

Review of “Crucial but often neglected: The important role of spatial autocorrelation in hyperparameters tuning and predictive performance for spatial data.”

General comments

I was reviewer #2 on the first version of this work. Overall, my positive review of the paper stands and I appreciate the authors revisions. I do have two major sticking points that I feel the author’s didn’t adequately address, one new major comment and a few minor suggestions.

In my previous review, my first major comment was:

- By using only a single data set, it is not clear how generalizable the results are. It seems that more than one data example should be used. I suggest framing the comparison using a common task framework (*sensu* Wikle et al. 2017). Heaton et al. (2017) provides an excellent example applying the common task framework to evaluate methods for spatial prediction.

The authors suggest that the focus of their work “is on the methodology (spatial tuning of hyperparameters, influence of spatial autocorrelation) rather than generalization across datasets” and provide many straw man arguments as to why they can’t make their study rigorous and complete (e.g., “there is only one dataset available” and “the processing power is limited (6 cores, 24 GB RAM)”) My first comment is that if the paper is really focused on methodology, why not just use, in addition to a real data example, a simulation study where “truth” is known? For example, it makes little or no sense to talk about “unbiased performance estimation” when you don’t have a gold standard (i.e., when truth is known). I think it is non-sense to say that the focus of this paper is on methodology (i.e., the description/development of novel methods), when clearly the authors are trying to evaluate which “methods” are best. Second, the reasons the authors list are trivial. For example, a modern laptop computer has 6 cores and 24 GB or RAM. Sure it may be a slight inconvenience to do a rigorous study, but surely it doesn’t take weeks to run (on a laptop) and if the author’s wanted they could use higher performance computing (e.g., a desktop or something like AWS; see http://www.louisaslett.com/RStudio_AMI/ for an easily accessible solution). Also, I don’t think saying “there is only one dataset available” is helpful. I am sure the authors could find another data set (e.g., take a look here <https://datadryad.org/>).

In my previous review, my third major comment was

- I suspect that the data used in this paper (e.g., Figure 1) are spatio-temporal (i.e., not all collected at the same time). Although spatial prediction is widely used in ecology, it seems the trend is towards making spatio-temporal predictions using statistical models and machine learning algorithms (e.g., Hefley et al. 2017).

The authors confirm that the data are spatio-temporal and collected over a 4 year time period. The authors response is non-sense (e.g., “all observations are unique in space, meaning that there is no spatio-temporal overlap.” and “due to the long-term average characteristic of most variables (e.g. temperature, precipitation, etc), the temporal aspect of the response variable becomes less important”). If the issue of spatial autocorrelation is such a huge concern when tuning hyperparameters (which is the main impetus for the authors work and even show up in the title!), then why wouldn’t temporal autocorrelation matter? I think the authors need to address the spatio-temporal aspect of their data example or just get rid of the data example and use simulated data in a spatial-only setting. As written, this paper is misleading.

Finally, on pg. 27 lines 540 - 543: The authors suggest that the predictive accuracy of parametric models (e.g., GLMs) with and without spatial autocorrelation structures is perhaps the same and that little research has been done on this topic. This statement is blatantly wrong. In most cases, a parametric model without an effect that accounts for spatial autocorrelation (e.g., a spatial random effect) is a special case of the spatial model (e.g., a spatial generalized linear mixed model is just a GLM with a spatial random effect). In almost all non-pathological

cases the spatial effect will increase the predictive accuracy. I suggest that the authors look at some of the formal statistical references in my first review. Overall, the quality of the entire paper could be improved if the authors paid some regard to the primary literature in statistics and machine learning rather than rely so heavily on the secondary (and often incorrect) ecological literature.

Specific comments

1. pg. 2 line 42: Can you give a citation for the difference between supervised learning techniques? For example, see <https://web.stanford.edu/~hastie/Papers/ESLII.pdf>
2. pg. 3 line 15: I think the assumption are “less restrictive” not “less important.”
3. pg. 7 line 109: “An uneven distribution of the binary response variable(25/75)” Can you please clarify? I have no clue what this means.
4. pg. 9. SVM has already been defined on a previous page.

References

- Heaton, M. J., Datta, A., Finley, A., Furrer, R., Guhaniyogi, R., Gerber, F., Gramacy, R. B., Hammerling, D., Katzfuss, M., Lindgren, F., et al. (2017). Methods for analyzing large spatial data: A review and comparison. *arXiv preprint arXiv:1710.05013*.
- Hefley, T. J., Hooten, M. B., Russell, R. E., Walsh, D. P., and Powell, J. A. (2017). When mechanism matters: Bayesian forecasting using models of ecological diffusion. *Ecology Letters*, 20(5):640–650.
- Wikle, C. K., Cressie, N., Zammit-Mangion, A., and Shumack, C. (2017). A common task framework (ctf) for objective comparison of spatial prediction methodologies. *Statistics Views*.