

Interactive comment on “Sensitivity to species selection indicates the effect of nuisance variables on marine microfossil transfer functions” by Lukas Jonkers and Michal Kučera

Anonymous Referee #1

Received and published: 16 October 2018

I have reviewed the manuscript presented by the authors and have found it to be well written, interesting and concise. The topic it addresses has indeed been generally overlooked by the wider community and there is no doubt that more work of this sort is necessary in the field of paleoclimatology.

They begin by ranking the species by importance. The method considered from Juggins et al. (2015) seems valid, but I wonder whether other methods could be considered and if the results would be significantly different; I imagine not so much. The authors then showed that for both MAT and WA, it is possible to obtain rather different paleo reconstructions when using a different number of species for the calibration of

C1

the model, even though those models had very similar error of prediction in the training dataset. This underlines that the prediction error in the training dataset is not by itself a sufficient metric for characterizing the uncertainty in reconstructions. This also applies to other microfossils. It would be interesting to analyze the spatial patterns of the prediction errors in order to identify the sources of variation related to species-pruning.

The manuscript leaves many questions unanswered, but I understand that not all questions can be answered in a single article. I will be looking forward for future work concerning the quantification of the species-pruning uncertainty, and the identification of non-climatic biotic and abiotic factors leading to the uncertainties.

I found the treatment of the authors satisfying and did not find any important flaw with the analysis presented. Therefore, I recommend that it can be published without major revisions, and below I include a list of specific points which should be taken into account to improve the quality of the manuscript, particularly many figure captions should be revised.

Minor corrections: Line 15: ‘information [about] all species’

Line 108: If they were not reported, I imagine they were not counted. I am not sure what the authors did, but for consistency, such species should then also be removed from the calibration dataset.

Lines 113: How many samples are kept for this part used to build the transfer function? I found the brief explanation hard to understand and had to read Juggins to understand. I would suggest rewriting this paragraph in a more straightforward way.

Line 118: I do not understand why the approach is repeated 10 times after the 1000 bootstrap. Couldn't the error estimate be obtained from the bootstrap estimates directly? Is it that the $\frac{1}{3}$ selection of species is the same for any given set of 1000 bootstraps?

Line 123: Does this mean you do not renormalize the species abundances after se-

C2

lecting a subset of species? If yes, then couldn't this lead to problems if for example we have two sites with a similar composition regarding the 'useful' species which are sensitive to say temperature, but one has a large amount of irrelevant species temperature-wise while the other has almost none.

Line 150: I would indicate the number of species in parenthesis for MAT (around 6 species) and for WA (around 9 species).

Line 201: Add comma after 'In WA'

Line 240: Is expatriation a synonym for sediment mixing or a different process? It is mentioned, but neither defined or discussed really.

Figure 1: Not sure that incremental is the right work, an incremental change could be large. Maybe "marginal" or a synonym would be more appropriate.

Figure 3: Could you define what years you are considering for the LGM. Even better would be to add a shaded background over that period which you averaged in the top row graphs.

Figure 4: The caption could be rewritten more clearly, it is not clearly stated that the difference is with the reconstruction with the full taxonomic resolution.

Figure 5: Is the prediction error calculated using the full timeseries or only the LGM as previously? Also, it might be useful to indicate the number of species used for the 1st point, does it start at $n=2$? I imagine that the increments of increasing number of species is simply 1, i.e. $n_{i+1}=n_i+1$. Cool figure.

Figure 6: Second sentence could be revised grammatically speaking.

Interactive comment on Clim. Past Discuss., <https://doi.org/10.5194/cp-2018-107>, 2018.