Response to C. Muller (gmd-2016-29)

I have, however, substantial concerns with respect to the validity of this "validation paper". First of all, I'm expecting that the MATCRO-Rice model is going to be applied at river catchment or continental to global scale. Even though the envisioned scale of future application is not explicitly mentioned, I'm assuming so, as the model development is motivated by the wish to better study the effects of agricultural production systems (here: rice production) on hydrology and climate. Yet, the validation only provides a comparison to a single site, which even lacks a central data element (crop yield) that had to be deduced by trend extrapolation. The model validation presented then turns out to be a demonstration of model calibration to a single site and then can reproduce much of the observed dynamics at this site. Yet, it remains unclear how the model would perform at sites where such intensive calibration is not possible for the lack of data. The authors seem to constrain their "validation" to this one site as it seems to be the only eddy flux measurement site for rice production systems, but clearly that is no proof of model skill. As suggested by the other reviewer, I would much appreciate if the improvements of the MATCRO-Rice model could be evaluated against the original MATSIRO simulations and if there was an evaluation of model skill apart from the calibration site. Authors are advised to consult e.g. Luo et al. (2012) for possible data sets and metrics and Iizumi et al. (2014) for yield data.

There seems to be a misconception on the net carbon flux between land and atmosphere. The authors claim that the models ability to reproduce the biomass accumulation is an indicator for its applicability to simulate the net carbon flux: "As indicated by the figure, the simulated total biomass was in good agreement with the observations. Hence, we conclude that the model has

The paper as presented here addresses an important in land surface and agricultural modeling but fails to validate the model or to evaluate the model performance. Model performance needs to be evaluated against independent data sources, the improvement compared to the original model should be quantified (and eventually assessed against possible shortcomings) and the validation/evaluation should be performed at the scale of envisioned application, not just at a single point. Also, the model evaluation should address the ability to reproduce spatial and/or temporal patterns.

Unless substantially extended to justify a publication on its own, this paper could well be merged into the model description paper at http://www.geosci-model-dev- discuss.net/gmd-2016-28/, as also suggested by a reviewer there.

 guideline of GMD admits the separate submission of the model description and evaluation papers, if the evaluation is extensive. Therefore, we will separately submit the revised manuscript of the model description and evaluation papers.