Luo and coauthors present a nice analysis that examines the challenges in parameterizing and reducing uncertainty in soil C models that are used for land management and policy decisions. Even with an spatially and temporarily robust dataset from agricultural sites around Australia they find that well calibrated models over the observational period still show significant uncertainty in trying to make future projections about the fate of C in a changing world. While I strongly agree that better understanding and structural representation of microbial physiology, C quality, and management effects are needed to reduce uncertainty in soil C projections (section 4), I'm not convinced the data presented clearly support these conclusions.

A significant amount of confusion is generated because the manuscript presently conflates model structural uncertainty, parameterization uncertainty, and forcing (or scenario) uncertainty in the analysis (see Hawkins and Sutton 2009). In me estimating focusing on the first one of two of these types of uncertainty would strengthen the conclusions being made here.

General comments

- *Parameter uncertainty*: A significant amount of effort went into reducing parameter uncertainty in the model at each site (Figs 1 & 2) I'm surprised that sensitivity analyses of temperature, moisture, and N scalars weren't considered in this model since previous work demonstrates that model results are strongly determined by these parameters (e.g., Todd-Brown et al. 2013; Exbrayat et al. 2014). Would consideration of these parameters in the optimization routines better constrain the projected uncertainty, or compound the equifinality problems mentioned in section 3.1?

  Fig. 2b shows a split in SOC projections for both high and low inputs. One is left to surmise this bifurcation in results is generated by the concurrent split in parameter space shown in Fig. 2d. The authors hint at this finding at the end of section 3.3, it's never adequately discussed in section 3.2, where optimization results are presented.

  Optimized CUE values seem quite high in Figs. 1ab, 2, especially given conclusions by Sinsabaugh and others (2013) that CUE values in soils should be considerably lower? I wonder if better constraints on the prior distributions of parameter values may lead to different conclusions? I'm not sure such analyses are warranted here, but discussing this dependency of prior distributions in Bayesian analyses seems warranted.

- *Structural uncertainty*: Similarly, it looks like all the sites have very different management practices (SI Table 1), but the effects of these different management practices are implicitly represented by site-level parameter estimation for steady state and temporal changes in soil C for each of the sites & treatments. Instead, I wonder if a single model would be better, with "global" decay constants that are modified by scalars for different management practices- in addition to temperature, moisture, and N scalars that already being used? With so many unconstrained parameters this approach may run into the same equifinality problems, but also may better constrain management effects on future soil C storage? I'm not asking that environmental or management scalars be evaluated here- but their potential importance should be discussed. Instead, my larger concern comes in how uncertainty analyses were conduced and the inferences drawn from them.

- *Forcing uncertainty*: It's not clear what actually generates the uncertainty shown in Fig. 2? It seems as though SOC parameters were optimized (Fig. 1), but that uncertainty in the crop response generated wide uncertainty in plant productivity, and therefore soil C inputs (which were not previously optimized). As the authors hypothesize in section 2.5, first order models are very sensitive to soil C inputs (again Todd-Brown et al. 2013). Projected inputs varied by more than a factor of two (section 3.2). Thus, uncertainty shown in Fig. 2a does not surprising- if this is what's actually generating the spread in projections? If so, I'm not confident that conclusions about persistent uncertainty in soil C projection (section 4) are well supported by this analysis?

  To control for differences in plant inputs, could the authors increase residue by 10% for different parameterizations of soil C at each site and quantify the variation in SOC projections? Subsequently, what if temperatures warmed [or soils dried] over the 100-year projection window, how would the temperature sensitivity of decomposition vary depending on tradeoffs between humus decay rates and partitioning to inert C? There is some speculation towards this effect in the middle of p 4261, but it's not clear how the authors generate climate uncertainty effects on soil C storage here? By isolating these variables, uncertainty in parameter estimates and/or model structure could be isolated (if this is the focus of the paper, as implied in the abstract), and would avoid confounding forcing uncertainty in the analysis.

- *Uncertainty attribution*: I have to admit that I'm not really clear what the intercept (α) and slope (β) parameters are showing (sections 2.6 & 3.3)? The authors conclusions seem to strongly rest on the change in α and β over sites and time (first paragraph of section 3.3). First order models can exhibit false priming (as in Fontaine et al. 2011) because initially increase soil C inputs enter pools with faster turnover times (FOM in this model), thus increasing soil respiration rates more than may be expected. Over time, however, as more C enters larger SOM pools with slower turnover times (humus and inert pools here) and the system begins to achieve a new equilibrium state the crop residue effects (C_r) on percentage uncertainty (U_p) should increase. It's not clear if this is what's going on here, but I'd suspect this may explain why α and β both increase over time (Fig. 4b)? Similarly, sites with "well behaved" parameter estimates that have a narrow range of values for rdhum and finert (e.g., Tarlee, Fig 2a,c) likely have low α & β values, whereas sites that generate bimodal distributions of parameter combinations (e.g., Brigalow Fig. 2b,d) will have larger α & β values? Is that what's being shown in Table 1?

- *Unsubstantiated claims*: In the same paragraph (bottom of pg 4259), there's discussion of 'optimal agricultural management'- which as something to do with residue management and N application? The authors also make what seem like widely speculative claims on the potential changes in agricultural soil C changes. Details of how these extrapolations were generated are lacking from the text, and I recommend removing this seemingly tangential finding from the text.

  Also, discussion of the potential effect of cultivation history on CUE seems very speculative (bottom of p. 4261). Although it's an interesting idea, with only a single site under "long" and "short" cultivation history, the results seems spurious at best, with no mechanism as to what would drive such changes in microbial physiology as a function of land use practices.

Technical comments

- *Precise language*: The phrase ‘carbon composition’ is mentioned several times in the abstract (p. 4246, l. 14, 22, & 23) as well as several times in the main text (e.g. p. 4250, l. 9; 4250, l. 14; etc.) but this term is somewhat ambiguous. Is this referring to the chemical quality of SOM, its physical accessibility to microbes, or something else? Can the authors use more precise language for this phrase?
- *Structural Clarity*: The authors refer to the ‘calibration’ of their model and the ‘calibration period’, however, this procedure is never really described in the methods. I suspect that ‘calibration’ and ‘optimization’ (described in section 2.4) are being used interchangeably here, but this may not the true? Care should be taken to clarify language so readers can accurately understand results and discussion in the context of the numerical methods being applied. Maybe subheadings in section 3 that match those in the methods would help clarify results. (e.g., 3.1 Sensitivity analysis; 3.2 Optimization; etc).

  Similarly, it seemed as though results from DE optimization were going to be compared to the Bayesian approach (top of page 4252). It seems like Fig. 2 presents results from the DE optimization and Fig. 3 shows results from the Bayesian approach; however, from a comparison of the two methods is not clearly presented. I think this is actually discussed at the bottom of page 4261, and in SI Fig. 4 (section 3.3), but this text should be move up to the optimization section (section 3.2), as described in the methods (section 2.4).

  Use of model abbreviations in the text that are not clarified in the model conceptual diagram (SI Fig. 2) unnecessarily obscures findings for readers who are not intimately familiar with the model. The model is simple enough to deduce the abbreviations being used, but could be made more direct by labeling parameters of interest on SI Fig. 2, and / or simplifying the parameter names (e.g. k_{carb}, $k_{cellulose}$, k_{lignin}… to describe the first-order decay constants of each pool).
- *Technical clarification*s: How does material get into the “Inert C pool”? This isn’t clearly described in the text of evident in SI Fig. 2, but it’s an important parameter in the model according to the sensitivity analysis (section 3.1, SI Table 2). Similarly CUE (which I’m used to seeing capitalized) is adequately described in the in the text, but not evident in SI Fig. 2. One is forced to assume that CO2 fluxes from each pool are equal to 1-CUE, and therefore the same for C losses out of each pool. But this should be clarified in the description of the model and it’s wiring diagram.

  I’m used to seeing plots like Figure 1a with the axes reversed, since here we’re interested in how the model (independent variable) can predict observations (dependent variable).

  Figure 1b is nearly unintelligible. Is this showing the 3 dimensional parameter space for the optimized parameters to generate steady state SOC pools in Fig 1a? The legend says that colors are described in Fig 2, but no description is provided there- forcing readers to assume that colors represent different ranges for the fraction of C allocated the inert C pool (Fig. 2c)? The one relevant finding one may draw from this figure is that turnover of the humic pool (rdhum, which I would suggest calling k_{humic}) is inversely related to the fraction of C allocated to the inert pool (fintert). This apparent covariation, however, is never discussed (e.g. section 3.3).

  It’s unclear how the spatial distribution of the uncertainty analysis (Fig. 3) adds to the story being told here since it’s never discussed in the text (section 3.2). As such does the map of individual study sites and their magnitude of SOC change communicate much? If not, maybe these projected results (and uncertainties) could just be added to SI Table 1, along with observed, optimized SOC pools?

Since Fig 4b is discussed before Fig 4a (section 3.3) can these panels be switched?

References:

Hawkins, E., and R. Sutton, (2009) The potential to narrow uncertainty in regional climate predictions. *Bull. Amer. Meteor. Soc.,* **90,** 1095–1107.

Exbrayat J F, Pitman A J and Abramowitz G (2014) Response of microbial decomposition to spin-up explains CMIP5 soil carbon range until 2100 *Geosci. Model Dev*. **7** 2683–92

Fontaine, S., et al. (2011) Fungi mediate long term sequestration of carbon and nitrogen in soil through their priming effect. *Soil Biology and Biochemistry*, **43**, 86-96.

Sinsabaugh R L, Manzoni S, Moorhead D L and Richter A (2013) Carbon use efficiency of microbial communities: stoichiometry, methodology and modelling *Ecology Letters* **16** 930-9

Todd-Brown, K. E. O., et al. (2013): Causes of variation in soil carbon simulations from CMIP5 Earth system models and comparison with observations, *Biogeosciences*, **10**, 1717– 1736, doi:10.5194/bg-10-1717-2013, 2013.