Atmos. Chem. Phys. Discuss.,
https://doi.org/10.5194/acp-2018-1081-AC1, 2019
© Author(s) 2019. This work is distributed under
the Creative Commons Attribution 4.0 License.

Atmospheric Chemistry and Physics Discussions
Open Access

ACPD

Interactive comment

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

Peter J. Rayner et al.

prayner@unimelb.edu.au

Received and published: 8 June 2019

Printer-friendly version

Discussion paper

CC BY

Response to Referees' Comments

Peter Rayner, Anna M. Michalak and Frédéric Chevallier

June 8, 2019

We thank the three anonymous referees for their comments which have allowed us to clarify various points in the paper. There are some common points made by several referees. We will deal with these first as general comments then with particular comments from each referee. We place referees' comments in typewriter font and our replies in Roman

General Comments

There is a general concern from all three reviewers about the purpose and target audience of the paper. The critiques diverge e.g. (paraphrasing) "The paper is not rigorous enough" (R1) or "Section 2.2 is too general" (R3). Unfortunately, as authors, we cannot assume from this divergence that we have the balance about right. It could well be that the "sweet spot" we seek does not exist. Obviously we do not think so and our experience in teaching, both at the summer school that spawned this special issue and more generally supports this view. The question of mathematical rigour is particularly difficult. The usual route to a well-grounded description of probability and its manipulation is via an undergraduate course in measure theory. Very few likely readers of

Printer-friendly version

Discussion paper

this paper will have undertaken such a course. Reviews or even text on data assimilation do not generally dedicate the space to provide this level of background. We have instead started from a few precepts:

1. The basic concepts are rather intuitive if demonstrated carefully;
2. The best starting point is the axioms of probability, Bayes Theorem is rather a consequence of these. This approach mirrors both Jaynes and Bretthorst (2003) and Tarantola (2005). We currently note this in Section 3 and have added a comment to the introduction.
3. there really is no distinction in principle between problems given names like data assimilation or inversion. As such we very much thank the reviewers for pointing out inconsistencies in language. We have corrected these and expanded on the problem description in the introduction to clarify these relationships.

The other general critique touches the role of the paper as a review, tutorial or perspective. Reviewers have suggested we should survey the current state of the art and describe future directions. This does not fulfil the paper's role within the special issue. The other papers in the issue do precisely this for their fields, usually by presenting a state-f-the-art example. There seems little value in summarising those results here. We do, however, accept that more forward-looking material on the general problem is warranted and have expanded the relevant section.

CC BY

Specific Comments

Reviewer I

First we have dealt with the general comments above although the reviewer makes some points as part of these to which we respond directly.

`For example, the term “target variable” is used on page 5, line 13 without any prior definition.` We have added text defining this and strengthening the link between the definition of events and PDFs.

`Furthermore, the manuscript makes references to "the inversion (or inverse) problem" many times without explaining what the inverse problem is (it is only mentioned briefly in Introduction on page 1, line 16).` We have switched to the term “data assimilation” or “assimilation” throughout and added text in the introduction on the equivalence of the various names used in the literature.

`For example, a schematic of the data assimilation process and the different components would be helpful and could be referred to when describing the different parts of the system.` This is an idea we have thought about during the preparation of the paper. Unfortunately there seems so little commonality in data flow among the various implementations of the theory. For example, highly efficient methods for calculating MAP estimates might use the adjoint of a cost function derived from some exponential PDF. They may never calculate a likelihood or probability. Other techniques that map or sample the posterior PDF may only calculate likelihoods or probabilities. Any form of the diagram would need so many variants that it would add confusion rather than clarity.

`I also think some concrete examples when describing the different implementations would help readers to follow the`

CC BY

whole process better. We have already noted that the article forms part of a special issue. The task of producing a new textbook in this area (essentially updating the 2000 AGU Monograph) is well beyond our scope here.

2. I think the manuscript would be clearer if the authors choose one perspective from the outset (e.g. data assimilation) and present all material from this perspective in a consistent manner. Currently the manuscript seems to borrow many terms from data assimilation, but frequently, especially in later parts, the language switches to that of inverse modelling (e.g. page 10, line 13: "We need therefore to incorporate these extra variables into the inversion process"; there are many other parts where "inverse problem" and "inversion" are mentioned). Section 6 is even named "Solving the Inverse Problem". I recommend to replace all occurrences of "inversion", "inverse problem" etc. with terms that are more common in data assimilation. see general comments above.

3. Data assimilation is probably best known for its applications in numerical weather prediction, where the technique is used mainly to improve meteorological initial con- ditions to produce better weather forecasts. In biogeochemistry, on the other hand, data assimilation (and related techniques) are more commonly used to constrain parameters. This difference is alluded to in the manuscript (e.g. page 15, lines 3-10). However, I think it would be better if this distinction is explicitly stated in the beginning of the manuscript. This difference explains e.g. why the manuscript does not focus on the dynamical model (the

CC BY

dynamical model for the target variables in biogeochemical applications is often unknown or assumed to be persistence, while the forecast model is an essential component in atmospheric data assimilation). A broader discussion about the choice of assimilation time window would also be helpful.
We do not accept the distinction drawn here between data assimilation as used to estimate different target variables, indeed one purpose of the paper is to demonstrate that the distinction is superficial. We now argue in an expanded section on future developments that so-called "dual state" approaches that seek to constrain the state and parameters of the model should be more widely used in biogeochemical problems. We have added further commentary on assimilation windows at the end of Section 7.

4. The authors write that "we will not be using mathematically precise language" (page 2, line 17). I can see where the authors are coming from, but I think this does a disservice to the readers. I recommend the authors to remove this sentence and to be mathematically rigorous to the extent feasible. I understand a mathematical precise language will take away some of the simplicity, but I think the benefits outweigh the added complexity; currently the reader may be left wondering where the language is imprecise and be less likely to refer back to the text when e.g. implementing a data assimilation system.
See general comments above.

5. The manuscript does not talk much about the issues of ill-posed problems, which are common in inverse problems, and the need for regularization (except for under "Historical Overview", page 19). I consider the use of prior error covariances, e.g. errors with a specified correlation length scale, to be a form of regularization. Even if the

Printer-friendly version

Discussion paper

"true" error correlation length scales are smaller, a larger correlation length scale may be necessary for the data assimilation system to converge to a solution. This constraint will on the other hand lead to larger aggregation errors. It may also be worth to add a discussion about how the number of observations influences the design of the data assimilation system, e.g. the choice of regularization. The term “ill-posed” has several meanings in mathematics, here we think the author means problems that are not numerically well-behaved, e.g. with poor convergence. We have dealt with the use of other regularisation methods in the historical overview. We do not know of cases where deliberately incorrect choices were made for prior uncertainties in order to facilitate convergence.

Specific comments

1. Page 3, lines 23-25: "As a practical example a frequentist may estimate a mean by averaging his sample while a Bayesian may calculate an integral over her probabil- ity density." I do not think this is a good example of the difference between Bayesian and frequentist statisticians. A better example could illustrate how Bayesian and fre- quentists interpret probabilities (e.g., a frequentist may only consider the long-term frequency of occurrence of a random event, while a Bayesian may draw from other prior information to assign probabilities, even for non-repeatable events). The reviewer’s example is correct, this *is* a difference between the two approaches but we don’t think it the most important one. We think the point important enough to defend. It is well-stated in the preface to Jaynes and Bretthorst (2003) commencing on p.xxii. Jaynes points out here that the machinery we use allows the use of prior information but its

Printer-friendly version

Discussion paper

most important distinction is more fundamental, working with the PDFs themselves rather than estimators pre-derived from assumed distributions.

2. Page 4, line 2: "we have followed it [the notation of Ide et al., 1997] here". It would be helpful to the reader to highlight exactly what is new with the notation introduced in this manuscript. Is it simply an extension of the Ide et al. (1997) notation for bio- geochemistry data assimilation? Is it a generalization? More about this in the next point. We have added "as closely as possible" to indicate we changed nothing substantial from Ide et al. (1997).

3.1. Consider adding a "Remarks" column and note when e.g. a notation differs from the notation in Ide et al. (1997). We have done this by separating the mathematical definition and textual description and expanding the latter where necessary.

3.2. For the definition of G, should mu and U be bold to show that they are a vector and a matrix, respectively? In that case also change "mean mu and covariance U" to "means mu and covariances U". We have changed the typeface. The pluralisation is not correct, the mean of a multivariate distribution is still a mean (singular).

3.3. Descriptions of superscripts "a" and "b": "Posterior or analysis" and "Background or prior". Change to "Analysis or posterior" (to be consistent with "Background or prior"). Done.

3.4 Symbol Q: From my understanding Q is often used to denote model uncertainty (for the dynamical model). "Forecast uncertainty" here seems to include uncertainties due to initial conditions and boundary conditions. This is a good point. $\mathbf{Q}$ should refer only to forecast uncertainty. This is too subtle a point to make in the table so we have referred to the relevant section on the Kalman Filter and spelled this out more clearly there.

`3.5. Description of R: Add "(Observation uncertainty)" or something similar.` Added.

`3.6. Symbol A: I do not think I have seen "A" used for "posterior uncertainty covariance" before. Maybe use` $\mathbf{P}^b$ `for background uncertainty covariance and` $\mathbf{P}^a$ `for analysis uncertainty covariance.` We considered this but it generates other inconsistencies, we should, for example, use the same superscripts for the MAP as well as the uncertainty etc. We believe adding one new symbol to Ide et al. (1997) notation is a better solution. This notation is common throughout NWP (e.g. Desroziers et al., 2005).

`4. Page 6, Figure 1: It took some time for me to interpret this schematic. It may be helpful to mention that the observation operator in this case is a simple 1:1 mapping to the system state.` An excellent idea, done.

`5. Page 10, lines 6-7: "then the dynamical model forms part of the mapping between the unknowns and the observations so is properly considered part of the observation operator". I know that this notation is common in e.g. atmospheric inversion, but I personally think it is unfortunate and easily leads to confusion (as the authors also note on page 9, lines 18-19, data uncertainty or data error may refer to the uncertainty due to errors in both observations and the observation operator, which is misleading). In e.g. 4D-Var for atmospheric data assimilation, I believe the observation operator and dynamical model are usually kept separate in the formulation of the cost`

Printer-friendly version

Discussion paper

function. Given that this manuscript focuses on clarity and fundamentals of data assimilation, I think it would wise to adopt the notation of atmospheric data assimilation and not conflate the observation operator and dynamical model. The reviewer is addressing two different problems here. The first is the role of dynamical models and observation operators. We disagree that the dynamical model error is treated explicitly in conventional 4dvar for numerical weather prediction. It is usually embedded in the background error. We wish to emphasise that the distinction is one of use rather than inherent and we have added a sentence to make this point explicit. The other question about observation and model errors is a different problem which we treat elsewhere. The dynamical model applies to the model state variable (see Eq (1) of Ide et al.). In the case of atmospheric inversion, the dynamical model is sandwiched between operations that are either upstream (interpolation, multiplication with a flux pattern, injection, etc.) or downstream (unit changes, spatial sampling) of it because the target variables $\mathbf{x}$ are not the model state variables $\mathbf{z}$. For this application, it is then appropriate to call "observation operator" the full operation chain, including the dynamical model. It is also important to account for the uncertainty in the dynamical model in the data error budget.

6. Page 10, lines 12-13: "Frequently we regard these [parameters] as fixed, which is likely to underestimate the uncertainty of the estimates. We need therefore to incorporate these extra variables into the inversion process." I think "need to" is a bit too strong; maybe say instead that some methods incorporate these extra variables. Agreed, changed.

7. Page 20, lines 16-19: The authors mention that the time delay in EnKF may be prob- lematic for tracers that live longer than the assimilation window. Is this not a common problem for all implementations? Yes, but it’s particularly severe for the KF which often uses

CC BY

short, rolling windows. We have clarified.

Technical Corrections

1. Page 1, footnote 1: I find the footnote unnecessary and suggest to put this informa- tion directly in the text. We disagree. Footnotes are for text which a reader might use but which would disturb the flow if included in the body. This seems to fit that description.

2. Page 5, line 11: "by an observation operator". May be worth mentioning that this operator is sometimes also referred to as the forward operator., agreed, done.

3. Page 5, line 13: "target variable". Define the term. Now done in the introductory section.

4. Page 7, line 18: "to find variables to which it is sensitive". Replace "it" with "the quantity of interest" or something similar. done.

5. Page 8, line 10: Capitalize "This". done.

6. Page 8, line 10: Fix citation format (no parentheses). done.

7. Page 13, line 14-15: "These estimates will generally yield larger variability than that from our most likely flux". I do not believe "flux" is defined in this context. Maybe change to "realisation". done.

8. Page 14, line 23: "superscript f indicates the application of the forward model". I believe the authors are referring to the dynamical or forecast model here. The forward model is,

from my understanding, often synonymous with the observation operator. This is indeed potentially confusing, we have replaced "forward" with "dynamical" and introduced the mnemonic "forecast" which is the meaning we intend for the superscript

9. Page 15, line 4: "For data assimilation our motivation is to hindcast the state of the system". Consider changing to "For data assimilation applied to biogeochemistry our motivation is often to hindcast the state of the system", or something similar. done as suggested.

10. Page 16, line 30: "Another advantage is that, ...". It is not clear from the previous sentences that the previous statement (need to run a dynamical model for each reali- sation of the ensembe) is an advantage. Maybe change to "An advantage of EnKF is that, ..." The real problem was with the previous sentence where we had not clarified the advantage of the EnKF, now done.

11. Page 19, line 7: Capitalize "possibilities". done.

12. Page 19, line 26: "(references (e.g. Gloor et al., 2000;". Remove "(references" or "(e.g.". removed "references".

13. Page 20, line 16: Change format of citations (no parentheses). done.

Reviewer II

We have responded to most of the general comments in our overall response. Some specific comments are addressed below:

The reviewer remarked that most of the literature cited was from the authors' own work.

CC BY

We don't agree with that critique but also note that, in a small research community, the authors are three active and longstanding researchers with broad collaborations. A strong representation of work which includes them is unsurprising.

General Comments

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 atmo- spheric 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. We acknowledge that there is a trade-off between comprehensiveness and methodological purity but think some of the reviewer's comments here are misplaced. Particle filters and hierarchical methods are firstly Bayesian (even if not all implementations are complete) and in growing use within biogeochemistry. To ignore studies like Manning et al. (2011) risks a false impression of complete methodological unity which is not the case. We would argue that "all of the methods in widespread use ..." is an accurate description at the moment.

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

CC BY

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 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. The reviewer makes two points here, first that the notation isn't good enough and second that we don't follow it throughout. We reiterate the general point that inventing our own notation which we thought perfect would be at best wasted effort and (more likely) add further confusion. The reviewer raises two examples of missing notation. The vector of hyperparameters and the maximum a posteriori (MAP) estimate. We have added θ to Table 1 and a brief mathematical development in the relevant section. We resist the idea of a special notation for the MAP. It is a useful quantity to know but it is often neither a parameter of the PDF and, for sampling approaches, not generated by the method. We believe referring to it as the mode of the posterior PDF is both accurate and parsimonious of notation. The reviewer has not cited a case where we do not follow our own notation. For example, we have a notation for Gaussian distributions and go on to use it being careful to state it is a special though widely-used case. The rest of the methodology we develop using the notation for probabilities we define.

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. Following Jaynes and Bretthorst (2003) and Tarantola (2005) we do not believe that Bayes Theorem is fundamental so do not introduce the topic this way. We do introduce the likelihood

later.

4. Section 5.6. Here is a good place to introduce the new notation suggested in Sec- tion 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). We have now added a brief mathematical development using θ as suggested. We already noted that Michalak et al. (2005) was a stepping stone and have added the classification suggested.

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 disad- vantage 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 informa- tion while sampling. It is important to note that these methods maintain ergodicity while sampling. We think this discussion is taking us too far into implementation details. Our solution is rather to cut back on discussion of different sampling strategies so we have removed the detailed discussion of the Metropolis-Hastings method and do not discuss other

CC BY

sampling strategies in any detail.

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. We have added a sentence to the end of this section.

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. We have added a new section based on the drivers identified earlier. It is necessarily speculative.

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. This section was added at the request of a previous reviewer who thought that historical context would support the rest of the paper. We have added a sentence to the opening paragraph explaining this.

Specific Comments

P1 Title: Data and Assimilation should be lower case done.

P1 Line 4: . . .for automating part of the process. i.e. the choice of prior distribution is not automated. This is probably line 14. Replaced with "building and using algorithms for ..."

Printer-friendly version

Discussion paper

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. Explanation added.

Page 2 Line 5: awkward wording "demonstration how these many methods are its implementations" changed to "these methods are implementations of that theory."

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. We wanted presentations that were extensive and in closely-related fields. This critique risks being open-ended; there is always one more paper.

P3 Line 7 (Eq. 1): Define xi. ξ plays the role of an integration variable which are customarily not defined.

Section 2.2 capitalize non-Bayesian done.

Page 3 Line 19: Explain what is meant by the "Replicate Earth Paradigm" We believe the explanation would be too long and take us too far afield, we have deleted the reference.

Section 2.2 should come before Section 2.1? This narrows the remainder of the discussion to Bayesian. We think the introduction of events and probabilities should precede the bifurcation between Bayesian and non-Bayesian approaches so have not made this change.

P4 Table 1: This isn't referred to anywhere in the text. Delete extra brackets in descrip- tion of R. Have added sentence pointing out that symbols used throughout are defined in Table 1, also removed extra

CC BY

brackets.

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

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. We have repositioned the captions and explained the values further.

Page 5 Line 26-27: describe what "Equation 2 not well applied" means This referred to poor approximations in the solution but this is not the best place to discuss that so we removed the reference.

Page 7 Line 5: Need to describe what you mean - what are the common misunderstandings We explained these below so have combined the paragraphs and explained the misunderstandings explicitly.

Page 7 Line 10: Reason 1: Why does limiting target variables underestimate the uncertainty? One might remove an important variable. We have explained this.

Reason 2: This is just repeating what was said above and is not a reason. Agree with the repetition, we have reorganised. We have also added that it is not always easy to guess in advance what will be constrained, especially in complex models.

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. We have used an example to illustrate the points below, this is an excellent suggestion. We also deleted the last element in the list since that concerns activities after the assimilation.

Page 7 Line 15: Point 1: Explain what this means. How do you make this decision? The choice is motivated by the science question we are addressing. We have added examples.

Point 3: An example would help. have added an example.

Point 4: What cut-off? We have lengthened the explanation.

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? Two points here. First on k the reviewer is correct, we have replaced "value" with "magnitude". We also adopted the second suggestion.

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. We have introduced this on page 5.

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) We don't think this is the right place to discuss aggregation errors which usually occur with the choice of target variables rather than prior PDF. We have added some more discussion in Section 5.1 including these references.

Page 8 Line 26: Another disadvantage is that it is not fully hierarchical and that hyper- parameters are not integrated out. This is true, we discuss it explicitly later.

Printer-friendly version

Discussion paper

CC BY

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. A good point, we have added a paragraph.

Page 9 Line 11: It is not only resolution that affects 'H' but model structures such as parameterizations. A good point, we had assumed it but now added an explicit comment.

Section 5.4. Need much more detail about how uncertainties are treated in H. We have added a paragraph. We also note this in Section 8 since it is a weakness in current work.

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 Added.

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. comment added.

Page 10 line 32: What is a "well-known" atmospheric inversion? TRANSCOM, now described in detail.

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. It's not clear what the reviewer is referring to here. If it is Wikle and Berliner (2007) then we have essentially done this in the expanded section with the solution for θ sketched. We would also argue that the general formalism we have used already embodies the hierarchical method. One

CC BY

simply augments the target variables with θ, solves for the joint PDF of the augmented target variables and integrates out whatever one doesn't want.

`Page 11 Line 12: A new altered Figure 2 would help show this point` We think the textual description is sufficient and that the space for a new figure is not warranted.

`Page 11 Line 13: Delete 'various of'` done.

`Page 12 Line 19 - Can talk about the role of thinning chains.` Added description.

`Page 12 Line 5- Need references using MCMC` We have cut back on the description of various Monte Carlo methods so believe the depth of referencing is sufficient.

`Page 12 Line 5 - Need to discuss limitations of MCMC such as convergence issues.` Again, we have added a comment that one must watch the sampling properties of whatever method one uses but have avoided technical detail of any particular method.

`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.` should have been "observation operator", now changed.

`Page 13 Line 5 - typo MCMC?` changed.

`Page 13 Line 16 - Needs more explanation - such as?` The sentence was meant to point to the section immediately below, we have made this explicit.

`P13 Line 27: The notation` $G(H(x) - y^o, R)$ `in this section is different to that defined in Table 1 and does not appear Ide et al (1997).` We have corrected the form to match the definition of G in Table 1.

`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 done.

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

P15 Line 20 (Eq. 7): Is the meaning of J described before this? What is J? What does it mean? Explain for a beginner. J is actually defined by that equation, we have now added text ahead explaining this.

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. corrected throughout.

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

P21 Line 10: Delete 'apparently' done.

Reviewer III

We refer to our general responses to the reviewer's overall comments.

General Comments

1) Section 2.2: Since the paper focuses on Bayesian approach, I don't think the paper needs to discuss non-Bayesian method. We don't discuss it in detail but think readers need to be aware of the difference, especially since Bayesian methods are still regarded as more obscure.

2) Section 3: It is not clear to me what is the difference between "x" (Target variables for assimilation) and "z" (model

CC BY

state variable). It is not explained in the paper. Later on, only "x" is used in the cost function. We have added a paragraph in Section 5.1 to clarify this distinction.

3) Section 5: this section has a lot of useful materials. I would suggest discussing how each element is addressed in specific examples. Section 5.1 gave a recipe to decide "target variables". It would be easier to understand if this recipe is discussed within a specific application. Reviewer II also suggested this and we have added an example.

4) Section 6.5: The Kalman filter is discussed in a very general concept here. How Kalman filter has been used in biogeochemistry, and what is the challenge? We have addressed the history in Section 7.

5) Section 7 discussed specific Bayesian methods from computation perspective. Again, I found the description is too general. The applications of these methods in specific problem could be very different. For example, in boundary con- dition estimation, the prior ensembles may not come from the posterior ensemble of previous step, since there is no dynamical model to propagate information forward. As a result, step 5 described in section 7.4 is not applicable. If there is no dynamical model for the target variables one might question whether the KF or any of its derivatives is a good choice for the assimilation algorithm. As we note later, even algorithms that did not seem to be using a dynamical model were often assuming a persistence model.

cc BY

References

Desroziers, G., Berre, L., Chapnik, B., and Poli, P.: Diagnosis of observation, background and analysis-error statistics in observation space, Quarterly Journal of the Royal Meteorological Society, 131, 3385–3396, https://doi.org/10.1256/qj.05.108, http://dx.doi.org/10.1256/qj.05.108, 2005.

Ide, K., Courtier, P., Ghil, M., and Lorenc, A. C.: Unified Notation for Data Assimilation: Operational, Sequential and Variational, 1997.

Jaynes, E. and Bretthorst, G.: Probability Theory: The Logic of Science, Cambridge University Press, http://books.google.com.au/books?id=tTN4HuUNXjgC, 2003.

Manning, A., O'Doherty, S., Jones, A., Simmonds, P., and Derwent, R.: Estimating UK methane and nitrous oxide emissions from 1990 to 2007 using an inversion modeling approach, Journal of Geophysical Research: Atmospheres, 116, 2011.

Michalak, A. M., Hirsch, A., Bruhwiler, L., Gurney, K. R., Peters, W., and Tans, P. P.: Maximum likelihood estimation of covariance parameters for Bayesian atmospheric trace gas surface flux inversions, J. Geophys. Res., 110, D24 107, https://doi.org/10.1029/2005JD005970, 2005.

Tarantola, A.: Inverse problem theory and methods for model parameter estimation, siam, 2005.

Wikle, C. K. and Berliner, L. M.: A Bayesian tutorial for data assimilation, Physica D: Nonlinear Phenomena, 230, 1–16, 2007.

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

cc BY