Biogeosciences Discuss., 10, C5897–C5904, 2013
www.biogeosciences-discuss.net/10/C5897/2013/

Biogeosciences Discussions
Open Access

Interactive comment on "Bayesian calibration of a soil organic carbon model using Δ^{14}C measurements of soil organic carbon and heterotrophic respiration as joint constraints" *by* B. Ahrens et al.

Anonymous Referee #1

Received and published: 18 October 2013

General comments:

Overall, this paper is a nice first pass at a Bayesian approach at fusing SOM pool and soil respiration data with 14C isotopic data on each of these in order to constrain a simple soil biogeochemistry model.

While I would agree that this paper is novel and an important contribution to the literature, relative to other applications of Bayesian stats in ecology there is an important gap between the approach used and and current best practices. In the approach the authors use, there is too much of a separation between the actual data and the statistical

model. The authors are constantly pre-computing factors that are part of their model, and then in the end combining things using an overly simplistic Gaussian likelihood that does not distinguish process error from observation error – in essence they're keeping everything bad about a cost function approach to model optimization, and greatly underutilizing the capacity for the Bayesian framework itself to characterize and propagate uncertainty. Ultimately, something like a state-space / Hidden Markov approach would be a much more robust way of dealing with process error than just a MCMC minimization of errors around a fully deterministic prediction (though such a radical change is probably beyond the scope of this specific paper).

The writing is generally not bad, though the grammar is rough in places.

Specific/technical comments:

Pg 13804

line 11: "allows determining"?

line 14: In the later discussion of this point, you don't give enough credence to the reality that the age of plant carbon can lag behind the atmosphere by years. There are clear cases of plants allocating carbon that's over a decade old (e.g. Carbone et al 2013).

line 20: "joint use all" ?

Line 21: "allowed constraining"?

Pg 13805:

line 18: "better accessibility"?

Pg 13807:

line 3: This isn't a "trade-off", it's a covariance between sink strength and turnover time. As you rightly note, it means that estimates will be biased. It also means that the

BGD

10, C5897–C5904, 2013

estimates that assume steady state will be falsely overprecise.

Pg 13808:

line 21: I'm surprised that you set r to 1 since there is a massive sensitivity or soil respiration to temperature, and a smaller sensitivity to soil moisture, both of which are well documented. It seems odd to assume that decomposition occurs at a constant rate in a boreal forest!!

line 25: First, Where is this term introduced? Later in the paper it appears that this term is multiplicative, not additive, but this is never written out. Second, Why is this term introduced? It seems odd to a priori assume that your estimates for these two terms (but only these two terms) are biased without any discussion from the literature about sources of possible bias that go above and beyond the sampling uncertainty (which would scale with the number of litter traps). From my personal experience in the field, I would not have thought this term necessary and would encourage running a version of the model without this term – I'm guessing that you added this as a solution to an error that you're not telling us about? In general, it should also be noted that bias terms are notoriously difficult to identify statistically without some other independent data source (or a strong prior), as they will frequently end up covarying with other model parameters? Furthermore, these terms should be part of your observation error model, not part of your process model, but your model as written doesn't separate these two (more on that latter). Finally, given the potential identifiability problems, I would strongly recommend that you perform a simulated data experiment to verify that if you simulate data from your model with known parameters that you can re-estimate those parameters with the statistical model.

Page 13809:

Line 18 – line 13 (following page): This single sentence is 16 lines long and includes three numbered equations!! That is way too long.

BGD

10, C5897–C5904, 2013

Page 13810 line 11: Setting these lags as fixes values is completely inappropriate, first because they are estimated from data (and thus has sampling uncertainty) and second because there is genuine biological variability in these processes. These should enter the model as parameters (i.e. as distributions) and the data you are using to estimate these fixed values should either enter explicitly into the statistical model or be used to estimate the informative priors.

Line 15: Yes, but that uncertainty in initializing those pools is real, while spin-up leads to a false overconfidence about the state of these pools. I think it would have been much better to have treated the initial conditions as part of the estimation problem than to rely on an equilibrium assumption, even if you introduce parameters to relax that assumption with a later multiplicative factor.

Pg 13811

line 4: “analog way”

Pg 13812, line 21: colon should be a period.

Page 13817:

line 20: These mixing models should be part of the statistical model (i.e. in the MCMC), not a calculation that is done a priori. In doing so you need to acknowledge the uncertainties in the end points as well as in the observations. An analytical error calculation is too coarse of an approximation for this nonlinear model. There is an extensive literature on Bayesian isotope mixing models that you should be utilizing (and citing) here.

Line 22-Line 4: “uncertainties” is a vague, imprecise term here. I suspect the equations you are using are for the propagation of variances. Also all these equations (e.g. eqn 20) assume zero covariance, which won’t be true for many of your data sets, and imply Gaussian distributions, which might not be appropriate. Furthermore, both here and in later applications, you are providing far too little information about what the calculations

were that you actually did. While I stick by my overall comment that I think that in general you shouldn't be doing these calculations at all (but instead include the data in the statistical model), however if you do perform these calculations you should have a DETAILED supplement that walks through each calculation. Remember that science has to be repeatable!

Page 13818:

line 11 Why would you exclud SOC measurements because they didn't pair with the 14C data – that C is still there and still respiring even if you don't know it's 14C.

Line 13: You can't cite personal communication with yourself. Sue Trumbore is a coauthor)

Line 15: are these other studies for the same site? This statement is ambiguous.

Page 13819

line 11-12: First, this statement needs to be much more precise about what you actually assumed (same order of magnitude is a pretty broad statement). Second, I'm not sure this assumption is true – you should provide some additional justification (e.g. root vs leaf litter in other conifer sites) for this.

Line 24: Why wouldn't this be representative for the year 2007? is the data not from 2007?

Page 13820:

line 2: ok, but what were the standard errors of the individual sampling dates and how were they calculated? Also, if you're measuring the same plots on different dates that's repeated measures data and Eqn 20 doesn't hold due to autocorrelation.

Line 15-17: Again, this should be a distribution and estimated in the model

Line 20-23: How were these measurements combined if they're from different forests

BGD

10, C5897–C5904, 2013

and different years?

Page 13821:

Line 6: Data for different horizons in the same pits are not independent. Eqn 20 doesn't hold.

Line 18: uncertainty on this estimate?

Page 13822:

Line 10: you can't prove convergence of a MCMC

Page 13823:

line 1-2: It doesn't make any sense to me to call Coarse Woody Debris a "bias" on you estimate of foliar litter. These are two different things.

Line 3: First, if the previous lines say that the bias is between 1.2 and 1.7, why would you then construct a prior that has a mean of 1 and a 95% CI from 2/3 to 1 1/3, and thus barely overlaps this range at all??? Second, this seems like a pretty informative prior on a parameter that has no quantitative information behind the construction of the prior.

Line 6: Why would you assume the same bias for root litter, the observation error on that is a completely different process and arguably much less constrained.

Line 12: Given all the thought put into the shapes chosen for the priors, I'm very surprised that you chose a Gaussian likelihood, which doesn't make sense given your data (you're allowing for negative SOM and HR).

Line 14: Define the subscript i. Also, this equation creates the false impression that you can fit each dataset separately in the MCMC and then multiply them together (Eqn 23). Also, somewhere you should be explicit about what parameters you are estimating by MCMC and which are fixed a priori.

BGD

10, C5897–C5904, 2013

Line 15: First, where are the priors on these sigmas?? Even if these were estimated from data, those estimates have uncertainties that will vary strongly with sample size. Second, and more importantly, this likelihood includes zero probability of process error in the model, which can't possibly be true. If you're disallowing the possibility of process error, and fixing the observation errors with zero uncertainty, then the uncertainties in your posteriors will be seriously biased.

Page 13824

line 18: what was the total sample size of the MCMC? What was the autocorrelation? Did you do any visual checks for convergence or just the SRF? Also, how was this computation done? What software or language was used? Is the code and/or data available in some public repository?

Page 12825

line 2: This paragraph breaks in the wrong place. It should break between these two sentences, and the second should be the start of the next paragraph.

Page 12826

line 20: The bias parameters are probably contributing to the nonidentifiability of these parameters

line 24: Good job including this section, too many people don't check/include the co-variances

Page 13827

line 21-22: No, it means that to get the pool size and respiration right you need to keep the total inputs constant and thus you need to have compensating shifts between the two bias parameters on your two inputs

line 29: "stronger constrained"

BGD

10, C5897–C5904, 2013

Page 13828

line 1: “shall be” → is

line 10: as you note later, strong correlations doesn’t mean overparameterization (e.g. a linear regression has very strong parameter correlations) but it may be that the model could be refactored to reduce the correlations somewhat (e.g. how centering reduces correlations in regression). Places that have parameters being added or multiplied together are likely to generate such correlations. Obviously some correlation is unavoidable in a multiple pool model.

line 18: drop “inadvertently” - this structure is clearly deliberate, and in fact is why you get so much power from isotope data

Page 13830

line 19-20: in general, prescribing a stronger prior is a bad solution to reducing uncertainties. Also, for this specific prior, I seriously doubt that knowing more about the site history would really improve this prior – if you told me exactly how a stand was managed down to the last stem I doubt I could tell you much about how much the soil carbon was out of equilibrium. This result (getting the priors back as the posteriors) also suggests that your results will be sensitive to these priors – it would be good to test this by re-running the analysis with different priors. Fo and fy will be convolved with model error as well (you introduce no possibility the real pools could be off the model equilibrium but actually be at equilibrium [i.e. that the model predicts the equilibrium wrong])

Page 13832

line 21-24: This is why this parameter should be treated as a distribution and estimated in the model.

Interactive comment on Biogeosciences Discuss., 10, 13803, 2013.