Reviewer #1 comments: Revision #2 of "Crucial but often neglected: The important role of spatial autocorrelation in hyperparameter tuning and predictive performance of machine-learning algorithms for spatial data."

Introduction

Remove references from the abstract as this is not necessary

Thank you, we removed the references.

line 5/6: Remove the "the" before soil science in accordance to the other examples. And I would prefer to leave the "of course" out. Sounds somehow misplaced

You are right, this is not needed here.

Is landslide modelling necessarily an ecological application? Why not mention a few biotic and abiotic examples like species distribution modelling but also modelling of e.g. plant or soil properties, plant diseases etc...

Thanks for spotting this. Adding more biotic examples fit better into the overall structure of this work.

We removed landslide modelling and added the following: “Ecological applications range from species distribution models \citep{Halvorsen2016, Quillfeldt2017, Wieland2017} over plant disease and soil type modeling \citep{heimDetectingMyrtleRust2018, brungardMachineLearningPredicting2015} to resource selection \citep{Baasch2010}.”

line 10/11: include a link here, e.g something like "One typical example for a spatial prediction approach in ecology is the detection of fungi infection on Monterey pines (Iturritxa et al., 2014)....

Thank you, this improves the reading experience. We added the following “One typical example for a spatial prediction approach in ecology is the detection of fungi infection on Monterey pines \ citep{Iturritxa2014}.”

Line 72: Rewrite "good performance results" to "robust performance results"

Rewritten to “To reach robust performance results with non-parametric models, their respective hyperparameters must be optimized.”

Main objectives: I think the brackets around the hypotheses are not necessary as the hypotheses are an important part of the introduction.

We removed the brackets and reorganized the hypotheses part as follows:

“\begin{itemize}

\item Comparison of the predictive performance of spatial and non-spatial partitioning methods. We assume that non-spatial partitioning methods will yield over-optimistic results in the presence of spatial autocorrelation.

\item Exploring the effects of (spatial) hyperparameter tuning for commonly used algorithms in the field of ecological modeling. We propose that optimal hyperparameter tuning has an substantial effect on model performance.

\item Comparison of the predictive performance of parametric (GLM, GAM) and non-parametric algorithms (BRT, RF, SVM, KNN). We assume that the predictive performance of non-parametric algorithms is substantially higher.

\end{itemize}”

Data and methods

line 106: change "used new predictors" to "introduced further predictors"

We followed your advice and changed it to “introduced additional predictors”.

Line 107: Make clear what the response variable is. It's not "Diplodia sapinea" but it's "trees infected by Diplodia sapinea or not"

This is correct, thanks for spotting this faulty wording. We changed the sentence to “In the present study we also introduced additional predictors (probability of hail damage at trees, soil type, lithology type, pH) to possibly enhance the predictive power of the trained models.”

Line 113: What do you mean by number(11)?

By this we referred to the number of variables in the dataset. We thought that this is clear due to the first part of the sentence. Apparently it is not and therefore we made it more clear. The sentence now reads as follows:

“It is representative for many other ecological datasets in terms of sample size (926), number of variables (11) and predictor types (numeric as well as nominal).”

Line 124 and also from 137 on: don't refer to the internal name of the variable if not necessary. Better: "Iturritxa et al. (2014) showed that the presence or absence of hail damage..."

This can be indeed confusing to the reader. Thanks for mentioning. We changed the sentence as follows: “\cite{Iturritxa2014} showed that the presence or absence of hail damage observed on trees explained best pathogen infections of trees in the Basque Country.”

Line 137: approx instead of ca.

We changed to “approx.”.

Discussion

Section 5.2.4. This sounds like the validation is very arbitrary. How is it possible that a algorithm is the best with AUC but worst with Brier. What is the reason for that and why is Brier more reliable?

Thanks for bringing this question up. It is a very interesting one to us as well.
And honestly, at the moment we have no solid explanation for it.

Answering it in sufficient detail would require a thorough comparison of both measures on datasets with different characteristics and across various algorithms.
Also, the characteristics of the measures would need to be investigated in more detail.
It might be simply related to the fact that AUROC is not a “proper scoring” rule (Gneiting 2007) and with these consequences might be in danger to produce biased results in certain benchmarking settings.
Reviewer #2 suggested this change to us and we were surprised to see the substantial changes for the GAM and GLM.

Even though the change in performance due to the new measure is a very interesting question, it exceeds the scope of this manuscript.

Gneiting, T., & Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. Journal of the American Statistical Association, 102(477), 359–378. doi.org/10/c6758w

The title

I think the title is misleading because it contradicts with the findings of the study. Yes Spatial CV is necessary for estimating the predictive performance but your results have shown that spatial hyperparameter tuning doesn't have an effect compared to non-hyperparameter tuning. So apparently it's not as crucial as claimed in the title. Maybe a question mark at the end of the title might solve the issue?

The title is very important and we therefore thank you for discussing this point. We also reconsidered again about the wording of “crucial”. Also, we decided not to focus on the word “spatial autocorrelation” as it is a bit vague what to expect. Finally, we changed to a more generic title which reads as follows:

“Hyperparameter tuning and performance assessment of statistical and machine-learning models using spatial data”.