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

Our final Author Comments are generally set in blue, while the original comments by the referee are set in *black italics*.

Reply to Referee #1

General comments

We would like to thank Referee #1 for his constructive comments and hints. Below we replied individually to the general and specific comments in more detail.

Besides other issues, Referee #1 criticized that we treated the lag-times between the atmospheric ^{14}C record and the ^{14}C of litter input as fixed parameters based on measurements at the different sites without allowing for uncertainty of these parameters.

In a revision we would include these lag-times as parameters that are calibrated within our Bayesian calibration framework. We will use log-normal priors for the two lag-time parameters $tlag_L$ and $tlag_R$ using the measured lag-times as modes for the prior. The 99th-percentile for these priors will be $tlag_L$ and $tlag_R$ + 2 years.

Additionally, Referee #1 criticized that the width of priors for the parameters that describe a possible bias in the litter input are not in accordance with the literature values we reported. Consequently, we will choose a broader, log-normal prior for $bias_{iL}$ with its 99th-percentile at 2. We will switch to a log-normal distribution for $bias_{iL}$ because the truncated normal distribution we used before, would have assigned too much probability to very small amounts of litterfall with a 99th-percentile of 2. Furthermore, we now choose a broader prior for the root litter input bias, $bias_{iR}$, compared to the bias for aboveground litter, $bias_{iR}$, with its 99th-percentile at 3. The distribution for $bias_{iR}$ will also be changed to log-normal.

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.

We assume that Referee #1 refers to the error propagation of soil carbon stock calculation and respiration partitioning, which could be integrated into the Bayesian framework. However, our study hinges on published mean values and their respective standard errors; we did not use all the original replicates of measured values which could allow for consistent calculation of

uncertainties within our Bayesian framework. Furthermore, we would argue that the "precomputing" of the respiration partitioning and of the respective uncertainty outside of the Bayesian framework is appropriate because we are actually neither calculating root respiration nor its ^{14}C signature within the model.

The Gaussian likelihood function we use here has been used in numerous studies by several authors for several years (van Oijen et al., 2013;Rahn et al., 2012;Santaren et al., 2007;Reinds et al., 2008;Yeluripati et al., 2009;van Oijen et al., 2005). We are aware of works of Jasper Vrugt, and discussions about model and input uncertainty and different forms of cost functions (Schoups and Vrugt, 2010;Kavetski et al., 2006), but we chose the Gaussian formulation because given the scarcity of our data it might be difficult to estimate process error from the data at hand. We would clarify that we assume that the observed variability in our measurements (mean + SDOM) dominates over process error.

We would be glad if Referee #1 could provide us with additional references where – in his opinion – current best practices in Bayesian calibration are showcased and compared to the approach we took here.

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).

Again, we would be happy to receive information about reference publications that use a state-space / Hidden Markov approach so that we can at least replace our Bayesian calibration approach with newer approaches in publications to come.

If we tried to adopt a Hidden Markov approach, the ICBM model would be the deterministic process underlying the observations. The observations in our case would not be truly "hidden" as they can be directly related to modelled states variables in ICBM. However, we would assume that we can only observe SOC, HR and the respective ^{14}C signatures with an error, which we put into a random measurement model, , with systematic and random errors, for example a normal distribution: $\varepsilon = N(\mu_{systematic}, \sigma^2_{random})$. Probably similarly, one could also treat the driving variables (litterfall) as a stochastic process with an error distribution, e.g. $\varepsilon = N(\mu_{systematic}, \sigma^2_{random})$. Please note, that we implemented something along the lines of this approach (at least for systematic errors) with the two bias factors for litterfall.

However, we are wondering how these additional parameters $\mu_{systematic}$ and σ^2_{random} would fit in our Bayesian calibration framework. Would we treat these parameters as nuisance parameters *sensu Gelman et al. (2004)*?

Overall, we doubt that given the inherent scarcity of our data (radiocarbon data), one could reasonably distinguish between systematic and random errors. Therefore, we chose to fix the error variance to the observed variability in our measurements (mean + SDOM) which is probably dominating over the process error.

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

We will try to rephrase some of the wording and check the grammar of the complete paper again.

Specific/technical comments:

Pg 13804
line 11: "allows determining"?

Will be corrected.

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).

We are aware of that. We would like to point out that the lag-times we had used are actually based on ^{14}C measurements of aboveground and belowground litter. Hence, the use of stored carbon to construct e.g. new roots is already factored into these lag-times. With the lag-time parameters we are only looking at the ^{14}C signature of the litter input, regardless if lag-times of 8 years for the root litter originated from the use of stored carbon to grow new biomass or from the length of the turnover time of roots. We thought that we clearly stated that in the respective sections in the material and methods sections, but we will try to emphasize this more (also when we describe the new priors for $tlag_L$ and $tlag_R$).

line 20: "joint use all" ?

Will be corrected.

Line 21: "allowed constraining"?

Will be corrected.

Pg 13805:
line 18: "better accessibility"?

Will be rephrased.

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 estimates that assume steady state will be falsely overprecise.

Will be rephrased.

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!!

We would like to point out that we are using yearly heterotrophic respiration as an observational constraint for the ICBM model which we also run at a yearly resolution. The ICBM model is essentially a simple 2-pool model, but Andrén and Kätterer (1997) used the parameter r to describe climatic and edaphic on the decomposition rates k_Y and k_O. We state in the

manuscript that we assume that the between-site variability of climate variables can be lumped into the other model parameters.

line 25: *First, Where is this term introduced?*

We will rephrase this sentence. What we wanted to say is that we additionally implemented these two parameters into the modelling setup (Fig. 1).

Later in the paper it appears that this term is multiplicative, not additive, but this is never written out.

It is actually written out in the sentence you are referring to "Additionally, we introduced the parameters $bias_{iL}$ and $bias_{iR}$ which should account for a potential bias in litterfall measurements by assuming that the actual litterfall is a **multiple** of the observed litterfall."

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?

We discussed the introduction of the bias parameters with literature references in more detail in section 2.3.1 where we define the prior distributions. We will make a cross-reference to this section. In the results and discussion we dedicated a whole section to "3.5 Interpretation of litter input bias parameters".

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?

We fully agree. The identifiability issue is discussed with the correlation matrices in section "3.2 Correlations between 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.

The identifiability issues are already discussed in great detail in section "3.2 Correlations between parameters" and with Figure 5. In the opening statement of this response we acknowledge that a Hidden Markov approach, which separates observation error and process model errors more explicitly, might also be an alternative to the usage of bias factors.

***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.

Will be split.

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.

Yes, we will introduce these as parameters in the calibration again. Consequently all Figures except Figure 1 will be redrawn to include also $tlag_L$ and $tlag_R$. Another section in the discussion will be introduced to discuss these lag times. The identifiability issues will increase – this will be discussed in the section on correlations between parameters "3.2 Correlations between parameters". The remainder of the text will be updated to account for this change.

We will use the measured lag-times as modes of log-normal priors for $tlag_L$ and $tlag_R$.

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.

We actually treat the initial conditions as a part of the estimation problem and do not perform a spin-up: we estimate Y_{ini} and O_{ini} by allowing them to deviate from steady state with the multiplicative parameters f_Y and f_O. In order to clarify this better, we will introduce another table where we list all parameters that are estimated: This includes the initial conditions of Y and O via the parameters f_Y and f_O (although this is already stated in Fig. 1)! We think that the f_Y and f_O parameters are well suited to treat initial conditions as unknowns in situations where one also wants to model radiocarbon.

Pg 13811
line 4: "analog way"

Will be rephrased.

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

Will be corrected.

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.

This is a simple rule of three with the error propagation described in detail by Phillips and Gregg (2001). Philips and Gregg (2001) show that the source partitioning we used is a **linear** two-source, single isotope mixing model. In our opinion, there is no need to treat it differently from what way it was proposed by Philips and Gregg.

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!

We will amend this section or make reference to a supplement that clarifies where we actually made use of Equation 20:

(1) The uncertainty of the SOC stock weighted $\Delta^{14}C$ value (equation 15) was calculated as:
$\delta q(\Delta^{14}C_{SOC,bulk}) =$

$$\sqrt{\sum_{i=1}^{Horizons}\left(\frac{\partial q}{\partial \Delta^{14}C_{SOC,i}}\cdot \delta\Delta^{14}C_{SOC,i}\right)^2+\sum_{i=1}^{Horizons}\left(\frac{\partial q}{\partial SOC_{stock,i}}\cdot \delta SOC_{stock,i}\right)^2} \quad (S1)$$

where the partial derivative in the first term under the square root is

$$\frac{\partial q}{\partial \Delta^{14}C_{SOC,i}}=\frac{SOC_{stock,i}}{\sum_j^{Horizons} SOC_{stock,j}} \quad (S2)$$

and the partial derivative in the second term under the square root is

$$\frac{\partial q}{\partial SOC_{stock,i}}=\frac{\sum_j^{Horizons} SOC_{stock,j}\cdot\left(\Delta^{14}C_{SOC,i}-\Delta^{14}C_{SOC,j}\right)}{\left(\sum_j^{Horizons} SOC_{stock,j}\right)^2} \quad (S3)$$

(2) The uncertainty of the proportion f_{HR} of heterotrophic respiration at soil respiration (equation 19) was be calculated as (Phillips and Gregg, 2001;Taylor, 1997):

$$\delta q(f_{HR})=\frac{1}{(\Delta^{14}C_{HR}-\Delta^{14}C_{RR})^2}\cdot\left(\delta^2_{\Delta^{14}C_{SR}}+f_{HR}^2\cdot\delta^2_{\Delta^{14}C_{HR}}+(1-f_{HR})^2\cdot\delta^2_{\Delta^{14}C_{RR}}\right) \quad (S4)$$

Uncertainties for SOC, HR and $\Delta^{14}C_{HR}$ were taken directly from the original publications we refer to in section 2.2.4 "Measurements and data processing" or were calculated from uncertainties reported therein using the basic rules for error propagation for sums and products.

We explain why we think that it is appropriate to assume zero covariance between the measurements to get a conservative and realistic uncertainty estimate in the reply to comment "***Page 13821:****Line 6*" and "**Page 13820**: *line 2*".

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.

Yes, that could be a point of discussion. Due to the fact that we are modelling bulk SOC stocks and $SO^{14}C$ stocks we decided to calculate SOC stocks and $SO^{14}C$ to the same depth. In our opinion this assures that the state that we compare the model to is consistent. It is not possible to calculate a bulk ^{14}C signature to a depth where ^{14}C was not measured.

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

Yes. This and later personal communications with ourselves will be removed.

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

These other studies are from the same site. We will clarify this statement. We decided to take more conservative (larger) SE than the other studies, because the SOC stocks we use at the Howland Tower site are only based on one soil profile.

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.

We will rephrase that. Already our other two sites show that aboveground and belowground litter input (Coulissenhieb II and Solling) is in the same range (also McClaugherty et al. (1984) and Persson (1978)).

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

Will be clarified. The data is 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.

We will include the reference to the original paper. We checked for autocorrelation between the repeated measurements, but partly because the measurements are quite scarce (^{14}C data), we did not find significant (positive) autocorrelation between the repeated measurements.

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

Yes, we will include $tlag_L$ and $tlag_R$ as additional parameters that are calibrated within our Bayesian calibration framework. The measured lag-times at the individual sites will be used to define priors for the parameters.

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

It is the same forest, but not part of Solling D0 experimental site. We will point out that we are assuming that SOC stocks did not change drastically within 4 years.

Page 13821:
Line 6: Data for different horizons in the same pits are not independent. Eqn 20 doesn't hold.

These are not the same pits. Please remember that we have 61 samples in the Oi + Oe horizon, 40 in the Oa, and 5 for the mineral soil for the SOC stocks. The ^{14}C was measured in three other pits. Furthermore, we used the reported mean values – this is why it is not possible to quantify if the uncertainty is reduced due to negative correlations or if it is increased due to positive correlations.

We would also like to point out that assuming independence between horizons and different measurements is already a conservative estimation of uncertainties because including the covariance between variables into the error propagation normally reduces the uncertainties due to negative correlations, i.a. between bulk density and SOC content (Panda et al., 2008;Goidts et al., 2009;Schrumpf et al., 2011). Because we do not have information about the actual covariance between measurements at the individual sites, we are not able use these directly in the error propagation.

As an alternative one could argue to use the maximum uncertainty equation (Taylor, 1997):

$$\frac{\delta q}{|q|} \leq \left|\frac{\partial q}{\partial x}\right| \delta x + \left|\frac{\partial q}{\partial z}\right| \delta z$$

In our opinion the use of this equation would lead to a massive overestimation of uncertainties because of the inherent assumption in this equation that the correlation between all variables used to calculate SOC stock or bulk $\Delta^{14}C$ is +1.

Line 18: uncertainty on this estimate?

Murach et al. (1993) report a standard deviation of the mean of about 100%.

Page 13822:
Line 10: you can't prove convergence of a MCMC

We will rephrase that sentence.

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.

That is a misunderstanding. We are actually looking at total aboveground litterfall, but e.g. at Coulissenhieb only foliar litterfall was measured, so that the total aboveground litterfall estimates are possibly too low. We will clarify the section by giving more specifics.

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.

We reported the 1.2 to 1.7 as a justification for using the bias parameter at all. Please note that the Gerstberger study is from Scots pine and not Spruce forest. We will increase the 99th-percentile 2 for $bias_{iL}$.

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.

Yes. That is good point. We will increase the 99th-percentile for the prior of $bias_{iR}$ to 3.

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).

No, you are probably mistaken here. We are not allowing for negative SOC and HR! f_Y and f_O are bound at zero (the priors follow a truncated-normal distribution, truncation at zero)!

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.

The subscript i designates one of the different datastreams (HR, SOC Δ^{14}C(HR) and Δ^{14}C(SOC)). We will reiterate in a Table here that "All parameters except $tlag_L$, $tlag_R$ and λ are calibrated" (Fig. 1). We had fixed $tlag_L$ and $tlag_R$ before, but given your comments we will also calibrate these two parameters. We will also clarify that we are not fitting each dataset separately, but did a calibration to all datastreams at once that were included in the respective calibration exercise (Table 2).

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.

In the next version we will state more prominently that we are assuming that we can neglect process error. Additionally, our observation errors are already quite large, but we have to emphasize that we compare our model against means + SDOM, so that we correct for different sample sizes of the different datastreams.

Given the scarcity of our data (radiocarbon, in some cases only one point per datastream), it seemed quite dubious to us to sample the σ of an individual data point (random or systematic error?). Furthermore, we would have to make sure for a prior of σ that σ cannot get smaller than the observed SDOM. Therefore, we did not sample σ in our Bayesian calibration for example as a "nuisance parameter".

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?

For each site and each calibration setup we ran 5 chains in parallel with 100000 iterations each. We discarded the first 50000 iterations of each chain and checked if the within and between chain point scale reduction factor < 1.025 *sensu* Gelman et al. (2004). Additionally, we visually checked for convergence with traceplots for the 5 chains and each parameter. Furthermore, we checked density plots for each parameter and chain to ensure that inferences from different chains would give the same results. The second halves of the chains were merged and thinned to a total sample size of 16666 (every 15th sampled value was kept). As already mentioned in the manuscript "we used the DRAM implementation of Soetaert and Petzoldt (2010)" to perform our calibration within the statistical software R-2.15.1.

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.

Yes. That makes more sense.

Page 12826
line 20: The bias parameters are probably contributing to the nonidentifiability of these parameters

Yes. We will add some lines discussing this.

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

Yes. We found that quite insightful.

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

We agree that this is the accurate interpretation for the correlation between the two bias parameters. We will change our previous statement along the lines you suggest here.

line 29: "stronger constrained"

Will be rephrased.

Page 13828
line 1: "shall be" → is

Will be corrected.

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.

We agree.

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

OK

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])

We meant that major disturbances could be included in a stronger prior, e.g. a wind throw or a clear-cut, and subsequent reforestation could call for prior with a mode that is not at equilibrium. We will clarify this point in the discussion.

We already showed with the steady-state run which does not differ much from the non-steady state run (Fig. 7, Fig. 8) that the data is just not enough to constrain the source-sink strength of these soils.

Furthermore, we did not use a spin-up to run the model into equilibrium, but used Eq. (8), (9), (10) and (11) to calculate the equilibrium of the young and old pool, and its respective ^{14}C valueanalytically. Hence, given a set of parameters we can accurately calculate the equilibrium states of the model.

Page 13832

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

Yes, we will do that in a revision.

References

Andrén, O., and Kätterer, T.: ICBM: The introductory carbon balance model for exploration of soil carbon balances, Ecological Applications, 7, 1226-1236, 1997.

Gelman, A., Carlin, J., Stern, H., and Rubin, D.: Bayesian data analysis, 2nd ed., Chapman & Hall/CRC, Boca Raton, 2004.

Goidts, E., Van Wesemael, B., and Crucifix, M.: Magnitude and sources of uncertainties in soil organic carbon (SOC) stock assessments at various scales, European Journal of Soil Science, 60, 723-739, 10.1111/j.1365-2389.2009.01157.x, 2009.

Kavetski, D., Kuczera, G., and Franks, S. W.: Bayesian analysis of input uncertainty in hydrological modeling: 1. Theory, Water Resources Research, 42, 2006.

McClaugherty, C. A., Aber, J. D., and Melillo, J. M.: Decomposition Dynamics of Fine Roots in Forested Ecosystems, Oikos, 42, 378-386, 10.2307/3544408, 1984.

Murach, D., Klaproth, F., and Wiedemann, H.: Feinwurzeluntersuchungen auf den Versuchsflächen des Dach-Experiments im Solling, Forstarchiv, 64, 188-191, 1993.

Panda, D. K., Singh, R., Kundu, D. K., Chakraborty, H., and Kumar, A.: Improved Estimation of Soil Organic Carbon Storage Uncertainty Using First-Order Taylor Series Approximation, Soil Sci. Soc. Am. J., 72, 1708-1710, 10.2136/sssaj2007.0242N, 2008.

Persson, H.: Root Dynamics in a Young Scots Pine Stand in Central Sweden, Oikos, 30, 508-519, 10.2307/3543346, 1978.

Phillips, D., and Gregg, J.: Uncertainty in source partitioning using stable isotopes, Oecologia, 127, 171-179, 10.1007/s004420000578, 2001.

Rahn, K. H., Werner, C., Kiese, R., Haas, E., and Butterbach-Bahl, K.: Parameter-induced uncertainty quantification of soil N2O, NO and CO2 emission from Höglwald spruce forest (Germany) using the LandscapeDNDC model, Biogeosciences, 9, 3983-3998, 10.5194/bg-9-3983-2012, 2012.

Reinds, G. J., van Oijen, M., Heuvelink, G. B. M., and Kros, H.: Bayesian calibration of the VSD soil acidification model using European forest monitoring data, Geoderma, 146, 475-488, 10.1016/j.geoderma.2008.06.022, 2008.

Santaren, D., Peylin, P., Viovy, N., and Ciais, P.: Optimizing a process-based ecosystem model with eddy-covariance flux measurements: A pine forest in southern France, Global Biogeochemical Cycles, 21, GB2013, 10.1029/2006GB002834, 2007.

Schoups, G., and Vrugt, J. A.: A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic, and non-Gaussian errors, Water Resour. Res., 46, W10531, 10.1029/2009WR008933, 2010.

Schrumpf, M., Schulze, E. D., Kaiser, K., and Schumacher, J.: How accurately can soil organic carbon stocks and stock changes be quantified by soil inventories?, Biogeosciences, 8, 1193-1212, 10.5194/bg-8-1193-2011, 2011.

Taylor, J. R.: An introduction to error analysis: the study of uncertainties in physical measurements, University Science Books, Sausalito. USA, 1997.

van Oijen, M., Rougier, J., and Smith, R.: Bayesian calibration of process-based forest models: bridging the gap between models and data, Tree Physiology, 25, 915-927, 2005.

van Oijen, M., Reyer, C., Bohn, F. J., Cameron, D. R., Deckmyn, G., Flechsig, M., Härkönen, S., Hartig, F., Huth, A., Kiviste, A., Lasch, P., Mäkelä, A., Mette, T., Minunno, F., and Rammer, W.: Bayesian calibration, comparison and averaging of six forest models, using data from Scots pine stands across Europe, Forest Ecology and Management, 289, 255-268, http://dx.doi.org/10.1016/j.foreco.2012.09.043, 2013.

Yeluripati, J. B., van Oijen, M., Wattenbach, M., Neftel, A., Ammann, A., Parton, W. J., and Smith, P.: Bayesian calibration as a tool for initialising the carbon pools of dynamic soil models, Soil Biology & Biochemistry, 41, 2579-2583, 2009.