

Interactive comment on “Reviews and syntheses: Parameter identification in marine planktonic ecosystem modelling” by Markus Schartau et al.

Anonymous Referee #2

Received and published: 4 August 2016

The manuscript discusses aspects of parameter identification, in particular the estimation of parameters in marine ecosystem models using data assimilation. Provided is a wide overview of topics like different error sources, variational estimation methods and the corresponding construction of cost functions, error models, and posterior parameter uncertainties. Further, typical model parameterizations are discussed as well as the aspect of cross-validation and model complexity, and space-time variability of parameters. As a further methodological aspect, emulator approaches are discussed. The manuscript also includes a discussion on some aspects of parameter estimation in large-scale biogeochemical ocean models. Overall, the manuscript considers a vast amount of topics and represents a quite comprehensive review of parameter estimation. The manuscript should fit well into the focus of the journal Biogeosciences. Despite the length of the manuscript, many topics are held very short. In this respect the

C1

manuscript represents more an overview paper than a review. Nonetheless, it should be well suited as an entry point to the topic. However, there are several weaknesses in the presentation and the review of the methods is lacking the inclusion of ensemble-based methods. These issues are described in more detail below and should be resolved before a publication of the manuscript.

A major issue of the manuscript is the strongly varying degree of detail in the review of methods and results. While many findings are simply mentioned by citing the corresponding study, others are worked out as examples in more detail. For the examples, for which also figures are included, it is not clear why these cases have been selected. From the text it is not evident that the examples are scientifically particularly relevant. Given that the chosen examples in the different sections appear to be the scientific work of the lead authors of the respective section, I have the impression that the examples are merely chosen to promote the section lead-authors' work. As such the manuscript leaves an odd impression in that the authors most prominently point out their own work, while keeping results from other studies on a shorter descriptive level or even reducing it to a sole citation of a paper. This varying degree of detail should be corrected. For the readers it will be most helpful, if the examples are clearly chosen to illustrate the scientific most relevant aspects.

With respect to varying detail and the provision of examples, Section 9 is a particular case. On the one hand, the section sticks out in the manuscript because it is written in a rather lengthy style compared to the other sections. Here, I see a good potential to be more concise. On the other other hand, the abstract suggests that Section 9 contains a "survey ... to studies that approached parameter identification and global biogeochemical modeling" and that "Parameter estimation results will exemplify some of the advantages and remaining problems in optimizing global biogeochemical models". Unfortunately, this is only partly true. The "survey" is mainly reduced to 13 lines of text in the upper part of page 41. Directly afterwards the text discusses the general issue that the models need long integration times to reach equilibrium (Sec. 9.1) and

C2

then focuses on the special topic of deep particle flux, which is started in Sec. 9.1 and worked out in quite some detail in Sec. 9.2. This is not marked as an example, but again given the many references to papers of the lead authors of this Section, one has the impression that the authors promote their own work instead of reviewing the state of the research. This is even more pronounced by the fact that from about line 13 of page 44 to the end of subsection 9.2 results from the not yet submitted study by Kriest et al. are discussed. Given that the described work is not yet published and hence not yet reviewed (and that I can not review it based on the incomplete description provided in the text) I can only recommend to reduce this part to one or two sentences stating the general finding. The full length of the description in the manuscript is scientifically not justifiable until the study itself is published. Overall, I expected from the description in the abstract a comprehensive overview of results from parameter estimation in large-scale models. Unfortunately, this is clearly missing in section 9. Rewriting Section 9 in this respect would significantly improve the manuscript.

With regard to the methods, the study misses ensemble-based schemes, even though page 4, lines 26-27 state that an "overview of major DA aspects concerning parameter identification" is provided. While several studies that used ensemble-based methods are cited throughout the manuscript, the methodology is only shortly described in Section 7.1.3 on time-varying parameters where the method of "Sequential Importance Resampling (SIR)" (nowadays usually called "particle filter") is shortly explained. A manuscript with the motivation to provide a comprehensive overview of parameter estimation in ecosystem models is clearly incomplete when ensemble schemes are left out.

I have the impression that the authors intentionally left out a methodological description of ensemble-based methods (next to the particle filter also including methods based on the Kalman filters) because the authors do not use these methods and because the methods cannot ensure mass conservation. The mass conservation is already mentioned in lines 20-25 on page. The argumentation in the text that filtering methods

C3

are "infringing" mass conservation, that mass conservation is relevant, and that one hence has to use methods that ensure mass conservation in the data assimilation, which "harmonise well" with corresponding methods in ocean state estimation, is part of a very old discussion, which is apparently followed emotionally (which is consistent with the words "infringing" and "harmonise" chosen by the authors). I'm not aware of any study that shows that the change of mass induces errors in the estimation of parameters or issues in the interpretation of the results. Even more, the methods could be used to estimate parameters alone, hence not changing the state directly such that the mass is conserved. My recommendation is that the authors simply avoid this discussion (unless they can provide scientific evidence) and revise the text accordingly.

While the main part of Section 10 summarizes the discussions of the manuscript, the sub-sections 10.1. and 10.2 do not fit with the main part. These sub-sections are not summaries a parts of the main text, nor do they show clear perspectives for the manuscripts' topic of parameter estimation. These aspects would better fit into the introduction section in order to discuss the different aspects of model parameterizations and the interplay of measurements, modeling and data assimilation.

Some more detailed Comments

Abstract; last sentence: I cannot see that the recommendation to find "...a good balance in the level of sophistication between mechanistic modelling and statistical data assimilation treatment..." is a result of the study. Either the authors should remove the statement or revise the text so that this statement results from reviewing the methods and application studies.

Page 6, lines 10-12: I have the impression that "weak constraint" and "strong constraint" are not general expressions used "In the geophysical community", but only in connection with data assimilation. Please consider changing the statement (Unfortunately, I cannot check the two cited books, as I don't have an easy access to them). BTW: Please cite books with providing a chapter or page range.

C4

Page 10, line 1: Here "model discrepancy" and "model inadequacy" are mentioned. For readers it would be very helpful if the text could actually explain what these quantities are. The text states that this is part of an "important initiative" (page 9 line 32), but the description is not really more than mentioning the expressions and referring to two papers.

Page 17, line 31-32: The text states that "The prominence of MCMC methods for data assimilation is described by Rayner et al. (submitted)." Actually, while doing data assimilation for quite a while, I'm not aware of any "prominence" (BTW: this is a typo, it should be "prominence") of MCMC in this field. As data assimilation is usually concerned with high-dimensional models, the application of MCMC is not feasible. Please correct your statement.

Page 19, lines 1-4: Here, it is mentioned that "Two approaches to point-wise approximation of U are found in ... modeling studies" followed by mentioning the approaches. Unfortunately, references are missing for this statement. While in the following subsections some references are provided for methods based on the Jacobian, no paper is cited for the Hessian-based methods.

Page 29, lines 27 and 33: Please provide a reference for the "Akaike Information Criterion" as well as for the "weighted AIC" and the "Bayesian Information Criterion"

Page 35, line 15: It is stated: "It's flexibility could equally well be increased by increasing the size of the parameter vector, rather than allowing it to vary in time". It is unclear whether this statement is a result of some study (which would require a reference), or whether it is just speculation?

Page 42, lines 20-24: Here, the text states that "... three types of data are considered essential for model assessment and calibration" and then lists the data types. I wonder what is the scientific basis for this claim? Unfortunately no paper is cited. Please provide references to support this claim.

C5

Page 47, line 20: The text mentions "dynamical and statistical emulators". Given that most readers are not familiar with these emulators, it would be helpful if each type is shortly explained.

For completeness of the review, please also consider the recent paper Simon et al. *J. Mar. Syst.* 152 (2015) 1-17, which is also concerned with parameter estimation in an ecosystem model.

Interactive comment on Biogeosciences Discuss., doi:10.5194/bg-2016-242, 2016.

C6