

Review for Laub et al.

A robust DayCent model calibration to assess the potential impact of integrated soil fertility management on maize yields, soil carbon stocks and greenhouse gas emissions in Kenya

The study uses a Bayesian calibration approach (sampling importance resampling) with leave-one-site-out cross-validation to calibrate the biogeochemical model Daycent to yields, biomass and SOC at four sites in Kenya. The authors addressed adequately the suggestions of previous reviewers and the community comment and improved the quality of the manuscript. Overall, the manuscript is well-written, methods are sound and described sufficiently. I have some suggestions and comments (see below). I suggest to publish the manuscript after minor revisions.

I refer to the track changes version with my line numbers.

Abstract

L33: Daycent is well-suited to estimate the impact of ISFM
The impact of ISFM on what? -> Please add.

Introduction

L82: so a propagation of errors is possible in upscaling exercises
We can be sure the errors propagate in upscaling exercises even if you don't track them, you probably mean: So an **estimation** of uncertainties is possible in upscaling exercises

L103: ISFM can.... but at the same time **mitigate CO2 emissions due to the mineralization of SOC**

That's an ambiguous formulation, please rephrase to an unmistakable sentence.

L105: displaying the confidence in model parameters by Bayesian calibration
Not clear what you mean by that

Methods

L253: ,, taken calculated with the equation
- Typo, remove 'taken'

L495: in CO2 eq kg-1 maize grain yield
- in kg CO2 eq kg-1 maize grain yield

Results

Figure 2

My visual impression is that prior and posterior distributions are quite similar.
Why is the posterior less narrow in Figure 2 compared to the prior? Wouldn't one expect the calibration to constrain the parameters and give a narrower posterior compared to the prior?

Figure 2 caption: Not clear what you want to say by 'uncertainty-based Bayesian model calibration', but since this is not a term generally used or a method description, I would leave out the term 'uncertainty-based'.

Figure 7: 'the black solid line the simulation by the best parameter set for each site'
You did not calibrate by site, but the caption can be understood as if you did. Since the panels are per site anyway, I would recommend to omit 'for each site' here in the caption.

Figure 8: Credibility intervals for cumulative fluxes are quite narrow, and do not cover the 1:1 line. Are these really credibility intervals? Unlike the other figures, N₂O was not calibrated. I think they are quite misleading here, since N₂O was not included in the calibration so of course they remain narrow if you put narrow posterior distributions. Or is it variance that is displayed? Please add explanation in the caption.

For claiming that the posterior distributions are suitable for upscaling this must also be true for N₂O, while my view for N₂O a realistic uncertainty estimate is not shown.

Which ISFM method is simulated with highest accuracy etc?

If you target a robust fit for upscaling the effect of different ISFM methods, then it might be worth presenting the bias and rmse per treatment across site.

Figure 9: Please explain 9b in the caption (Mention 9 a b c in the caption.)

In several table & figure captions you explain the lowercase letters:

Same lowercase letters indicate the absence of a significant difference in XYZ

Easier to read would be a positive formulation: Different lowercase letters indicate a significant difference in XYZ between ...

Discussion & Conclusion

These sections make sense to me and I have no further comments.