Reply to authors final response

Michel Bechtold, KU Leuven, December 20, 2017

I thank the authors for their kind reply on my comment and for bringing up the interesting reference on "misconceptions of spatial analysis" by Kühn and Dormann (2012).

I am writing on response to that to clarify the issue. I do not consider spatial auto-correlation as a problem 'per se'. And I agree that peatland locations are very much spatially correlated in reality. A good model should thus also show this property and authors' check of the spatial autocorrelation of the residuals is a good idea. But it cannot be used as a proof for a well-calibrated model alone. My concerns were about the **validation approach.** The cross-validation followed a fully random design and must, however, be clustered or blocky, to guarantee that validation data is independent. There is a new paper published by Roberts et al. (2017 in Ecography) that in detail explains the issue. "Random hold-out data are too optimistic and favour overfitted models" (Carsten Dormann, personal communication, Dec 19, 2017). I think this is a very important point we must all be aware of because machine learning applications are spreading and often used with little care in that respect.

Sincerely,

Michel Bechtold

References:

Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J. J., Schröder, B., Thuiller, W., Warton, D. I., Wintle, B. A., Hartig, F. and Dormann, C. F. (2017), Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. Ecography, 40: 913–929. doi:10.1111/ecog.02881