Reply to comments on Braakhekke et al, Modeling the vertical soil organic matter profile using Bayesian parameter estimation, Biogeosciences Discussions, 2012

General comments

We thank the reviewers for their constructive comments, which often focussed on very detailed and technical aspects of the study. They have contributed to a great improvement of the paper.

Due to an error in the measurement data for Hainich, the calibrations needed to be rerun. For these reruns we followed reviewer #1's suggestion in comment 2.3, and used a different likelihood function, in which the uncertainty of the model residuals is integrated out. This is described in section 2.4.1 of the revised manuscript. The calibrations for both sites were rerun using this new likelihood function. The calibration results for Loobos were virtually identical to those presented previously. For Hainich there were several differences but the tenor of the results stayed the same. The most important differences are: (i) the minimum misfit of the modes. Mode B is now more clearly favourite, while for the mode C the misfit has become comparatively larger. (ii) the modeled advective fluxes for mode B are now more strongly overestimated compared to the measured DOC fluxes.

In 2011 an similar version of this manuscript was published in Biogeosciences discussions (Vol. 8, 7257-7312, 2011). We refer to this document in several places as the 2011 version of the manuscript.

Reply to review 1 by B. Scharnagl

Specific comments

2.1 Both reviewer #1 and #2 commented on the definition of the likelihood in Bayes theorem (Eq. 6). However, their advice is conflicting (also with the advice of reviewer #1 of the 2011 manuscript). Furthermore, review of literature shows that there is in general no consensus on the exact formulation of Bayes' theorem (Mosegaard & Sambridge, 2002, Inverse Problems, 18, R29-R54; van Oijen et al., 2005, Tree Physiology 25, 915-927; Tarantola, 2006, Inverse Problem Theory). We follow the suggestion of reviewer #1 with the modification of writing probability as lower case p, as reviewer #2 suggested, which is customary for probability density functions:

$$p(\boldsymbol{\theta}|\mathbf{O}) = c\, p(\boldsymbol{\theta})\, p(\mathbf{O}|\boldsymbol{\theta}).$$

2.2 The reviewer criticised our choice for a log-normal likelihood function for the Bayesian calibration, as described in section 2.4.1. We must first note that the description of the likelihood function in the manuscript as a log-normal distribution was incorrect. Rather, we used a normal distribution for the log of the measurements (except for the $^{210}\mathrm{Pb}_{\mathrm{ex}}$), which is not the same. Nevertheless, this still implies a right-skewed likelihood function for the model residuals of the untransformed variables. We agree with the reviewer that normally distributed model residuals is the natural starting point for most calibration exercises. However, we have several reasons to expect that a log transformation is appropriate in our case and chose to do so again in the reruns. Our reasons are as follows:

1. All of the observed variables are expected to have right-skewed distributions because they cannot be less than zero and have large variances due to spatial heterogeneity. The measurement error coming from other sources (e.g. the C/N analyser) may be normally distributed, but these errors can be assumed to be negligible compared to the spatial variation. Analysis of measurements from a large campaign at Hainich with 100 cores (Schrumpf et al., 2011 Biogeosciences, 8, 1193-1212) confirms this for the organic carbon stocks and mass fractions, and shows that the distributions become closer to normal when a log transformation is applied.

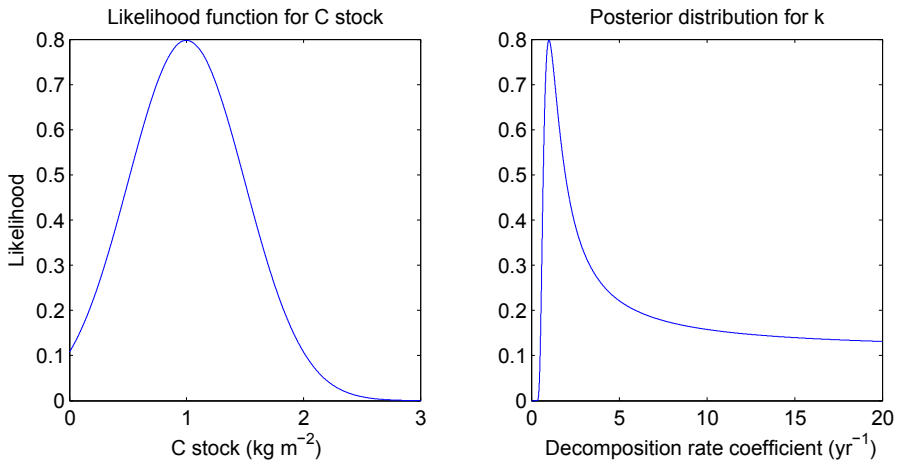

Figure 1: Example of how an improper posterior distribution can result from using a normal likelihood function for the modeled C stock using a simple soil carbon model. The model estimates the steady state C stock as $C = {}^{I}/_{k}$, where I is the litter input and k the decomposition rate. The left graph shows the likelihood function used for the steady state C stock ($C \sim N(1, 0.5)$), and the right graph the resulting posterior distribution for k using $I = 1$ and a uniform prior on the positive reals.

2. In the new calibrations we treated the soil carbon mass fractions and effective decomposition rate coefficients in the mineral soil as one data stream with the same variance. Since these mass fractions show clear heteroscedasticity, taking the log is required to obtain similar variance for all depths.
3. There is a theoretical reason to apply a log transformation, which relates to our use of a first-order decomposition model. In such models the decomposition rate coefficients are in general inversely proportional to the modeled soil carbon quantity. For a simple one-pool model, a carbon stock of zero can only be produced for an infinite decomposition rate coefficient, if litter input is larger than zero. This means that using a likelihood function that does not approach zero at zero carbon stock yields an improper posterior distribution for the decomposition rate, that does not asymptotically approach zero at infinity, if an uninformative prior is used (see Fig. 1). SOMPROF is obviously much more complex than this example, but it does lend theoretical support to our choice for applying the log transformation because the resulting likelihood function for the untransformed variables approaches zero at zero carbon stock.

Although we agree with the reviewer that posterior checking of the assumptions concerning the distribution of the model residuals is prudent, this is difficult in our case. Since the residuals for the different data streams are expected to be differently distributed (see also next point, and section 2.4.1 of the revised manuscript), checking the distribution would have to be performed separately for each data stream. However, per data stream there are very few data points available. Checking the distribution of the residuals for all replicate measurements would be possible, but this distribution would be determined by the distribution of the replicates (which we already know), not the modelling error.

Section 2.4.1 has been revised and we extended section 2.3.1 to explain our reasons for the log transformation.

2.3 The reviewer rightfully pointed out that σ used in the likelihood function (section 2.4.1) refers to the variance of the model residuals, not the observations. By using the standard deviation of the observations for σ we pretend that our model is perfect, which is usually not true. For the reruns we followed the reviewers suggestion of using a modified likelihood function where the

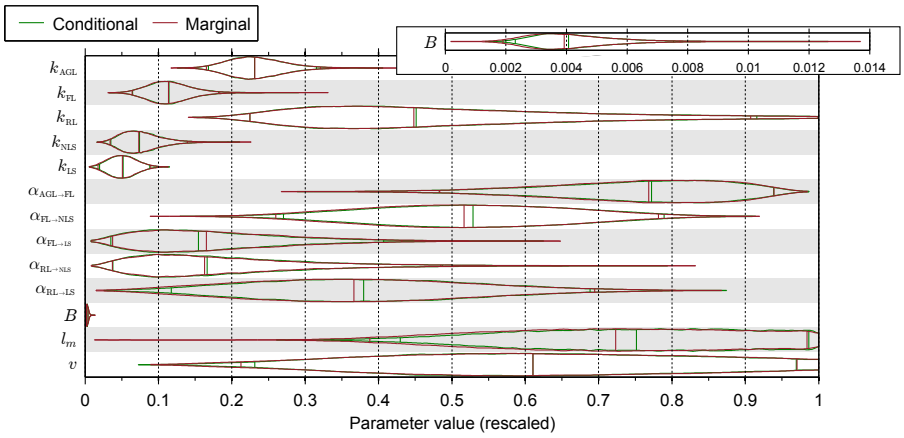

Figure 2: Comparison of posterior distributions for Loobos, calibration setup 3. The green line shows the conditional distributions derived with the previous likelihood function for given values of σ equal to the standard deviation of the measurements. The red line shows the marginal distribution over all values of σ.

unknown sigma is integrated out based on the uninformative Jeffreys prior. The new results for Loobos are virtually identical to those of the previous runs (Fig. 2). (For Hainich the new results cannot be compared with the previous runs because different observations were used; see general comments above.) This shows that the modelling error has limited influence on the distribution of the model residuals in our case. Several reasons may be put forward for this.

First, as explained in section 2.4.1 of the manuscript, replicate measurements from different soil cores were used individually to calculate the sum of squares. However, for most data streams these were all compared with only one model prediction, hence the spread of the model residuals is fully determined by the spread of the replicate measurements, which was already accounted for in our previous approach. The exception to this are the profile variables (mineral soil C and $^{210}Pb_{ex}$ fractions), for which measurements at several depth levels were included. But the fit to this data is generally quite good (see Fig. 5 in the manuscript), i.e. modelling error is relatively small.

Second, the number of estimated parameters is relatively large compared to the number of observations, which partially explains the poor constraint on some the model parameters. This is likely to cause over-fitting, which means the modelling error is small. On the other hand, the spatial heterogeneity of soil carbon stocks and fractions is notoriously large, which is expressed by the spread of the replicate measurements. Therefore, the spatial uncertainty is likely to dominate distribution of the residuals.

2.4 The reviewer commented on the use of the correction factor which accounts for the effects of generating samples in log- or logit-transformed space (Appendix A2). The introduction of the Jacobian of transformation follows from the rules for change of variables for probability density functions[1]. It can easily be demonstrated by generating a large random sample of numbers x from a uniform distribution on any interval, and plotting a histogram of $\exp(x)$, which follows $1/x$. We improved the explanation in Appendix A2 and we included a reference to Box and Cox (1964).

The description in the previous version of the manuscript was misleading because it implied we did not apply the correction factor for calibration setup 3, which is not true. For *all* calibrations

[1] see `http://en.wikipedia.org/wiki/Probability_density_function#Dependent_variables_and_change_of_variables`

we sampled in transformed space and we applied the Jacobian correction. However, in calibration 3 the correction factor cancels out with factors in the log- and logit normal probability density functions which were used as priors. For example, the log-normal distribution is defined as:

$$f(\theta) \propto \frac{1}{\theta} \exp\left(-\frac{(\ln(\theta)-\mu)^2}{2\sigma^2}\right),$$

and the correction factor in case of a log transformation is equal to θ. Thus we simply used the corresponding normal distribution for the log of the parameter, without the $\frac{1}{\theta}$ factor. Essentially, we modified the prior such that it already included the correction factor. Since this is trivial to the description, we shortened it and simply stated we used the correction factor for all calibrations.

We did not refer to the correction factor as the "Hastings factor". The confusion may have been caused by the fact that we previously (in the 2011 version of the manuscript) termed it such and because we denoted it by "J", which is often used to indicate the Hastings factor. The Jacobian correction *could*, however, be construed as such since the Hastings factor is introduced to account for an asymmetric proposal distribution. A normal proposal distribution in e.g. log-space, leads to a log-normal proposal distribution in untransformed space, which is asymmetric. Nevertheless, to avoid further confusion we changed the symbol to λ.

2.5 The reviewer commented on the misfit quantity, used to compare the modes, as described in section 3.2 and table 3. The caption of table 3 and the description in the text were indeed incorrect. The misfit function is defined as the *negative* log density of the unnormalized posterior, hence a minus sign was missing in the formula. This was corrected and the explanation in section 3.2 was extended (see also reply to point 5 of reviewer 2). For normal likelihood functions and priors, the misfit can be interpreted as a weighted sum of squared residuals against observations and prior parameter estimates (Tarantola, 2006, Inverse Problem Theory, p.36). In our case this interpretation does not hold, but it is still a valid statistic to compare the modes.

2.6 The reviewer asked about the usefulness of the effective decomposition rate observations, in the context of this research. It is difficult to assess the information content of the effective decomposition rate measurements without performing a separate calibration study without these observations, which is outside the scope of this study. However, we believe that this information is important for constraining the decomposition rates of the organic matter pools. As can be seen in Figs. 4 and 7 (and Fig. 7 in the supplementary material), the marginal posterior distribution of the decomposition rate coefficient for the dominant pool LS is much more tightly constrained for Hainich (mode B) than for Loobos. It is likely that this is largely due to the presence of the effective decomposition rate measurements for Hainich, because in the deep soil, this pool almost completely determines the model prediction for this variable. Furthermore, for Loobos we see strong correlations between k_{LS}, $\alpha_{\mathrm{RL}\rightarrow\mathrm{LS}}$ and v (supplementary Fig. 6), which indicate that the formation, loss and transport of this pool cannot be constrained individually. For Hainich these correlations are less strong, which is presumably also caused by the effective decomposition rate observations. It also suggests that through the correlation structure these observations can also help to constrain other parameters.

It is true that these measurements involve considerable efforts. However, we believe that when studying vertically explicit soil carbon cycling these measurements are worth the effort. We added additional discussion related to this question to section 4.3.

Technical corrections

Unless otherwise mentioned all suggestions were followed.

p.11270 l.16 The reviewer asked about the factor used to artificially inflate the variance of the posterior as described in appendix A3 and the caption of Supplement Fig 2. (now Fig. 3). The paragraph

in appendix A3 describes the initial "exploratory" runs intended to search for multiple modes. The run performed for Supplement Fig. 2. (now Fig. 3) was a separate run meant to demonstrate that the multi-modality is not an artifact of the sampling, in view of the comments from the reviewers of the 2011 version of this paper. For this run we did not want to inflate the variance of the posterior more than necessary to sample over all modes.

Reply to review 2

Specific comments

1. The reviewer commented on the definition of Bayes theorem in Eq. (6). The advice of the reviewers #1 and #2 on the formulation of the likelihood conflicts. We chose the following compromise: $p(\boldsymbol{\theta}|\mathbf{O}) = c\, p(\boldsymbol{\theta})\, p(\mathbf{O}|\boldsymbol{\theta})$. See also discussion of point 2.1 in the reply to review #1, above.

2. The reviewer commented on the statement that an analytical solution of the posterior does not exist for our study in section 2.4 (now in section 2.4.3). We followed the reviewers suggestion and moved all discussions about the MCMC algorithm to section 2.4.3, which is now a general section on Monte Carlo simulations, both inverse and forward. Furthermore, we added a sentence to explain more clearly why an analytical expression of the posterior is not available.

3. The reviewer commented on the choice of a log-normal model for the likelihood function described in section 2.4.1. In the reruns a different likelihood function was used. This is discussed in the reply to point 2.3 of reviewer #1 and in section 2.4.1 of the revised manuscript. Note that for most data streams the measurements were log-transformed for the calibrations. Our reasons for this are discussed in the reply to point 2.2 of reviewer #1.

   Concerning the formulation of the likelihood function, in the revised manuscript we use the proportionality symbol to express the likelihood function, which means the normalizing constant can be omitted (Gelman et al., 2004, Bayesian data analysis, Ch. 1).

4. The reviewer pointed out that the several parameters for Loobos are poorly constrained by the data while this is only noted for several of them, in section 3.1. This is correct. The text has been modified.

5. The reviewer commented on the definition of the misfit, as described in section 3.2 and Table 3. The misfit was indeed incorrectly defined, both in the text and in the caption of Table 3. This has been corrected, and we followed the reviewers suggestion of adding an exact definition of this quantity. See also reply to point 2.5 of reviewer #1.

6. The reviewer pointed out that Table 2 needs more explanation on the prior distributions and upper bounds of the parameters. Additional information has been added.

7. The reviewer pointed out the conflict between Fig 3. and Table 2. related to the upper bound of parameter v. Fig 3. has been corrected.

8. The reviewer commented on the used of the correction factor in the calculation of the posterior density in appendix A2. This section has been modified. See also reply to point 2.4 of reviewer #1.

Technical corrections

All suggestions were followed.

Reply to review 3

Specific comments

1. The reviewer asked why an average annual cycle was used to force the model (section 2.1.1), rather than using the full time series of data that were available.

   We must first note that since we did not use time series measurements in the calibration, we had no constraints on the dynamic behavior of the model over time, only of the model state at the end of the simulation. It is true that inter-annual variability of the forcing may affect organic carbon stocks over long time scales. The reviewer is correct in stating that this could have been somewhat better accounted for, had we used the full time series of measurements available to us. Our reasons for not doing so were as follows:

   - The time series of forcing data (litter fall and soil temperature and moisture) available to us spanned different time periods and had many gaps. Using the full time series would be difficult and require harmonizing the of the different variables and performing gap filling.
   - Our main reason for considering the seasonal cycle was to account for variability in the forcing data. The seasonal (intra-annual) variability typically represents the largest part of this variation. Over the complete length of the simulation the inter-variability is presumably much larger (particularly for Hainich), but the time series of the forcing measurements provides little information about this. The uncertainty that results from using the same annual cycle over the whole simulation is discussed in section 4.6.
   - Test simulations where we used a constant average annual values for the forcing gave very similar results compared to runs with an average annual cycle. This shows that even the seasonal climate fluctuations do not have large effects on the model results.

   Section 2.1 has been extended to more clearly explain our choice for an average forcing cycle of one year.

2. The reviewer pointed out that the vertical distribution function of root litter input for Loobos discussed in section 2.2.1 should be better explained. The root litter distribution used for Loobos was not derived by fitting to a data set, since no suitable measurements were available for this site. Rather, the function was based on information from Janssens et al. (2002, Forest Ecology and Management, 168: 231-240) for the Brasschaat site in Belgium which is very similar to Loobos. Furthermore, personal communication from I. Janssens and J. Elbers, helped to establish the vertical distribution function. In the study for the 2011 version of this paper all root input was distributed according to the function used for understorey which is very shallow. The results of this study suggested that liquid phase transport of organic matter was mainly responsible for the OM in the deep soil but we realized this could have been caused by the fact that we ignored deep litter input by the pine roots. Hence, we modified the distribution function to account for this. Despite the different distribution, very similar results were obtained, which strengthens our conclusions concerning the role of liquid phase transport at Loobos. We modified section 2.2.1 to better explain background of this function.

Technical corrections

Unless otherwise mentioned all suggestions were followed.

P 11247 L13-15 The reviewer asked about the lower boundary conditions for the diffusion-advection model. These are described at the end of the paragraph.

P 11284, Fig 1 The reviewer pointed out that diffusion can also lead to upward transport, which is not shown in Fig 1. This is correct. Although net transport will in general be downward, diffusion in principle also lead to upward transport. We added additional arrows to Figure 1. to indicate this.

Reply to short comment by O. Dilly

The commenter asked for better representation of the soil profile measurements at both sites. We added a table with all the information used in the calibration to the supplementary material (Supplement Tab. 1), as well as a graph depicting the measured soil carbon stocks and mass fractions (Supplement Fig. 2). For additional information about the Hainich site we refer to Schrumpf et al. (2011, Biogeosciences, 8, 1193-1212).