The paper describes the capability of DayCent model to simulate yield and SOC development of the different ISFM practices in SSA and its improvement after cal-val. So as presented, the paper is quite long and verbose, resulting quite hard to follow. The figures do not follow a chronological order and are often hard to interpret (see fig. A5). While authors report in M&M a wide description of parameters selection and initialization values which is appropriate and detailed, results are not very clear, often reporting average data which do not highlight the model's ability to reproduce the different selected managements. Also, the mismatch in N2O simulations make hard accounting the GWP here reported. Based on these premises, I recommend a major revision before to be acceptable for publication.

Comments:

L118: …CH oxidation4. Typo.

L241-243: As authors state, DayCent needs to initialize the SOM pools to equilibrium using the typical input of biomass of the native vegetation. However, simulating native vegetation in SSA is not plausible since it is characterized by tropical evergreen forest, dry savanna and humid savanna that, with the only exception of savanna systems which was partly simulated in literature using the grass and tree layers, DayCent is not able to well simulate forest production (Gathany and Burke, 2012). Also, to my knowledge, DC was never tested over tropical environments. Authors should better explain what they used as vegetation for model spin-up.

L335: Authors should consider replacing the term GWP with GHG balance. Despite the likely low effect of CH4, the model is not able to predict CH4 emissions, that therefore they cannot be considered in the whole balance. In this context, would be better to define the GWP as GHG balance since, in any case, the contribution of CH4 cannot be measured neither excluded.

L338: Figure 1 is included in M&M, please move below in Results.

L390-393: Authors can remove this part since calibration is widely recognized to improve model performances.

L394: ….and for aboveground biomass for all sites except Machanga. You mean Aludeka?

L399-401: please, when cited into the main text, report the supplementary figures in chronological order (why A9 before A4, etc…?). Also, why fig.4-5-6 in paragraph 3.5? It's quite hard to follow this flow….

Major weaknesses:

a) In Fig. 3 authors reported all together sites and management for comparing not vs calibrated model. To my opinion, this representation of model calibration is misleading. Firstly, looking at the performances for each site (Fig. A9), model calibration only little improve the model performances found using default values, with statistics confirming the improvement is quite low and lower for each site compared to when assessed overall. This confirm that averaging all sites make unclear to evaluate the model performances under different conditions. Also, it is not clear the ability of the model to reproduce different type of management after calibration process (Fig. A5 is poorly readable, and statistics should be reported. From a visual analysis, variability seem not well simulated). So, from the whole study, does not clearly emerge how the model is able to reproduce yield and AGB for each ISFM at each site. This do not allow to discuss why model does or does not work at each site and for each

management, which could be the limitations and weaknesses, which should be the best practice to use and its response at each site. Averaging all yield data does not clarify the efficiency of the model to be suitable as tool to assess the potential of specific ISFM management practices (as stated by authors in introduction) to cope with food insecurity or further issues. Authors should revise all this part to provide a more accurate response to what they stated in the introduction.

b) The GWP discussion is another major point of weakness. Results clearly showed as N2O is not well simulated neither at daily scale (Fig. A10) nor as cumulated (Fig. 7). Despite in discussions authors state that simulated N2O emissions were generally reasonably well predicted with this current DayCent calibration, looking at Fig. 7 emerged as at Aludeka and Embu the measured N emissions were more than double than those simulated. This clearly affect GWP analysis, especially considering the role of N in GWP analysis, thus making these results very uncertain. Authors should exclude GWP analysis from this study or should much better calibrate the N response to better fit with observations, otherwise GWP discussion risk to be highly speculative due to low level of confidence in N emission outcomes.