

Geosci. Model Dev. Discuss., referee comment RC1
<https://doi.org/10.5194/gmd-2021-218-RC1>, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on gmd-2021-218

Anonymous Referee #1

Referee comment on "Nested leave-two-out cross-validation for the optimal crop yield model selection" by Thi Lan Anh Dinh and Filipe Aires, Geosci. Model Dev. Discuss., <https://doi.org/10.5194/gmd-2021-218-RC1>, 2021

The authors show the value of a two out of sample approach for statistical modelling. The formal approach is convincingly motivated and presented. The substance of the modelling, however, is completely inadequately presented. One would expect a brief characteristic of the modelled crops and their cropping specifics (annual, perennial; sensitive phases for weather dependence), a characterisation of the climate at the site considered, information about the crop relevant weather variability in the periods considered, a motivation of the predictors, and the margins for determining the predictors, a discussion of model errors. The superficiality of the problem analysis becomes particularly clear in the classification of the model quality:

'It was shown that monthly mean precipitation and temperature could explain more than 30 % of the coffee yield anomaly variability. The 70 % remaining variability is due to non-climatic factors (agricultural practices, diseases, or political and social context). '

This conclusion presupposed that the authors had fully explained the weather-related yield variability with their approaches. However, this cannot be assumed given the selection of climate variables, their coarse resolution, the static view on the underlying processes, the negligence of their interplay, the linearity restrictions applied, the exclusion of spill over effects from previous seasons and of soil characteristics.

Nevertheless, the presented approach shows which formal possibilities exist to obtain a 'best' model without an in-depth examination of the object to be modelled. This is quite interesting for practical statistical modelling. However, the examples presented here (maize, coffee) should be motivated much more comprehensively and classified in a more differentiated way. This does not mean that the authors have to make a comprehensive modelling claim. But they should be able to justify and classify the range of the model configurations considered in their modelling attempt.

The general quality of the manuscript is currently at the border between 'fair' and 'poor', but has potential for moving to the better direction when the authors motivate more explicitly their modelling approach for the selected climates and crops.