# Reply to reviews of Braakhekke et al, Modeling the vertical soil organic matter profile using $^{210}Pb_{ex}$ measurements and Bayesian inversion, Biogeosciences Discussions, 2011

## Relationship between current and 2011 manuscript

A very similar manuscript has been published previously in Biogeosciences Discussions (Vol. 8, 7257-7312, 2011[1]) and was reviewed by two referees. This manuscript essentially concerns the same study as described here, but a revised manuscript was not submitted so far because an error was discovered in the code which required all calibration runs to be redone. The current manuscript presents the results of the corrected runs, with several additional improvements. The most important differences with respect to the 2011 manuscript are:

- An error was removed from the model code.
- A better Metropolis algorithm was used for the calibration (DREAM(ZS)).
- The root litter input rate and its vertical distribution for Loobos changed to more realistic values.
- The results have changed to some extent, which required the results and discussion sections to be largely rewritten.
- With additional analysis we calculated the amount of material derived from roots and above ground litter. Furthermore, the relative contribution of different processes for SOM profile was estimated.

Because of the similarities between the two manuscripts, the reviewers' comments to the 2011 manuscript partly apply to the current manuscript as well, hence, we considered their comments, when applicable. Below a reply to many of the comments of the reviewers is given. If, for a specific comment, no reply is given, we simply followed the reviewers suggestion, or it was no longer applicable in the current manuscript.

## Reply to comments regarding multimodality of the posterior

Both reviewers pointed out that the multimodality of the posterior distributions has technically not been proved, since the Markov chains were not able to sample multiple modes simultaneously, i.e. the algorithm had not converged globally (specific comments 6 and 8 in review 1; specific comment 28 in review 2). It should be noted that for the calibrations for the current manuscript multiple posterior modes were found only for Hainich. For Loobos the posterior appeared to be unimodal for all calibration setups.

It is correct that the individual chains have generally not sampled more than one of the modes (for both the previous and the current study). Since we treated the modes individually, a multi-modal sample was not required for our analysis. In fact, performing separate samplings of the individual modes was preferable. We believe that the fact that a multitude of chains converged on each of the modes demonstrates that they are not an artifact of the algorithm. To further confirm the multimodality we attempted to perform additional MCMC runs aimed at global convergence. This proved to be very difficult even with the DREAM(ZS) algorithm, which is designed for sampling multimodal distributions. Since, for technical reasons, sampling is more difficult with uniform priors, we managed to obtain global convergence only for calibration setup 3, (including $^{210}Pb_{ex}$ observations and strong priors). Furthermore, several additional modifications were made to speed-up convergence: The calibration was limited to four parameters that differ strongly between the modes, while the other nine (for which the distributions overlap) were fixed at the average over the modes. Furthermore, a cost

[1] http://www.biogeosciences-discuss.net/8/7257/2011/

reduction factor was applied, flattening and widening the distributions. As can be seen in supplemental figure 2, the three modes are quite apparent and Markov chains are able to jump between them. Since the prior distribution is uni-modal, the multi-modality must be caused by the model-data likelihood function. Furthermore, the results show that both the prior distributions and the $^{210}Pb_{ex}$ data did not have strong effects on the posterior. Thus, in spite of the limitations, these results strongly support the multi-modality for all three calibration setups for Hainich.

## Reply to comments regarding (root) litter input

Both reviewers commented on the fact that the parameters related to litter input, particularly the root litter input rate and its vertical distribution, were held fixed and not included in the calibration (specific comment 10 in review 1; specific comment 1 and general comments in review 2).

We agree that the measured root litter parameters represent a source of uncertainty which should be considered when interpreting the results. However, there are many more uncertainties related to other aspects of the model, such as the history of the sites, the forcing data, and the general validity of the model structure. There are limits to the number of uncertainties that can be considered in the calibration. The posterior distributions are already quite complex and therefore very difficult to approximate. Even with the superior algorithm DREAM(ZS) sampling proved difficult and often 500 000 iterations per chain were needed (taking weeks of calculations on over 30 parallel CPUs) in order obtain a good approximation. In short: already with the current setup we have approached the limits of what is feasible with the amount of computing power available to us. We expect that adding more parameters to the calibration would make convergence in a reasonable amount of time impossible.

Considering this limitation, the obvious question is: which uncertainties should be considered in the calibration? We decided to focus on model parameters that are inherently unmeasurable, i.e. those describing decomposition and organic matter transport, and fix the litter input parameters, for which we had reasonable estimates. Reviewer 2 noted that "litter input constants are more uncertain than decomposition constants". It is true that overall decomposition rates coefficients can be estimated, for example, from lab incubations. In fact, such measurements were performed for Hainich and used as observations in the calibration. However, the decomposition rates of the individual pools can in general not be measured because the pools themselves are model abstractions that do not exist in reality.

For the current manuscript we were better able to quantify the effects of the different processes on the SOM profile. This showed that root litter input is in fact very important for soil carbon, particularly at Hainich. We believe that these results are robust against reasonable changes in the litter input parameters. In preliminary tests with the model it was found that a moderate ($\sim$25%) increase of the root litter input rate and changes to its vertical distribution, caused a slight shift in the posterior distribution but it did not change the tenor of the results (i.e. which organic matter pool was the most abundant, and the possible existence of multiple modes).

Hence, we decided not to do reruns that include the litter input parameters, as reviewer 2 requested, even though we agree that this would be favorable. The uncertainties associated with the fixed parameters (and from other sources) are discussed in section 4.6 (page 11266, lines 11-26). Despite them we believe our results are still interesting, and contribute to the understanding of SOM profile formation.

## Reply to review 1 by M. van Ooien

### Specific comments

1. Finding correct initial values is a problem for all soil organic carbon models, and is the subject of much research. In general one has two choices: (i) starting with zero organic carbon and simulating the complete buildup of the profile, and (ii) starting from a non-zero initial state and running the model only for the period of interest Since formation of the vertical SOM profile

involves very slow processes, we are interested in the complete period of soil formation, hence we chose the first option.

For the Loobos site, the simulation covered the period since forest plantation, approximately 95yr before the sampling date. Since the area was covered by shifting sands (indeed similar to a desert soil) until the forest plantation, the assumption of zero initial organic carbon is quite reasonable. The effect of the initial development of vegetation on litter production was accounted for by multiplying the average litter input with linear function increasing from 0 to 1 in 60yr.

For the Hainich site soil profile development started much longer ago, hence we assumed the soil to be near equilibrium. We simulated using a constant annual cycle of forcing data, for 1000yr, which (as demonstrated by test runs) is sufficient to bring the model close to steady state for typical parameter values. We admit that the assumption of equilibrium and semi-constant past forcing adds uncertainty to the predictions. However, the Hainich site has a relatively homogenous site history, having been covered with forest for at least 250yr and has been unmanaged for the last 60−70yr. Several changes were made to sections 2.1 (page 11245, lines 6-11) and 2.2.2 (page 11251, lines 13-19) to more clearly emphasize these points. Furthermore, the site history, together with other uncertainties is discussed in section 4.6 (page 11266, lines 11-26).

2. The only process in the model that was influenced by temperature and moisture is decomposition. Since the response functions that correct decomposition ($f(T)$ and $g(W)$ in eq (1), sect. 2.1.1) for these variables are quite non-linear, smoothing can indeed lead to an aggregation bias. Therefore, the response factors were calculated based on unsmoothed daily measurements of these variables. These response factors were subsequently averaged to monthly values and from several years of data an average annual cycle of monthly values was derived, which was input into the model. Section 2.1.1 (page 11246, lines 2-6) has been updated to include this information.

   Since our aim is to quantify the average processes involved in SOM profile formation over long time scales, the transport rates were held constant during the simulations and were not influenced by temperature and moisture. Hence the choice for smoothed forcing data would not have affected these processes.

3. A reference has been added (page 11253, line 13). Piecewise Hermitian interpolation is similar to spline interpolation, but preserves monotonicity and has less overshoots and oscillation for noisy data.

5. The transformation factors split the decomposition flow into loss as $CO_2$ and fluxes to other pools (Sect. 2.2.1). If the transformation factors for a single pool sum to more than 100%, it would mean that net organic matter would be created. The term “transformation factor” has been changed to “transformation fraction” to emphasize this.

6. See section “Reply to comments regarding multimodality of the posterior” above.

7. In the new algorithm the Metropolis rule was not split, hence this section was removed.

8. For the calibrations presented in the 2011 paper, a simple non-adaptive random walk Metropolis algorithm was used, which is poorly suited for sampling multimodal distributions. This, together with the “isolation” of the different modes in parameter space, made it infeasible for a single chain to sample across multiple modes within a reasonable number of iterations, thus global convergence was not possible.

For the new calibrations we were able to obtain global convergence under certain conditions using the DREAM(ZS) algorithm (see "Reply to comments regarding multimodality of the posterior" above), which we believe, confirms the multimodality of the posterior for the Hainich calibrations. However, the marginal and bivariate distributions, as well as the results of the forward simulations, in the current manuscript are still derived from unimodal samples, to get a better approximation of the modes.

We also increased the number of iterations to a minimum of 500 000 for calibration setups 1 and 2, and 200 000 for calibration setup 3, and used a more stringent maximum of 1.01 for the Gelman-Rubin index. This is discussed in the appendix A of the manuscript.

9. For the runs with strong priors (calibration setups 2 and 3), only informative prior distributions were used for the decomposition rates and transformation fractions. For the transport parameters, the same distributions were used as with the runs with weak priors: a uniform distribution within the bounded region. This is discussed in section 2.4.2, (page 11256, lines 10-12), and illustrated in figure 3e.

10. See "Reply to comments regarding (root) litter input", above.

### Technical corrections

1. All indicated errors were corrected, except for those on page 7269, line 6; page 7276, line 11 (2011 manuscript), since it is not clear to us what errors on these lines should be.

## Reply to review 2 by A. de Bruijn

### General comments

- As the reviewer suggested, the answers to the research questions are now clearly discussed in conclusions section (page 11267, lines 18-24).
- The number of figures has been reduced from 14 to 9. A number of figures has been moved to the supplemental material. We did not follow the reviewer's exact suggestions for merging figures, since there are a number of new graphs, but we believe new arrangement is more convenient.
- We chose to still discuss the calibration setups 1 and 2, since part of our aim is to demonstrate the value of $^{210}Pb_{ex}$ data and prior knowledge for constraining the parameters. However, we removed the figures depicting the results for the forward simulations for these setups (the posterior distributions are still shown), and also discuss these results less extensively.

### Specific comments

1. See discussion in section "Reply to comments regarding (root) litter input" above.

2. Tables 1 and 2, deal with distinctly different model inputs, and have as such quite different columns. Merging the tables would reduce the readability, hence we prefer to keep them separate.

5. It is true that CENTURY version 5 estimates also the soil carbon contents as a function of depth. However, soil carbon dynamics are only simulated for the top 20 cm—an exponential function is used to extrapolate simulated soil carbon contents to deeper soil layers[2]. Hence, the processes and properties of the top 20 cm are assumed to be representative for the whole profile, which

[2] See http://www.nrel.colostate.edu/projects/century5/documents/ExponentialCDistribution.pdf

is similar to what we meant. We've modified this sentence (page 11241, lines 15-19) to better reflect our point.

The reference to Schimel et al., 1994 also concerned the CENTURY model. It was replaced with a reference to a publication concerning the Yasso07 model, which also does not consider the vertical SOM profile.

6. This is correct; the performance SOMPROF had until this study not yet been evaluated for multiple sites and this was part of the aim of this study. We expanded the introduction (page 11243 lines 27-29 and page 11244 lines 14-15), to discuss this. Also, the contribution of the current work to evaluate SOMPROF is discussed in section 4.6 (page 11244, line 27 - page 11244, line 9).

7. We altered the sentence to something similar to what the reviewer suggested (page 11243, lines 23-25). However, we feel it is important to mention models here, since the point is that equifinality hampers parameter estimation.

8. Although it is correct that we did not use $^{13}C$ and/or $^{14}C$ here, almost all past studies on calibrating SOM profile models have used carbon isotopes. Hence, this concerns an important part of the history of this research and deserves attention. Hence, we prefer to leave this paragraph unaltered (page 11243, line 26 - page 11244, 3).

9. We agree that the importance of the distribution of organic matter over particulate and potentially mobile fractions should be discussed to give a background to question ii). Rather than what the reviewer suggested we expanded an earlier paragraph (page 11243, lines 10-11) to clarify this.

13. See reply to specific comment 1. of reviewer 1.

15. The advective transport in SOMPROF represents movement of dissolved organic matter (DOM) with infiltrating water. Accurately simulating DOM dynamics requires simulation of ad- and desorption of organic matter to the solid phase and short timescale water fluxes, which is infeasible and outside of the scope of SOMPROF. We are only interested in the effects of DOM leaching on the SOM profile development over long time scales. We therefore define a pool which is potentially mobile, i.e. leachable slow (LS) organic matter. This pool represents both the organic matter that is in solution and that which adsorbed to the solid phase (but can enter the solution through desorption), and is thus transported by both liquid phase transport and bioturbation. The advection rate represents the effective movement of this pool: the average water infiltration rate, corrected for the fraction that is not in solution. The difference with the non-leachable slow (NLS) organic matter pool in the model is that the material that comprises the LS pool is potentially subject to leaching, whereas the NLS pool cannot be leached. The rationale behind this approach is further described in Braakhekke et al. (2011).

    The line the reviewer referred to was added to clarify that LS does not represent only DOM, but also (and mainly) the adsorbed fraction. We do not know exactly how much of this pool is fact dissolved and how much adsorbed, but it is reasonable to assume that in terms of mass the dissolved part is negligible compared to the adsorbed part. Therefore, we compared the modeled total organic carbon stocks and fractions (sum over all pools) to measurements of solid organic matter, and did not include measurements DOM.

21. The measurements from Smit et al. (2001) were presumably not taken at the exact same location as the soil samples and the other measurements (the coordinates mentioned in the paper were not precise enough to exactly determine this). However, they were done in the same forest, at presumably not more than 1-2km distance. The uncertainty of the root liter input measurements was not explicitly listed in Smit et al. (2001). The standard error of the total above and below ground litter input from canopy and understorey was approximately 18%. Accounting for a somewhat higher uncertainty of root litter input measurements, our best guess of the standard error of the root litter input would be between 20 and 30%.

    Our reasons for not including this uncertainty in the calibration are discussion in the section "Reply to comments regarding (root) litter input", above.

22. Organic carbon measurements were always compared to *total* modeled organic carbon, i.e. the sum of all pools, hence the simulated amounts of the NLS and LS pools were not evaluated individually. Since LS represents mainly material associated with the mineral phase, this pool likely correspond roughly to organic carbon in the heavy fraction, as determined by density fractionation. The other pools in the mineral soil (FL, RL and NLS) would then correspond to the light fraction. Such a comparison was made for Hainich, for which these measurements were performed (Section 4.2, page 11261 lines 19-26). However, it is unsure if the LS pool can directly be compared to the heavy fraction, hence we did not include these observations in the calibration.

    Additional description of the model variables that were compared to measurements has been added to sections 2.1.1 (page 11246, lines 12-17) and 2.1.3 (page 11249, lines 3-5).

25. Since, in addition to negative values, decomposition rate coefficients of zero are also not allowed, the probability must go to zero, when approaching $0\,\mathrm{yr}^{-1}$. Hence a log-normal distribution is justified. As small change was made to the sentence starting on page 11255, line 20-21) to emphasize this.

26. The phrase "maximum likelihood" was replaced with "mode" to avoid confusion with the phrase "likelihood function". The prior distributions were not specifically constructed to have the modes at specific points, but as expressions of our prior expectations. We mention the modes of the priors simply as an indication of the central tendency of the distributions.

28. See section "Reply to comments regarding multimodality of the posterior" above.

31. For the new calibrations $^{210}\mathrm{Pb}_{\mathrm{ex}}$ did not lead to significant additional constraint, because the $^{210}\mathrm{Pb}_{\mathrm{ex}}$ measurements were already well reproduced by the model without including them in the calibration. While, for the sites in this study, this indeed makes this tracer less valuable than we had hoped, it does suggest that it could be useful for other situations where the organic carbon measurements alone cannot constrain the processes. This is discussed in section 4.5 (page 11264, lines 17-24 and page 11265 lines 19-23).

# References

Braakhekke, M. C., Beer, C., Hoosbeek, M. R., Reichstein, M., Kruijt, B., Schrumpf, M., and Kabat, P.: SOMPROF: A vertically explicit soil organic matter model, Ecol. Model., 222, 1712–1730, 2011.