We would like to thank the editor for taking the time to handle our manuscript and for finding three very constructive reviewers. We also want to thank all reviewers for taking the time and reviewing our manuscript to help improve its quality. We are grateful for the honest and thorough feedback. The suggestions were highly useful and provided us with information, where misunderstandings could be possible and where we needed to make our message clearer and to discuss the limitations of the DSI in more detail. They helped to further improve the quality of this manuscript and we hope that we addressed concerns to a satisfying extent. Our comments to the reviewers in the following are in blue color. We made use of the constructive criticism and altered the text of the manuscript, where applicable. We added screenshots of alterations in the text related to the comments. These are displayed in green color.

## Comments of Reviewer 1: Sander Bruun

**General comments**

The papers deals with initialization of pools in the soil organic matter model of Daisy. The paper is using unique datasets for long-term fallow treatments to test a new way of initializing the soil organic matter pools based on specific peaks in the DRIFTS specta of the soils. Pool initialization of SOM model is an important issue that is still causing some difficulties with the currently used approaches. The paper is therefore very timely and present an interesting approach that could be useful in many situations. The work is of a high quality and based on high quality data and the manuscripts is well written.

We are glad that reviewer found our work useful and interesting.

**Specific comments**

Line 78. I agree that the DSI can be better than the steady-state assumption, but perhaps it is worth discussing this in a little more detail. If information about the history of the site is available then that method should work. This require that the history is known for millennia, and that is rarely the case.

We agree and added a corresponding sentence as suggested, at line 56.

> lack of data of residue input and weather data for the required long-term timescales (from > 200 years to millennia). So, while the approach should work in theory, the history of a site is usually not known for the timescales that SOM needs to equilibrize. Therefore, the simulation of past carbon inputs and the assumption of

Line 118: Was soil samples from throughout the experimental period analyzed? Please specify.

Yes, from throughout the period. We specified this now more clearly.

> Bulk soil samples from the start and throughout the simulation period of all experiments were analyzed for total carbon and DRIFTS spectra; samples from the Kraichgau and Swabian Jura sites were additionally analysed for soil microbial biomass carbon (SMB-C). After sampling, all bulk soil samples (except for SMB-C) were passed

Line 129: The spectra were not recorded in absorbance, but subsequently converted to absorbance units, right?

Yes – the wording was changed

combination of 16 co-added scans with a resolution 4 $cm^{-1}$. Spectra were recorded and then converted to~~in~~ absorbance units (AU); the acquisition mode "double-sided, forward-backward" and the apodization function

Line 130: I wonder how much this way of determining the DSI is affected by the instrument i.e. if somebody took the same soils and did the measurement on another instrument would the get the same DSI and pool sizes. I am afraid that it would be quite much affected by that especially if you use other IR detection techniques. Maybe it would be worth addressing this in the discussion.

Indeed, at least to our experience, there are some differences between the spectra of different spectrometers, especially between detectors. We added a sentence addressing this in chapter 4.1. As we already tested different temperatures for drying, which we found to be the most dominant factor affecting DSI, it was beyond the scope of this publication to test the effect of the spectrometer. We were first and foremost interested in, whether the DSI approach adds value in general to SOM initialization, which we think it does.

There are some remaining questions that should be answered to standardize the application of the DSI for model initialization. Those are related to how the type of spectrometer influences the spectra, as well as how water and mineral interferences (Nguyen et al., 1991) in the spectra can be eliminated or at least be further reduced. We had the experience, that spectra and therefore peak areas vary to some degree between different spectrometers (mostly due to different detectors, types of detector cooling and resolution). Hence, it will be necessary to either use the same spectrometer, or to develop techniques to standardize spectra across a large number of instruments.

Line 181: It says 84% and not 83% in Table 2. Please correct where appropriate.

We have done so

Line 196 to 209: I am not entirely sure I understand what the point of analyzing the SMEx with a statistical model is. I think you should consider whether it add enough understanding to warrant inclusion. Alternatively explain the point a little better.

We wanted an analysis of the model error which could give us a better measure of model uncertainty, and since in some experiments (Swabian Jura and Kraichgau) we had several fields, make use of the statistical power provided by the experimental design. The second advantage of a statistical analysis of model error was, that we could analyze for a time trend (increase with time) of the model error.

mixed model with $SME_x$ as response was then used to test for significant differences between initialization methods. This approach allowed us to make use of the statistical power of the three Kraichgau and Swabian Jura fields to analyze which initialization was most sound and for a trend of the model error with increasing simulation time. In some cases, $SME_x$ was transformed to ensure a normal distribution of residuals (square root

Line 236-237. The necessity of constraints on the fSOM-Slow parameter is a little problematic. I cannot help thinking that it means that the data, which is used for calibration, is insufficient. With these restraints, I guess you are likely to end up with a value of 0.35 which is rather arbitrarily chosen by you.

From our perspective, rather than a data limitation, this is an indicator how model structure affects the results of Bayesian calibration. In the initial first Bayesian calibration without limits, fSOM_slow was well constrained by the calibration (Figure S5 in the manuscript), but to a value we consider implausible (~ 95%). Therefore, we suggested a possible alternative formulation of DAISY (Figure 7 in original text). While recently testing the proposed revised model structure of DAISY, we found that with this new model formulation, fSOM_slow does not have a trend towards the upper constraints ( >= 80%) anymore (the high humification efficiency values here, are because little new SOM is coming in within the bare plots) , even without artificial constraints. See as an example the results of the new structure with (2) and without (3) the fSOM_slow constraints compared to (1) the initial BC of this study:

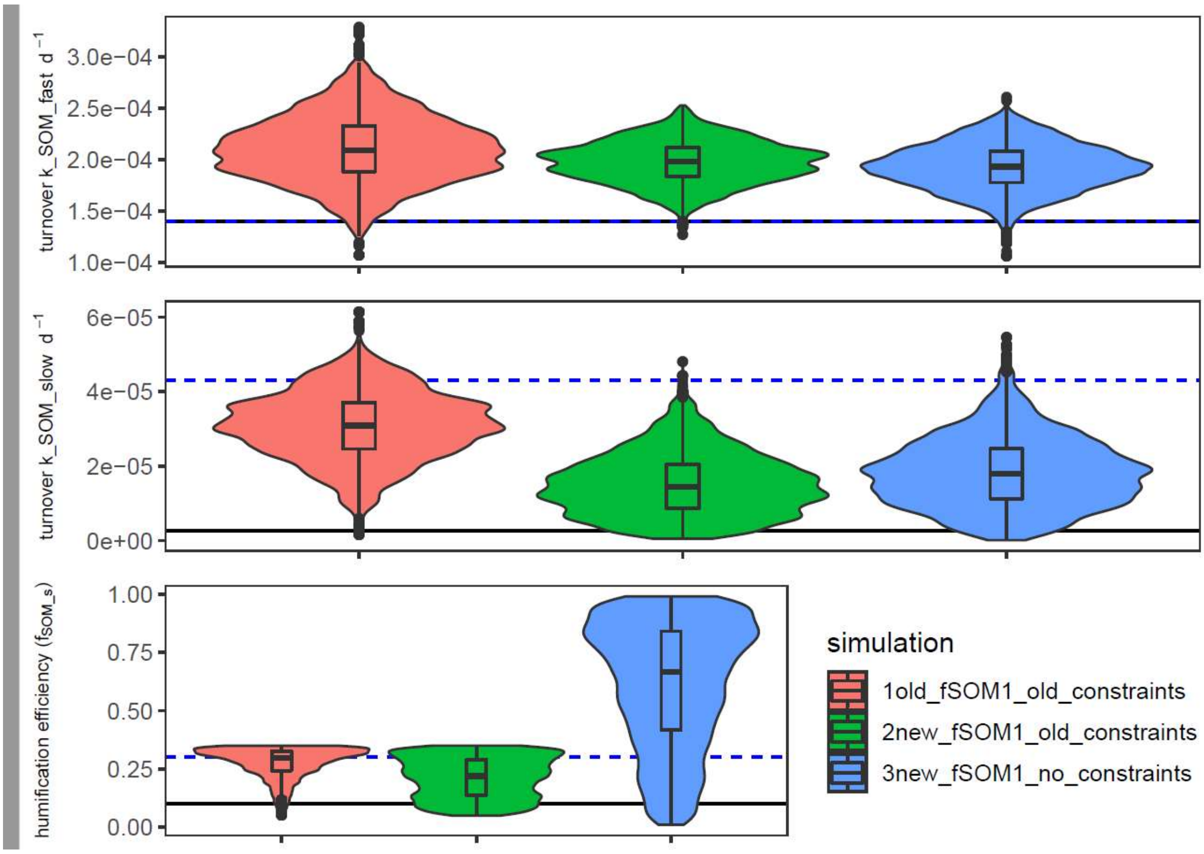


**Figure S 1 Violin plots of the parameters, obtained by the Bayesian calibration using the new suggested model structure (Old constraints are 0.05 and 0.35, no constraints means 0.01 and 0.99.). The black line corresponds to the parameters of Mueller (1997), the blue dashed line to the parameters of Bruun (2003).**

We also added two more sentences to discuss this points.

> SOM aligned with the earlier published rates. If a parameter is problematic, such as $f_{SOM\_slow}$ it could mean that there is a lack of data, especially if it is not identifiable by the Bayesian calibration. However, if parameters are clearly constrained by the Bayesian calibration, but those constraints are implausible, it usually means that the model structure is suboptimal (Poeter et al., 2005) and should be altered.

Line 364-365. I agree that even though we have had the same management for a longtime the steady-state assumption is not valid. However, I believe that the reason for this has to do with longer-term effects rather than the smaller effects that you mention i.e. variation in climate agricultural management. If you look at a longer terms, most sites would probably have been deforested within the last 2000 years. Because of the high inputs from the forest, this could have resulted in an unusually large fraction of resistant organic matter that has not been degraded from that period. Also it is very common with drained soils soil. This means that the soil at some time it its history has had a very high water table and perhaps even been inundated. We know that this can result in significant accumulation of organic matter. After the soil has been drained, this has led to a large residual of resistant C again. The same could happen if there has been a history of fires with inputs of charcoal. Perhaps this is worth discussing a bit more.

We agree and added these possibilities to the main text.

> particular field. This is particularly relevant, given that the changes in genotypes of crops, agricultural management, crop rotations and the rise of average temperatures in recent decades as well as stronger land use changes such as organic soils draining or deforestation in recent centuries probably have affected the past quality and quantity of carbon inputs to soil. Consequently, the steady state assumption for model initialization is not

Line 373: I cannot help it thinking that it is somewhat of a coincidence that you get better model performance with the DSI as long as you have not recalibrated the model. Of course using more data as for example DSI to restrain the model should improve the model, but only after it has been recalibrated.

We interpreted this from the fact that SMB-C simulations were best when using the DSI as indicator, even if the turnover rates are unclear. As SMB-C is a much faster reacting pool than TOC, which did not change that much in our trials in Kraichgau and Swabian Jura. The DSI at 105°C was consistently lower in model error for simulated SMB-C than the steady state initialization, which should indicate that it is a useful proxy regardless of turnover rate, as long as there is a clear distinction between fast and slow pools.

It is not entirely clear what data were used for the calibrations based on DSI. As far as I understand, you measured DSI of all the soil samples and that means that you can compare the simulated distribution between fast_SOM1 and slow_SOM with the one measured and calculated using formula (2) and a similar formula for fast_SOM. Is this the case? And if it is why have you not shown the “measured” value of fast and slow SOM and compared it with the modelled?

You are correct, we used the measured DSI throughout the simulation period for the Bayesian calibration. We are happy to provide the modelled vs measured DSI throughout the simulation period – we also added it to the manuscript:

> The resulting amount of SOC in the slow pool according to the computed DSI changed from the initial range of 54 to 80 % to the range of 76 to 99% at the end of the observational period (Figure S 7). The SMB-C reacted

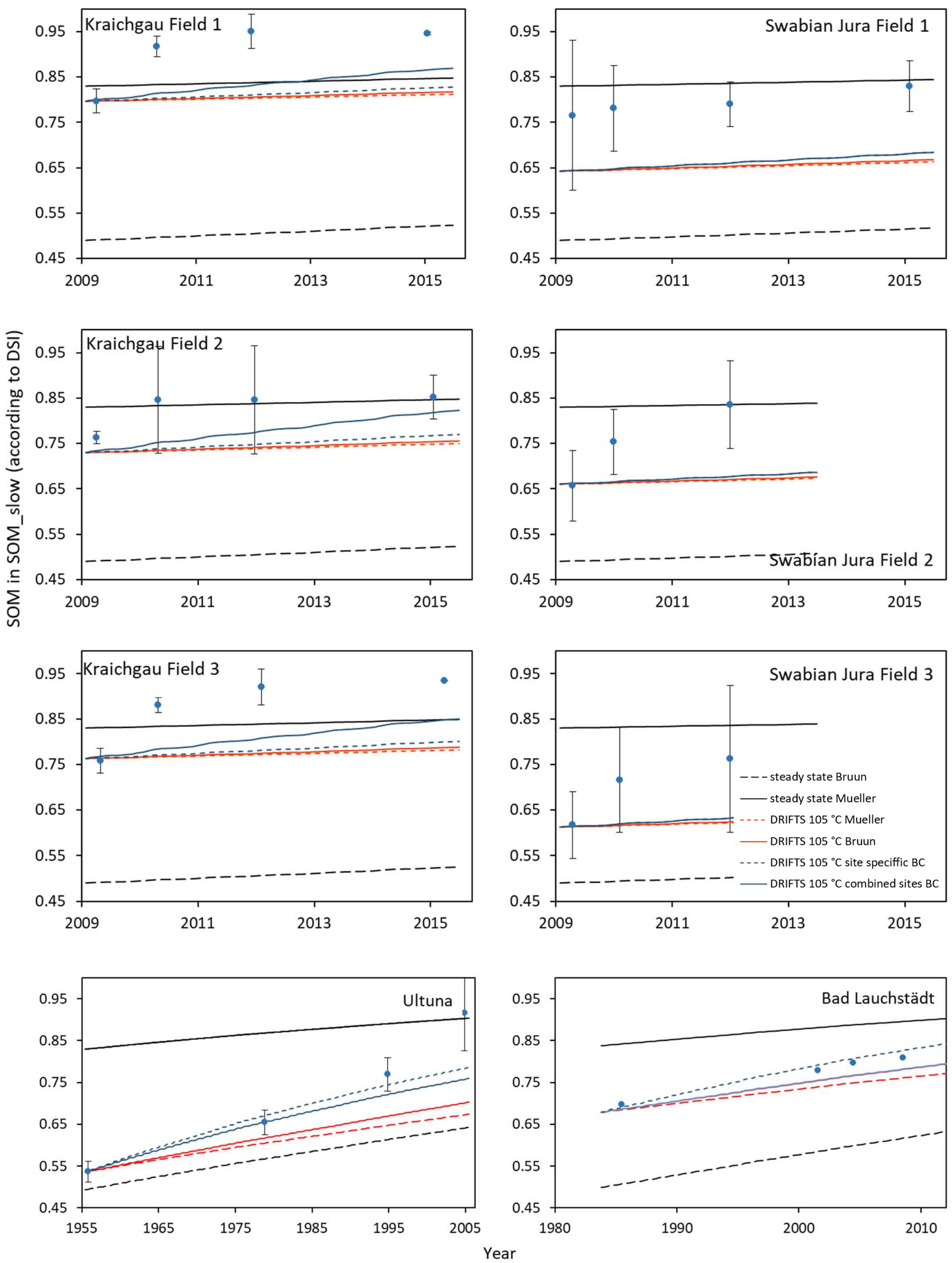


**Figure S 2 Development of simulated vs observed SOM in the slow pool, according to DSI division throughout the simulation period (for brevity only for 105 °C). Bars indicate standard deviation of all plots per field.**

Is it worth publishing the optimal parameters selected by the Baysian calibration based on DSI?

While we think that the ideal way to use our results is using the posterior probability distributions of our parameters, we have mentioned the parameter set of the maximum likelihood from our

Bayesian calibration in chapter 3.3 (0.34, $2.29 * 10^{-4}$, $3.25 * 10^{-5}$ for the original weight calibration and 0.06, $9.58 * 10^{-5}$ and $5.54 * 10^{-5}$ for the calibration using original weights and no DSI) and in Table 5.