Geosci. Model Dev. Discuss.,
doi:10.5194/gmd-2016-148-RC1, 2016
© Author(s) 2016. CC-BY 3.0 License.

Geoscientific Model Development Discussions
Open Access
EGU

Interactive comment on "Fundamentals of Data Assimilation" *by* Peter Rayner et al.

Anonymous Referee #1

Received and published: 22 July 2016

Comments on "Fundamentals of Data Assimilation" by Rayner et al

General Comments

The authors embark on an important mission, namely to provide a common language with which all data assimilation practitioners in biogeochemistry can speak with one another, as well as detailing certain techniques for this. Though this reviewer acknowledges the difficulty of the task, the present effort is a bit rough around the edges and needs quite a bit of work before it will be useful to the community at large. As such, I recommend major revision before publication.

Generally speaking, my major complaints about the manuscript are threefold: 1) There are very few places where examples of the techniques described being used in biogeochemistry are cited, except those by the authors themselves. Readers of this document would be best served to have a list of examples in hand to better understand

Printer-friendly version

Discussion paper

CC BY

how the field has employed these techniques. Particularly in Section 7, where specific algorithms are detailed, at least a modest list of relevant papers is called for. 2) The manuscript seems to meander between vague generality and over-specificity. In particular, the bits that describe ideas relevant to the relaxation of prior assumptions and dependence on priors are much more specific and thoroughly cited than the rest. 3) The discussion of MCMC methods in 6.2 would be better placed with the material in section 7, which should be renamed something like “Implementations of the Theory” or something similar. This is a computational approach to sampling the posterior distribution, rather than a “general principle”. This is likely to help the reader to better understand the notion of the posterior distribution as “the solution”, but it will also likely tempt them to believe that this sampling method is the “right way” to get at the solution, which is certainly a problem specific conclusion to draw.

Specific Comments

Page 2, Lines 29-: The task of defining probability measures for the non-specialist is certainly nontrivial. The discussion of probability density function here is a bit confusing, especially since most will not understand what you mean by “continuous space”, which I believe is that every interval in (0,1) has a preimage in the probability space, so that it makes sense to define a CDF. Without having to go into Radon-Nikodym derivatives, it’s enough to define a CDF, and then define the PDF as its derivative, which is what you’re doing. Perhaps a rearrangement of the words here would serve this purpose.

Table 1, entry for R: I believe the notation should be U(y-H(xt)), where the parenthetical bit is not a subscript.

Section 4: citation of Tarantola (2004) should be for the year 2005. I also think that given the heavy reliance on his developed theory, it may be worth pointing to his original 1982 paper as well as the edition of his book from 1987 (?), both of which are more readable by those new to the field.

Printer-friendly version

Discussion paper

CC BY

Page 7, Line 16: Remove “target” from the first sentence and merely reference a quantity of interest.

Page 7, Line 18: It’s probably worth stating outright that the PDF being computed is a marginal PDF, since you later say in step 5 to calculate the PDF for the quantity of interest. Another reason for this computation is that it’s the product of the sensitivity and uncertainty that matter, and ensures an “apples to apples” comparison between different potential parameters.

Page 8, Line 2: Though the example is instructive, I’m not sure what purpose the last sentence serves, unless more information about the recommendation is given, such as what he’s trying to optimize with this choice.

Page 8, Line 19: “more limited formal link” If the point is to remove reliance on subjective priors, then what are they being replaced with? A more honest sentence would be something like “replacing a formal link with an empirical one” or something similar.

Page 9, Line 7: “Absent such direct verification calculations like sensitivity analyses or ensemble experiments give incomplete guidance” This sentence would read better if the “like” were replaced with a comma.

Page 10 Equation 3: Should it be p(B) rather than P(B)?

Page 10 Line 6: Should it be p(x|B) rather than p(B)? I’m not sure how we infer the MLE of p(B) from equation 3.

Page 13 Lines 14-17: This seems like a very good place to cite the synthesis inversion literature as a great body of examples of this technique for the biogeochemistry applications.

Page 15, Line 27 Across all fields, the common nickname for these techniques is “EnKF”. To enable readers to connect this text to others in their area of specialty, it seems using the more common name would be most useful.

CC BY

Page 15, Lines 32 to Page 16, Line 3: This was true for the initial formulations of the EnKF by Evens and others. Modern implementations favor a “deterministic” formulation that doesn’t perturb observations, such as the Ensemble Adjustment Kalman Filter (EAKF) and the Ensemble Square Root Filter (EnSRF). Tippett et al (2004) is a good reference for this topic.

Page 16, Line 6: “ensemble method may capture nonlinear impacts on the state covariance” I have heard this but never seen evidence. Is there a relevant citation? Mathematically, the covariance in equation 5 is exactly the covariance of forecasted state, using the jacobian rule for propagating uncertainty.

Page 16, Line 15: p(x)ˆn should be p(xˆn)

Interactive comment on Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-148, 2016.

Printer-friendly version

Discussion paper

CC BY