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 2:

General comments

Laub and Colleagues present interesting ideas how DRIFTS spectra could be used to initialize and calibrate soil organic matter models. What warrants more discussion is that with their results we should put again more emphasis on the chemical recalcitrance hypothesis, i.e. that molecular properties determine the persistence of organic matter in soils. The literature seems to disagree (Schmidt et al., 2011). If we indeed assign the aromatic peak to slow cycling pools with a turnover time of 426 years and the aliphatic peak to a fast cycling pool with 47 to 90 years, the authors would contradict the synthesis of Schmidt et al. (2011) (their Figure 1, for example).

We do not think, that our results contradict Schmidt et al. (2011). Rather, the DSI seems to point towards the same direction as other measures of SOM quality, such as the amount of SOM in different aggregate sizes and density fractions. This was actually shown in our previous works (Demyan et al., 2012). We have to keep in mind that the DSI is still only a proxy and dividing the whole continuum of SOM quality into two discrete “qualities” is a strong simplification of the real world. However, we think it seems to be a valid one, especially when two pool SOM models are to be used, which anyway divide SOM into two pools. Additionally, a physical protection of SOM is implicitly included in DAISY, in the form of a clay function reducing SOC turnover.

> 65°C > 32°C drying temperature (differences being sometimes but not always significant). It has to be noted, that the DSI is not purely related to chemical recalcitrance of SOM, as it also correlates with the level of SOC protected by aggregation (Demyan et al., 2012). Hence, it is likely that aggregation and chemical recalcitrance are related.

In my opinion, it would be interesting if the authors could at least discuss how their DRIFTS peaks could be useful for the new class of microbial-mineral models such as Tang and Riley (2015) or Sulman et al., (2014)

We think that DRIFTS could also be useful for those models, because of a good correlation of the DSI to size density fractions (Demyan et al., 2012), which is thought more representative of structural protection mechanisms. We added one sentence about this in the discussion.

> physically meaningful and to reduce model uncertainty and equifinality. As the DSI also had a good correlation with structurally protected SOM (Demyan et al., 2012) it could also fit very well to models that directly simulate the protection of SOM as a function of microbial activity (Sulman et al., 2014). A better understanding and the

Specific comments

The authors state that “the DRIFTS initialization of SOM pools significantly reduced model errors of poor performing model runs assuming steady state, irrespective of the turnover rates used, but the faster turnover parameter set fit better to all sites except Bad Lauchstädt. This suggests that soils under long-term agricultural use were not necessarily at steady state.” In my opinion this statement is not backed up by their results. The Bruun parameters with steady state assumption perform better at Ultuna and Kraichgau + Swabian Jura (Table 4) for SOC stocks.

We agree that our original text could be misinterpreted, so we altered the wording. For this statement we placed more weight on the Kraichgau + Swabian Jura sites (because those consisted of six fields) and assumed that Ultuna with Bruun turnover rates were not performing poorly. We observed there a significant improvement of the estimation of the sensitive SMB-C pool, for both turnover rates, while for SOC stocks there were only significant differences in model errors between turnover rates but not between initialization methods.

> efficiency, respectively). The DRIFTS initialization of SOM pools could significantly reduce~~d~~ model errors of ~~poor performing~~ model runs assuming steady state if they were performing poorly, irrespective of the turnover rates used, but the faster turnover parameter set fit better to all sites except Bad Lauchstädt. This suggests that

The authors also state that “[...] two approaches [...] significantly reduced parameter uncertainty and equifinality”. One of the approaches was the inclusion of DRIFTS. But looking at the violin plots in Figure 5, only the humification efficiency seems to be better constrained. I suggest modifying the statement towards this direction.

It is true, that humification efficiency was the parameter most seriously constrained by the DSI, also the turnover of the slow carbon pool was more strongly constrained (standard deviation of $9.3 * 10^{-6}$ with DSI vs $12.3 * 10^{-6}$ without DSI). We altered the wording in the sentence, to be more accurate.

> individually and for a combination of all sites. The two approaches which significantly reduced parameter uncertainty and equifinality were: 1) the addition of the physico-chemically based DRIFTS stability index (strongest for humification, but also for slow SOM), and 2) combining several sites into one Bayesian calibration, as derived turnover rates can be strongly site specific. The combination of all four sites showed that

I agree with the other reviewer, Sander Bruun, that analyzing the squared model errors with a statistical model should at least be better explained.

This was done, see the comment to Sander Bruun.

> 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

The manuscript would benefit from a thorough spell and language check.

This will be done on the final reviewed version manuscript.

References:

Demyan, M. S., Rasche, F., Schulz, E., Breulmann, M., Müller, T. and Cadisch, G.: Use of specific peaks obtained by diffuse reflectance Fourier transform mid-infrared spectroscopy to study the composition of organic matter in a Haplic Chernozem, Eur. J. Soil Sci., 63(2), 189–199, doi:10.1111/j.1365-2389.2011.01420.x, 2012.