We thank the two reviewers for their positive and constructive comments on our manuscript.

We hereafter address the comments made by the reviewers. Our replies are given in italic.

Reviewer 2

Page 4730, Line 3: "spatially-structured" –> "spatially structured"

Page 4732, bullet item iii. Would it possible to add one or two sentences to clarify the underlying reasoning of the variance-explained test? From the authors' description, I can readily follow what is done in this test, but I struggle to understand why this is a useful way to estimate the correct h.

We will clarify the rationale of using the variance explained method to estimate h.
We simulate environmental variables with the same spatial structure as the variable influencing species composition. The species composition is therefore not related to the simulated environmental variable. The only possible relation is through a correlation with the variable of interest. We therefore compare the r2 of the transfer function with the r2 between the two variables. If the data were independent we would expect a maximum r2 of the transfer function close to the r2 between the two variables. Therefore transfer function r2 larger than r2 between the two variables is indicative of over-optimistic performance estimates.

Page 4735, Line 6: Regarding loess, the authors say that "shorter spans are expected to remove more local variance." Is this backwards? I'd expect a short span to remove less of the local variance.

With loess smoothing a short span (proportion of data considered in smoothing) results in heavier smoothing, closer fitting to the data. Hence a short span removes more of the local variance.

Figure 1: Question for authors. It seems to me based on this figure that for many datasets the variogram-distance method often considerably overestimates the optimal h compared to the other two methods. The authors suggest using both the variogram- distance and variance-explained methods, and choosing the smaller h. But in addition to that, would it be possible to roughly estimate by how much the variance-explained tends to overestimate optimal h? I ask because of the two methods suggested, the variogram-distance method is by far the easier one to run. The variance-explained method, in particular, seems like it might be very calculation-intensive. It can be run for MAT, probably the least calculation-intensive reconstruction method of all, but what if the test needs to be done for another reconstruction approach? Thus it might be helpful, in some cases, to be able to run the variogram-distance method only, and have some rule of thumb about how much the h is likely to be overestimated. Or does the relationship between the h suggested by the two methods vary too much between individual datasets to give any such guideline?

Yes indeed, the estimates obtained with the variogram range method are usually longer than the ones obtained with other methods. Unfortunately, the relationship between h suggested by the two methods is too uncertain to give guidelines on how much the variogram range method overestimates h.

Figures 2-3: There appear to be two sets of results for the sum of variogram ranges of 30. Is one set of results perhaps for another x value?

The two values of 30 are caused by two possibilities of getting a total variogram range of 30 with the nuisance variables used in this study: 5+25 and 15 + 15

Figure 5: I don't see this figure referenced anywhere in the text

We will fix this

We will also upload the code needed to use the variance explained method with the revised manuscript.