Atmos. Chem. Phys. Discuss.,
https://doi.org/10.5194/acp-2018-1081-RC2, 2018

Atmospheric Chemistry and Physics Discussions

Interactive comment on "Fundamentals of Data Assimilation applied to biogeochemistry" *by* Peter J. Rayner et al.

Anonymous Referee #2

Received and published: 6 December 2018

The manuscript by Rayner et al. presents an overview of data assimilation methods that are applicable to biogeochemistry and how these methods have roots in Bayesian inference. The manuscript covers many aspects on the topic and the writing style is easy to read. I feel, however, that the overall message of the paper is unclear, and the contribution of the manuscript is somewhat lacking. This paper straddles the line between being a tutorial for a novice versus being a review of studies that have employed these methods. In terms of being a tutorial, this manuscript would not be detailed enough for a beginner – both in terms of the mathematical notation not being rigorous enough as well as not showing each step of the implementation of different methods. Because this is meant to be applied to biogeochemistry (although I would argue that that title should be expanded to include atmospheric sciences as well for NWP, trace

Printer-friendly version

Discussion paper

gases, etc) and not be a generic statistics manual, what is needed is a practical guide on this field. What are examples of physical concepts in biogeochemistry/atmospheric science for each component of the assimilation system (e.g., models, target variables, uncertainties)? What are the hurdles? Which methods work best in which situations? Schematics for teaching purposes? In terms of it being a review paper, it is not nearly comprehensive enough. The examples are often (though not entirely) from the authors' own groups (and therefore one perspective), and that is not enough for a review paper. As such, I found it difficult to review because the intention was not clear. I would recommend this paper to be re-written with a clear purpose in mind, such as being a very detailed tutorial about how to apply different methods to biogeochemical/atmospheric problems with rigorous notation (such that a user wouldn't inevitably end up having to go to other sources to practically implement their problem). This would greatly increase the value of the paper.

I will elaborate further in the major and specific comments below.

1. I was not convinced that that 'all of the methods in widespread use within the field are special cases of the underlying Bayesian formalism'. The manuscript often switches between Bayesian methods not found within biogeochemistry but found within atmospheric sciences more generally (e.g. particle filtering), hybrid Bayesian methods (e.g. Michalak et al. 2005) and non-Bayesian methods (e.g. Manning et al. 2011) with little distinction between them.

2. Section 3 has huge potential but does not deliver as one of the main contributions listed in the abstract. The notation needs improving rather than simply reiterating. For example, I do not agree that the notation is 'sufficient for most practical cases' as it is neither followed throughout or sufficient for a tutorial. An example is the discussion of hyperparameters – there is no notation available in Table 1 to represent a vector of hyperparameters (I would suggest a bold theta). I think that this is a good opportunity for an explicit notation for the MAP estimate vs the mean. The notation in Table 1 is not precise enough. For example, some of the notations are specific to Gaussian

Printer-friendly version

Discussion paper

distributions. This would be fine if Table 1 were being used to only discuss a special case of Gaussian problems, but later in the manuscript, non-Gaussian formulations are also discussed, which would require a new notation.

3. Section 4. As this section is fundamental Bayesian theory and if this is a tutorial, there should be a rudimentary explanation of the notation of p(A given B) and how (in basic terms) this forms Bayes' theorem. It then worth reserving the notation p(x) for the prior probability of x.

4. Section 5.6. Here is a good place to introduce the new notation suggested in Section 3 and to expand the definition of Bayes' theorem from Section 4. The references used as examples are not conducive to the narrative thus far. The paper Michalak et al. (2005) does not integrate out the hyper-parameters and is therefore this is not a hierarchical method but an empirical hierarchical method (see e.g. Statistics for Spatio-Temporal Data by Cressie & Wikle).

5. Section 6.2. Gibbs samplers always have higher acceptance rates in that by design, the Gibbs sampler has an acceptance rate of 1. The real gain is not due to sampling from a univariate distribution but sampling from distributions where there is a closed form expression for the conditionals that can be sampled from directly. The disadvantage is that for many situations, this is not known a priori. It is not clear whether the work 'adaptive' on page 12, line 3, is meant to refer to Adaptive MCMC. Adaptive MCMC methods have a particular meaning, which is not the same as that described here. It would be better to simply state that improved strategies use gradient information while sampling. It is important to note that these methods maintain ergodicity while sampling.

6. Section 7.2. This needs to discuss non-Gaussian problems, especially given the focus on Bayesian inference. For example, should this posterior uncertainty represent the highest posterior density region around the mode or an equal weighted probability region around the median? This brings to light the difficulties when quoting the

posterior mean for non-Gaussian posteriors.

7. Section 8. The final sentence on 'future methods' could have much more discussion. At the moment, only one study is referenced but not really discussed, so it reads more like an afterthought.

8. Section 8 reads like a 'review paper', but it isn't comprehensive enough. I would suggest that if this is meant to be a tutorial, to not try to be a review also, because section 8 could be a whole paper in and of itself.

Specific comments:

P1 Title: Data and Assimilation should be lower case

P1 Line 4: . . .for automating part of the process. i.e. the choice of prior distribution is not automated.

Page 1 Line 12-14: An idea is introduced here but the readers are left hanging. Why introduce this example of improving a model but not testing it, but not provide any explanation of what is meant here.

Page 2 Line 5: awkward wording "demonstration how these many methods are its implementations"

Page 2 Lines 21-23: Why are only these few papers referenced? There is a wealth of literature on applications of the theory to different fields and it is unclear why the authors select only four to represent their fields.

P3 Line 7 (Eq. 1): Define xi.

Section 2.2 capitalize non-Bayesian

Page 3 Line 19: Explain what is meant by the "Replicate Earth Paradigm"

Section 2.2 should come before Section 2.1? This narrows the remainder of the discussion to Bayesian.

P4 Table 1: This isn't referred to anywhere in the text. Delete extra brackets in description of R.

P5 Line 20 (Eq 20): Make clear that the earlier reference to one variable, one observation has now been expanded into vectors.

Figure 1. Put all text into the caption below. Why is the x axis labelled "unknown" in the bottom panel? Label panels (a) and (b). The numbers 1.2 and 0.8, 0.2, -0.2 are not immediately clear what they are showing.

Page 5 Line 26-27: describe what "Equation 2 not well applied" means

Page 7 Line 5: Need to describe what you mean – what are the common misunderstandings

Page 7 Line 10: Reason 1: Why does limiting target variables underestimate the uncertainty? Reason 2: This is just repeating what was said above and is not a reason.

Page 7 Line 15: Point 1: Explain what this means. How do you make this decision? Point 3: An example would help. Point 4: What cut-off?

Page 7 Line 15: Walking through an example setting up this "ideal" world would be much more helpful for a novice reader than only providing the instructions.

Page 8 Line 9: K can be either sign so why would sK be more likely to increase rather than decrease? i.e. a larger s means a more negative sK if K were negative. I may be misunderstanding what is being said here, but isn't it simpler to say that positive scaling has a minimum of 0 so skewed in one direction while log(s) can be both positive and negative?

P8 Line 24: The text does not describe what a uniform prior is (i.e. a uniform distribution). If this is for a novice, needs to be explicit.

Page 8 Line 19: Discussion about aggregation errors missing. No reference to methods that try to diagnose these (e.g., Turner et al., 2015, Lunt et al., 2016)

Page 8 Line 26: Another disadvantage is that it is not fully hierarchical and that hyper-parameters are not integrated out.

Section 5.2. Need a discussion about the uncertainties assumed in the prior as this makes a very large impact. In practice, this is not well known. This could be a practical note on the application to biogeochemistry.

Page 9 Line 11: It is not only resolution that affects 'H' but model structures such as parameterizations.

Section 5.4. Need much more detail about how uncertainties are treated in H.

Page 10 Line 25: Need to state that this is not a fully hierarchical Bayesian method if hyperparameters are not integrated out, and state that the impact is likely an underestimation of uncertainties

Page 10 Line 29: This study did integrate hyper-parameters in a hierarchical sense and propagated these uncertainties through to fluxes, but requires MCMC calculations with potentially higher computational cost.

Page 10 line 32: What is a "well-known" atmospheric inversion?

Page 11, line 1: If this is a tutorial, then the description and notation should be included in this paper and not just referenced to another paper.

Page 11 Line 12: A new altered Figure 2 would help show this point

Page 11 Line 13: Delete 'various of'

Page 12 Line 19 – Can talk about the role of thinning chains.

Page 12 Line 5- Need references using MCMC

Page 12 Line 5 - Need to discuss limitations of MCMC such as convergence issues.

P13 Line 3: It isn't clear what the word 'model' means here. For a tutorial, an example of what is meant would be helpful.

Page 13 Line 5 – typo MCMC?

Page 13 Line 16 – Needs more explanation – such as?

P13 Line 27: The notation G(H(x)-yˆo, R) in this section is different to that defined in Table 1 and does not appear Ide et al (1997).

P13 Line 29 (Eq. 5): I would change p(x) to something else here, e.g., p(x|y), to distinguish it from the prior probability of x

P14 Line 16: Specify where this will be described rather than simply 'later'.

P15 Line 20 (Eq. 7): Is the meaning of J described before this? What is J? What does it mean? Explain for a beginner.

Section 8: Reserve the term 'we' for subjective choices made specifically by the authors rather than the community in general, as e.g. "We can identify...", "We now know. . ." etc.

P 19 Line 23: Capital T needed at start of sentence.

P21 Line 10: Delete 'apparently'

Interactive comment on Atmos. Chem. Phys. Discuss., https://doi.org/10.5194/acp-2018-1081, 2018.