General comments

This is a revision. The paper addresses the very important question of model complexity. How can one best choose the level of model complexity when the objective is to minimize error of prediction for out of sample cases. The paper presents two practical cases (prediction of coffee yield in Viet Nam, prediction of maize yield in France) and two prediction methods (linear regression, artificial neural network). The main objectives of the paper are to explain the leave two out cross validation approach (LTO) as a method of evaluating and choosing between models of different levels of complexity, and to compare it with the leave one out cross validation approach (LOO). The presentation of LTO is useful and interesting. The review of the original manuscript pointed out some important problems with the specific cases used to illustrate LTO. These problems have not been adequately addressed in the revision.

We thank the referee for his/her comments. These comments have been carefully considered, and modifications have been made accordingly.

Specific comments

The major problem is in the definition of model complexity. The linear models in this study have a fixed number of explanatory variables (3 for LIN3, 5 for LIN5), chosen from n candidate explanatory variables, where increasing values of n are tested. The main measure of model complexity in the paper is the number of potential explanatory variables. As pointed out in the review of the original manuscript, this is not a usual definition of model complexity. Normally, it is the number of explanatory variables in the model that is the measure of model complexity. Not only is the number of potential explanatory variables not the usual measure of model complexity, it in fact is a very problematic measure of model complexity. Using this measure of model complexity, one could have two identical models and conclude that one is much more complex than the other, if the explanatory variables were chosen from among a larger pool of candidate explanatory variables. That is, model complexity would no longer be determined by the model itself, but also by the exact history of how the model was developed. I think that to illustrate LTO as a way of choosing the best level of model complexity, the authors need to use a more accepted measure of complexity (i.e. the number of explanatory variables in the model, for a fixed set of potential explanatory variables).

We agree that the number of potential explanatory variables (i.e., potential predictors) is not a usual definition of model complexity. However, as described in our manuscript (Sect. 2.3), the list of potential predictors is not fixed, and thus establishing this number is a crucial modelling step. Especially, in agricultural applications, ***with a very limited number of samples****, it is inappropriate to consider a large number of predictors. This* *large number of predictors issue was also identified in previous studies (Ambroise and McLachlan, 2002; Hastie et al., 2009).*

In addition to the number of potential predictors, we did test several (accepted) measures of complexity: increasing the number of explanatory variables (i.e., inputs) for a fixed set of potential predictors) or considering a more complex neural networks model instead of a linear regression one.

In the revised version, we showed more results of the common measures of model complexity: the number of inputs, model types, the number of neurons in the hidden layer, as shown in Sects. 4.1 and 5.1. Also, we introduced the number of potential predictors as one factor that drives the model selection and model quality. We also used the term "model selection" instead of "model complexity" to avoid confusion for the reader and better present the problem.

References:

Ambroise, C. and McLachlan, G. J.: Selection bias in gene extraction on the basis of microarray gene-expression data, Proceedings of the National Academy of Sciences, 99, 6562–6566, https://doi.org/10.1073/pnas.102102699, 2002.

Hastie, T., Tibshirani, R., and Friedman, J.: Model Assessment and Selection, in: The elements of statistical learning: data mining, inference and prediction, pp. 219–260, Springer, 2009.