENERAL THOUGHTS:

This is an enjoyable read, with a worthwhile discussion of the conditions under which computational endeavors in ecology can exist, prosper and expand the field in general. The calls for 1) more open, inter-communicable and reusable (hence far less wasteful in terms of scientific effort) data practices and 2) more integration of data and theory (in training as well as research), are both very important and ever timely.

Being myself computationally minded, and therefore part of the cheer squad here, I defer to more worldly ecologists on the question of whether this paper's content is adequate for its goals. I do wonder if these concerns have not been voiced before in a similar fashion, and whether this paper is constructed so as to change minds that were not already on the authors' side. Of course, it is always useful to restate the problems at stake here, and the review-like quality of parts of this paper provides a trove of references and semi-concrete suggestions.

Less importantly, I am not so clear on what the initial sections, trying to define or characterize computational ecology, really achieve, and in particular how computational ecology is meant to differ from any theoretical work that is not entirely data-free.

> **RESPONSE:** We have made numerous changes to the introduction, and rewritten most of section 2. We hope that this important change will clarify how computational approaches differ from and complement other approaches – see in particular the new figure.

An obvious distinction would be a pragmatic one of techniques and training - crudely, a question of volume of data (and code required to handle said data) - rather than content. But I feel that the paper is also trying to suggest such a

distinction of content, which I cannot quite grasp. For instance, would Hubbell's neutral theory, or Maximum Entropy style models, be considered computational ecology? If they do have such a distinction in mind, I would like the authors to find a starker way to express it, as it could improve the dialogue between computational and theoretical ecologists in the future (an interface which is far less addressed than the one with empiricists).

Some questions:

What's new since Pascual 2005? Many references and specific proposals listed here are recent, but has the field of computational ecology in general changed shape or direction somewhat since then?

> **RESPONSE:** We have added a few discussions of the changes since Pascual 2005. Basically, there has been an explosion in techniques and computing approaches, which ecologists have been slow (or at least slower than other fields of ecology) to react to.

Optionally, I think it would be interesting to provide some discussion of epidemiology here, since I feel it already does (for a certain type of population dynamics) a lot of what is proposed for the intersection of "action" and computational ecology, in particular short-cycle predictions, tools for stakeholders, and so on. Looking at these practices, if they are not too far removed from the authors' field of expertise, could help make new concrete suggestions, and perhaps recognize some opportunities or problems that computational ecology may face in the future.

> **RESPONSE:** This is a great point. We discuss computational epidemiology as a good roadmap for ecologists to follow at the end of section 2.2.

MINOR CONTENT COMMENTS:

Fig 1: The locations (or labels) of meta-analysis and observation are a bit perplexing at first. If anything, I would expect the typical observational work to have typically less integration with models than a meta-analysis (depending on when statistics qualify as modelling in your perspective, which seems to me to oscillate a bit throughout the text) If by "observation" you mean the paragraph about Sallan et al., then I understand better, but I would suggest making the relationship between the figure and the main text more obvious. Perhaps due to this, the paragraph on meta-analyses in page 6 seems slightly disconnected from the rest.

RESPONSE: Figure 1 has been removed.

Table 1: I am not sure what you mean by accuracy, for it to be high for mathematical models, observation and meta-analysis all at once. I am also unsure about what is "indirect" suitability for prediction. As for the invisible costs of computational infrastructure, I can see that they are immense at the social level, as noted at the end of the discussion, but its seems to me that most of these costs are not borne by researchers and therefore not really relevant to choosing a research practice (obviously pure math is somewhat cheaper, but to caricature, even a pure mathematician requires a lot of infrastructure to access publications from across the world, to say nothing of chalk and coffee)

RESPONSE: Table 1 has been removed.

page 8: I would like a bit more clarity on Ovaskainen's proposal (translating ecological concepts into digital representations); right now it seems very vague.

RESPONSE: We have added more detail about this study.

Paragraph "With or without a common data format…" I am not clear on what you have in mind here, talking about prediction propagating "from one analysis to the other" Do you mean the fact that computational ecology reuses pre-analyzed rather than raw data? That a single computational ecology project itself tends to have multiple analytical steps? Or somtehing else?

RESPONSE: We have expanded this section (but it has been heavily modified during the re-writing).

WRITING

Throughout, you probably mean "quadrant" rather than "quadrat", although this is a very ecological slip to make. Of course, if this is conscious and tongue-in-cheek, feel free to go with it.

RESPONSE: We have removed this from the revision.

page 2 Computational science is one of the ways to practice computational thinking: Makes me really perplex about what the other ways are "to the point where one would almost…" : That's a very mild statement "ecological systems as complex and adaptive […] a certain degree of stochasticity" -> the link is not so obvious. "modelling, that can often function on {its} own"

RESPONSE: Wee have clarified these statements.

page 3: P(S|E=1) should be P(S=1|E) de novo field samplig -> sampling

RESPONSE: Corrected.

page 6: "it is paradoxically the high degree of model abstraction … that gives them generality". It is not entirely clear to me how 1) this is obviously paradoxical, depending on hwat is meant by generality, and 2) this paradox is explained by the sentences that follow (they are about the fact that data is hard to collect, which is tangentially relevant).

RESPONSE: Clarified.

page7 computaitonal research is dependant-> ent

RESPONSE: Corrected.

page9: I simply wish to wholeheartedly support this beautiful defense of parasites. conservation target -> targetS

RESPONSE: Corrected.

page 10: "and how data were transformed during the analysis" is probably lacking a verb before the "how".

RESPONSE: Left as is after two native English speakers found no issue with the sentence – we remain open to suggestions.