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

Markus Schartau et al.

mschartau@geomar.de

We appreciate the time spent and the comments provided by Referee #2. Major concerns raised by Referee #2 were respected and we introduced considerable changes to our manuscript.

Comment* 1 *by Referee #2:
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.

Author's response: Our aim was to provide examples that describe and stress particularly relevant aspects. In practice, the lead author (M. Schartau) asked each section to include an example figure that both he and the section author(s) considered to be particularly relevant/illustrative. In most cases these examples have come from the sections authors' work or extensions thereof. This is understandable since each section author is a specialist in that area and is most familiar with his or her own work. However, we appreciate that this may have led an unfortunate impression of promoting our own work, and we thank the reviewer for pointing this out. To counteract this impression we have made a number of changes to improve balance in coverage of different approaches and to maximise breadth of relevance in the figures. However, we feel it is acceptable for section authors to use examples from their own work since it is for these that they have best understanding and control. Here is a summary of the figures in the revised manuscript:

1) We have moved the original Fig. (1) with the example of a variable lag fit (VLF) to the Appendix (Fig. A1) and substituted it with a figure from the study of Simon and Bertino (2012, Journal of Marine Systems, 89, 1-18). This figure is a nice example of the improved asymptotic behaviour of a deterministic Ensemble Kalman Filter (DEnKF) when using log-transformed observations and model results to realise the analysis

step. Ehouarn Simon and Laurent Bertino kindly provided their results so that we could redraw the figure. We will send a request to Elsevier for using the redrawn figure. The figure is referred to in Sect. (4) about Error Models. The new Fig. (1) (send upon request) is based on results from a sequential, ensemble based, data assimilation (DA) approach, which should further improve balance with respect to ensemble based DA methods.

2) In Sect. (7) we address space-time variations in model parameter estimates and we find it appropriate to include a figure (Fig. 2 of Losa et al., 2006), based on results from Losa et al. (2004). It is prominent and illustrative example of variable parameter values in the North Atlantic. We will send a request to Elsevier for using their figure. In the text we refer to this figure (Fig. 5) as follows:

Nevertheless, Losa et al. (2004) were able to document the plausibility of their posterior photosynthesis parameter values for the maximum phytoplankton growth rate (μ_m in Sect. 3.1) and intial slope of the P-I curve (α_{phot} in Sect. 3.3) by comparison with observational estimates of Platt et al. (1991). Six parameters were optimised in all and the posterior parameter fields were cross-validated in a 3D version of their model by comparing the output with an independent SeaWiFS chlorophyll data from 1997-2003 (Losa et al., 2006). The spatially-varying parameter set of Losa et al. (2004), obtained by assimilating Coastal Zone Color Scanner (CZCS) data for the period 1979-1985, was interpolated and extrapolated onto the spatial grid of the 3D model as shown for the two parameters relevant for phytoplankton growth, μ_m and α_{phot} respectively (Fig. 5).

3) We eliminated the orginal Fig. (7) in Sect. (9) on the probability densities of the climatological phosphate, nitrate and oxygen data. Instead we put more emphasis on explaining the relevance of Fig. (8) with the projections of the parameter-cost function manifold.

***Comment* 2 *by Referee* #2**:
With respect to varying detail and the provision of examples, Section 9 is a particular case. One the one hand, the section sticks out in the manuscript because it is written in a rather lengthly style compared to the other sections. Here, I see a good potential to be more concise.

Author's response: Sect. (9) is revised entirely. The text has already been considerably reduced by removing details and by skipping the original Fig. (7) that showed probability densities of nutrient concentrations. The detailed illustrative example in Fig. (8) is kept. This example nicely connects aspects explained in new Sect. (3) (former Sect. 5, Typical parameterisations of plankton models and their parameters), recalled with an example in new Sect. (5) (former Sect. 4, Posterior parameter uncertainties).

***Comment* 3 *by Referee* #2**:
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.

Author's response: Ensemble-based schemes were in fact briefly referred to in

the Theoretical Background section: "Similar concerns apply to parameter estimation for stochastic dynamical models, where fully Bayesian approaches appear to be favoured using computational strategies based on sequential Monte Carlo methods (van Leeuwen, 2009; Jones et al., 2010; Dowd, 2011; Doucet and Robert, 2013; Dowd et al., 2014). "

but we admit that this was not adequate coverage given the importance of these methods. We have therefore replaced this sentence with a new subsection "Sequential methods" (Sect. 2.2.2) in the Theoretical Background, which expands on the methodological material previously included in Sect. (7.1.3). This new subsection (2.2.2) reads as follows:

"In some problems, assimilating all the data at once from all available sampling times can be computationally impractical. This is particularly likely for models with stochastic dynamics ($\eta \neq 0$ in Eq. 1), if the data are clustered in time, or if model states need to be repeatedly updated as new data come in. In such cases a sequential approach can be expedient. The basic idea is to break the large integration problem defined by Eq. (7) into a number of smaller problems by sequentially assimilating observations in subsets defined by sampling time. The method comprises a consecutive sequence of two major steps, a forecast- and an analysis step respectively. If the sequential approximation or 'filter' is accurate, it should approximate the posterior distribution defined by Eqs. (6 and 7), when all data have been assimilated by the end of the assimilation period. To see how this works, suppose we know the probability density $p(\vec{x}_j^t \mid \vec{y}_{1:j}, \Theta)$ of the true state at sampling time t_j (possibly an initial condition) for a given value of the uncertain parameters Θ and given all the previously assimilated observations $\vec{y}_{1:j}$ (possibly null). The probability density at sampling time t_{j+1} is given by the forecast density:

$$p(\vec{x}_{j+1}^t \mid \vec{y}_{1:j}, \Theta) = \int p(\vec{x}_{j+1}^t \mid \vec{x}_j^t, \Theta) \cdot p(\vec{x}_j^t \mid \vec{y}_{1:j}, \Theta) \, \mathrm{d}\vec{x}_j^t \qquad (8)$$

In general this integral can be approximated by an ensemble of Monte Carlo simulations, sampling an initial condition from $p(\vec{x}_{j+1}^t \mid \vec{y}_{1:j}, \Theta)$ and then running the model to

the next sampling time t_{j+1} (possibly including stochastic dynamical noise, and possibly accounting for kinematic model error). Next, in the analysis step, the new observations are assimilated by applying Bayes' theorem:

$$p(\vec{x}_{j+1}^t \mid \vec{y}_{1:(j+1)}, \Theta) \propto p(\vec{y}_{j+1} \mid \vec{x}_{j+1}^t, \Theta) \cdot p(\vec{x}_{j+1}^t \mid \vec{y}_{1:j}, \Theta), \qquad (9)$$

which again can be approximated e.g. by Monte Carlo sampling. The forecast and analysis steps can then be repeated until all the data are assimilated. A seldom-discussed assumption here is the conditional independence of the observations, allowing us to write $p(\vec{y}_{j+1} \mid \vec{x}_{j+1}^t, \Theta)$ instead of $p(\vec{y}_{j+1} \mid \vec{x}_{j+1}^t, \vec{y}_{1:j}, \Theta)$ in Eq. (9). This amounts to assuming that the observational errors are independent between sampling times (Evensen, 2009), which may not be strictly true if sampling is frequent and if there is a noticeable contribution from representativeness/undersampling, or from errors in conversion factors (see Sect. 2.1.3).

Once the predictive filtering densities $p(\vec{x}_{j+1}^t \mid \vec{y}_{1:j}, \Theta)$ have been approximated for all sampling times (t_j with $j = 1, \ldots, N_t$), these can be used to approximate the likelihood in Eq. (7), since:

$$\begin{aligned} p(\vec{y} \mid \Theta) &= \prod_{j=1}^{N_t} p(\vec{y}_j \mid \vec{y}_{1:j-1}, \Theta) \\ &= \prod_{j=1}^{N_t} \int p(\vec{y}_j \mid \vec{x}_j^t, \vec{y}_{1:j-1}, \Theta) \cdot p(\vec{x}_j^t \mid \vec{y}_{1:j-1}, \Theta) \, \mathrm{d}\vec{x}_j^t \qquad (10) \\ &= \prod_{j=1}^{N_t} \int p(\vec{y}_j \mid \vec{x}_j^t, \Theta) \cdot p(\vec{x}_j^t \mid \vec{y}_{1:(j-1)}, \Theta) \, \mathrm{d}\vec{x}_j^t \end{aligned}$$

For j=1 in Eq. (10) we have a set of zero members and $p(\vec{y}_j \mid \vec{y}_{1:j-1}, \Theta) = p(\vec{y}_1 \mid \Theta)$. In the third line of Eq. (10) again some conditional independence of the observations is assumed and the final integral can in general be approximated using the predictive

ensembles (see Jones et al., 2010; Dowd, 2011; Dowd et. al., 2014). This procedure can be repeated for different values of Θ and combined with Eq. (6) to assess posterior probability. Alternatively, $p(\Theta \mid \vec{y})$ can be calculated from a single application of the filter using a 'state augmentation' approach whereby the parameters Θ are appended to the vector $\vec{x}$ as additional state variables with zero dynamics. In practice, random parameter noise may need to be added to avoid filter degeneracy, such that this approach may be considered a separate estimation method (Dowd, 2011). However, if such ad hoc noise can be avoided, or if the parameters are in fact assumed to vary stochastically, then the augmented-state filter at the end of the assimilation interval should approximate the theoretical Bayesian posterior for this time. For other times, a 'smoother' algorithm would be required. A further benefit of the augmented-state filter is that the parameter estimates for intermediate time periods may show temporal patterns that expose deficiencies in the model formulation and provide useful information for model development (e.g., Losa et al., 2003).

The various types of filter differ essentially in terms of how the integrals in Eqs. (8) and (9) are approximated. Particle filters (van Leeuwen, 2009) use Monte Carlo sampling for both steps while the Ensemble Kalman Filter (Evensen, 2003; Evensen, 2009) uses Gaussian and linear approximations for the analysis step, enabling the use of smaller ensembles but at the cost of lower accuracy in strongly nonlinear/non-Gaussian problems. The (Extended) Kalman Filter applies when the model dynamics are (quasi-) linear and both model and observational errors are Gaussian. These conditions allow both integrals to be evaluated analytically, but appear to be rarely applicable to parameter estimation in marine ecosystem models. For reviews of sequential approaches the reader is referred to Dowd et al. (2014) for marine biogeochemical modelling and to Bertino et al. (2003) for oceanography in general. "

Note that we do not aim to provide technical details on the various filters, partly because these are already discussed in other reviews (to which we have directed the

interested reader), and partly because we want the Theoretical Background to focus on models and general methods, not algorithms and techniques.

***Comment* 4 *by Referee* #2**:
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 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.

Author's response: This discussion, although old, is still of importance. Whether mass balance is achieved or not is relevant, in particular when communicating model results, other than simulated fields of chlorophyll *a* concentrations, to biological oceanographers and marine biogeochemists. The property of mass conservation is more fundamental than any detail in the parameterisations. It is not our intention to evoke any emotional commotion and have therefore rephrased respective sentences. In the second paragraph of the introduction we introduced the following change:
"So far no fundamental ecophysiological principle has been further exacted beyond

the conservation of mass. Whether a balanced mass budget needs to be achieved depends on the scientific problem addressed. Some ecosystem model applications may not critically depend on mass conservation, e.g. when simulating plankton growth to act as food source in regional simulations of fish stock size and recruitment. In biogeochemical models the conservation of mass can be essential, in particular for large-scale or global ocean simulations. A consistent theme running through most ecosystem models is the determination of mass flux of certain biologically important elements, such as nitrogen, phosphorus, iron and carbon (N, P, Fe and C)."

Furthermore, we revised the subsection (Sect. 1.4, Inferences from data assimilation): "Much of the literature on DA in oceanography is focussed on state estimation (e.g., Allen et al., 2003; Natvik and Evensen, 2003; Dowd 2007; Nerger and Gregg, 2008; van Leeuwen, 2010). In these studies, the primary objective is to improve hindcasts, nowcasts, or forecasts of time-dependent variables such as chlorophyll *a* (Chl*a*). However, many of the DA methods originally developed for state estimation have more recently been adapted to estimate static parameters, especially for stochastic models where random noise is injected into the model dynamics. Stochastic noise offers a plausible way to represent model error, but it should be noted that it can lead to violations of mass conservation unless it is injected in certain ways (e.g. by perturbing growth rate parameters). Deterministic plankton ecosystem models guarantee mass conservation and have a longer tradition in parameter estimation for marine ecosystem models, although they imply a less explicit treatment of model error. To identify and gradually eliminate model deficiencies it can be helpful to analyse model state and flux estimates while mass conservation is imposed as a strong constraint. The optimisation of only parameter values assures that simulation results remain dynamically and ecologically consistent, which is comparable with those DA approaches in physical oceanography that produce dynamically and kinematically consistent solutions of ocean circulation (e.g., Wunsch and Heimbach, 2007; Wunsch et al., 2009). "

***Comment* 5 *by Referee* #2**:
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.

Author's response: The sections (10.1 and 10.2) would not fit to the introduction section. Section (10.1) clearly refers to perspectives and it should reflect current tendencies in the development of planktonic ecosystem models. We actually complemented Sect. (10.1), in response to comments by Referee 3. In Sect. (10.2) we summerise our major impression after literature search and after reading many papers that covered diverse aspects. One important take home message is that we found, on the one hand, many studies (with DA methods applied for parameter optimisation) where biological aspects (e.g. basic model assumptions) remained undifferentiated (undiscussed). On the other hand, a series of biologically motivated studies did not consider aspects of parameter identifiability, let alone of DA. We think we have expressed this in Sect. (10.2) and stressed the need to find a good balance between the different scientific communities to which we refer to in our manuscript.

—

***Specific comment* 1 *by Referee* #2**:
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.

Author's response: The implication here is that there is frequently an imbalance in the level of sophistication in these two areas. This was a general impression that we gained from reading the literature, and we feel it could be helpful to report this impression to readers. Probably this imbalancedness is driven by the fact that it is easier to publish a paper and write a successful grant proposal if it purports to be "cutting-edge" and "state-or-the-art" in some particular way, rather than putting a balanced level of effort into *all* methodological aspects. We therefore do not want to follow the Referee's suggestion and leave the statement as it is.

Specific comment 2 by Referee #2:
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, a I don't have an easy access to them).

Author's response: Corrected to "In the geophysical data assimilation community". The first citation in this instance is a paper: Sasaki, Y.: Some basic formalisms in numerical variational analysis, Monthly Weather Review, 98, 875–883, 1970. We have added chapter/page ranges to the book references where possible.

Specific comment 3 by Referee #2:

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.

Author's response: This paragraph has been expanded to:
"Another important initiative is the estimation of hyperparameters of the kinematic error model along with the ecosystem parameters (Arhonditsis et al., 2008). The posterior of the kinematic model error provides an estimate of the model discrepancy, introduced by Kennedy and O'Hagan (2001) and originally referred to as model inadequacy. The model discrepancy is defined as the model error for the "true" values of the model parameters, i.e. the unknown values of the parameters for which the model best represents $\bar{x}^t$. Estimates of model discrepancies may thus provide useful diagnostics for model skill assessment and development. "

Specific comment 4 by Referee #2:
Page 17, line 31-32: The text states that "The prominance 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 "prominance" (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.

Author's response: DA is not exclusively concerned with high-dimensional models and MCMC methods have been and still are commonly applied. We admit that the statement is awkward and the reference is not exclusive. We suggest to simply remove the sentence.

Specific comment 5 by Referee #2:
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.

Author's response: Yes, some references would be helpful. We added two useful references:
" The matrix $\mathcal{H}_\Theta$ is the Hessian whose elements are second derivatives of $J(\Theta)$ with respect to the parameters (e.g., Tziperman and Thacker, 1989; Matear 1995): ... "

Specific comment 6 by Referee #2:
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"

Author's response: We included the original reference (Akaike, 1973) for AIC, refer to the work of Johnson and Omland (2004) for the log-likelihood approximation based on residual sum of squares, and use Burnham and Anderson (2004) as an example reference for the bias-variance trade-off:
"One of the simplest techniques (in terms of its applicability), is the Akaike Information Criterion (AIC, Akaike, 1973). The AIC considers two opposing terms corresponding to the maximum log-likelihood of the parameters given the data ($\ln[L(\widehat{\Theta} \mid \vec{y})]$, measuring model data misfit) and a bias-correction factor, that increases with the number of free parameters (N_Θ).

$$AIC = -2 \ln \left[L \left(\widehat{\Theta}_p \mid \vec{y} \right) \right] + 2N_\Theta \qquad (11)$$

C13

Note that for a model fitted by least-squares, the log-likelihood can be approximated by the residual sum of squares (RSS), following Johnson and Omland (2004): $\ln[L(\widehat{\Theta}_p \mid \vec{y})] \approx -N_y/2 \cdot \ln(RSS/N_y)$, with N_y being the total number of observations. The AIC, and alternative techniques (weighted AIC, or Bayesian Information Criterion, BIC), seek to quantify the trade-off between bias and variance (e.g., Burnham and Anderson, 2004). "

Specific comment 7 by Referee #2:
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?

Author's response: The statement queried is not the result of a study but was intended simply as a reference to a general concept that is covered in other parts of the text. It is not needed here specifically and has been removed.

Specific comment 8 by Referee #2:
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.

Author's response: We thank the referee for identifying this mistake. Here, the verbalism is inappropriate and we realised that it does not reflect what we intend to state. There is no such "claim". We suggest to revise the entire Sect. (9) and to correct this statement in a new subsection (9.2 Data availability):

"In regard to the ocean's key role in global carbon cycling and hence for the climate system, three different types of data can be considered for model assessment and calibration: 1) data of dissolved inorganic tracers, e.g. distributions of nutrients, oxygen, alkalinity and dissolved inorganic carbon, 2) measurements or data products of rates, e.g. of planktonic primary- or net community production, and 3) observations of the gravitational flux of organic particles to the ocean interior, transporting particulate organic matter through the water column. "

Specific comment* 9 *by Referee #2:
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.

Author's response: The "dynamical" emulator was/is already described in Sect. (8.1) and the "statistical" emulator was/is explained in Sect. (8.2). We see no need to again explain the differences between the two in Sect. (10).

Specific comment* 10 *by Referee #2:
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.

Author's response: We thank the referee for alerting us to this useful paper. It is now cited in Sect. (7).