

We would like to thank the reviewer for their critical look at our manuscript and are grateful for spotting an unexplained mistake in our analysis (species ranking in the North Atlantic). We have repeated all analysis and will make the code publicly available. Whilst the species ranking for MAT in the North Atlantic is indeed different from our original submission, the conclusions of our study remain unaffected: the sensitivity to species pruning provides a useful means to assess the influence of other environmental variables in transfer function-based reconstructions. However, the now correct (and replicable) species ranking leads to changes in the discussion. Below we have copied the reviewer's comments and provided our response in red and proposed changes are indicated in blue.

With kind regards,

Lukas Jonkers and Michal Kucera

=====

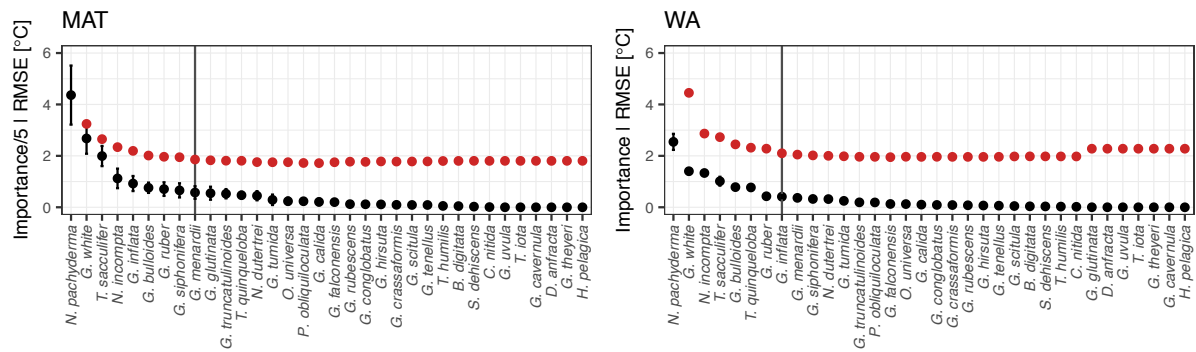
This manuscript uses the methods developed by Juggins et al (2015) to determine the importance of each planktic foraminifera for the performance of transfer functions calibrated against sea-surface temperature. These results are used to make models with a greatly reduced set of species which perform almost as well as models based on the full set, however, the reconstructions were different. The manuscript is generally well written.

I found some of the results surprising. In particular, the low importance of *N. pachyderma* with the modern analogue technique (MAT). This is an abundant (up to 100%) and common taxon with a clearly defined thermal niche. I found the explanation in the manuscript reasonably persuasive for transfer function performance, but not for importance.

Intrigued, I re-implemented the method and, with a smaller training set, and found *N. pachyderma* to be among the most important taxon for MAT. Subsequently, I was given access to the authors' code which gave the same result. The authors blame a glitch and now get the same result. I encourage the authors to adopt the techniques described in the British Ecological Society guide to reproducible code.

At least some of the subsequent results will be affected by this glitch.

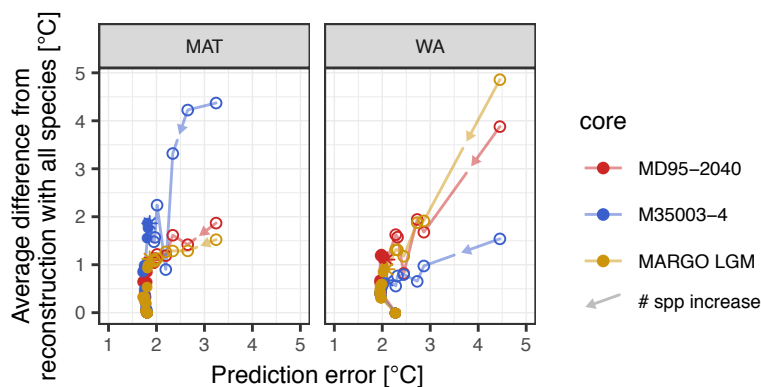
Repeat analysis showed indeed that the species importance ranking for MAT in the North Atlantic was erroneous. Our new analysis shows indeed that *N. pachyderma* is the most important species (see figure below).



This ranking has some consequences for the remaining part of the manuscript:

- i) it affects the number of species required to achieve a prediction error similar to the solution with all species;
- ii) the species importance ranking is now more similar for both methods;
- iii) it affects to some degree the reconstructed temperatures (Figs 3 and 4).

However, the most important finding of our analysis remains unaffected: we still find many different reconstructions with the same prediction error (see figure below)



Based on the prediction error alone, these reconstructions appear equally valid, thus highlighting a previously unrecognised source of uncertainty in transfer function based environmental reconstructions. We will update the manuscript to reflect these changes.

Despite extensive efforts, we were not able to reconstruct where the error occurred. It probably resulted from using erroneously an earlier version of the analysis, or an issue with file naming. However, the error only affected the one analysis and after sharing the code between us and the reviewer, we could establish that the new ranking is replicable. We apologise for this mistake and applaud the reviewer for their stringency.

The manuscript gives some consideration as to why some taxa are not important, but it would be interesting to see this expanded. I can think of several reasons. Taxa with low importance might be nuisance taxa, these should have negative importance. Taxa with poorly defined niches, or broad tolerances will have low importance and may be good to exclude from transfer functions. In contrast,

taxa with low abundances or occurring in only a small number of sites may have low importance as the analysis will give these low weight even though they might be valuable indicators. It would be useful to try to quantify how much these factors contribute to low importance.

The reviewer points out an interesting aspect of our study and given the comments by reviewer 1 we expand the discussion on this topic. We agree with the reviewers' categorisation, although we would like to point out that we did not find any strict nuisance species with a negative importance. We assume that a species' importance for transfer function is dependent on i) its abundance, ii) its niche width and iii) its sensitivity to the environmental variable that best predicts the entire assemblage. To assess how these factors influence species importance we calculated average and maximum (99th percentile) abundance (in %), the thermal niche width (in deg C) and the temperature sensitivity (expressed as the goodness of fit (r^2) of the species abundance to a Gaussian curve) for each species and assessed how well these parameters explained species importance using linear regression with increasing number of variables. We acknowledge that some of these variables are autocorrelated (e.g. temperature sensitivity and abundance are positively correlated, suggesting that the distribution of rare species is not primarily governed by temperature (in agreement with their low importance), or that our metric of temperature sensitivity is inadequate), but we believe this is of minor importance for the exploratory analysis we carry out.

For all ocean regions and for both MAT and WA we find that abundance appears to explain most of the variance in species importance. For the North Atlantic for instance the r^2 of a linear model that only includes maximum abundance is 0.79 and 0.91 for MAT and WA respectively. Addition of niche width raises the r^2 slightly to 0.83 and 0.93 and further addition of temperature sensitivity has a negligible effect (r^2 0.84 and 0.94). This indicates that species abundance is the most important factor controlling species importance. This is probably due to the simple fact that abundant species are consistently present and allow best definition of thermal niche. We agree with the reviewer that rare taxa with a well-defined narrow niche might be good indicator species, but (if they exist) their usefulness for transfer functions is limited by their low abundance, leading to inconsistency incidence. We will add the following to the discussion:

We will add this discussion to the section where we consider species importance:

To understand why some species are more important than others, we consider their overall maximum abundance, the width of their thermal niche in the training set and their temperature sensitivity as potential predictors of importance. We define temperature sensitivity based on how well the species abundance in temperature space can be described by a simple Gaussian curve. This analysis reveals that abundance (Fig. 2) is the best predictor of species importance (Table 1). Indeed, multiple regression models that include all three variables perform only marginally better than a

model using abundance alone. However, we note that all three variables are correlated to some degree. Interestingly, abundance and temperature sensitivity are positively correlated ($r = 0.84$), implying that the thermal niche of abundant species is better defined compared to rare species (Fig. 3A). We also observe that temperature sensitivity and thermal niche width are correlated. Counterintuitively, this correlation is positive: species with a narrow thermal niche appear less temperature-sensitive ($r = 0.60$; Fig. 3B). We attribute this pattern to a combination of low abundance of species with narrow thermal niches (sensitivity being correlated with abundance) and the possibility that their distribution is not primarily governed by temperature, assuming that the narrow thermal niche may be an artefact of adaptation to specific oceanic regions or regimes, only secondarily correlated with temperature.

Whereas importance can be evaluated objectively, the effect of species inclusion rules on reconstructions cannot be as the truth is not known. Adding more taxa to the reconstruction will obviously tend to make the reconstruction more similar to the all-taxa reconstruction. But any individual taxa could make the difference larger. Contrary to line 192, I do not regard the increase in difference with the inclusion of *G. glutinata* as evidence that it is a nuisance taxon. It might be more powerful to run this analysis on the calibration set.

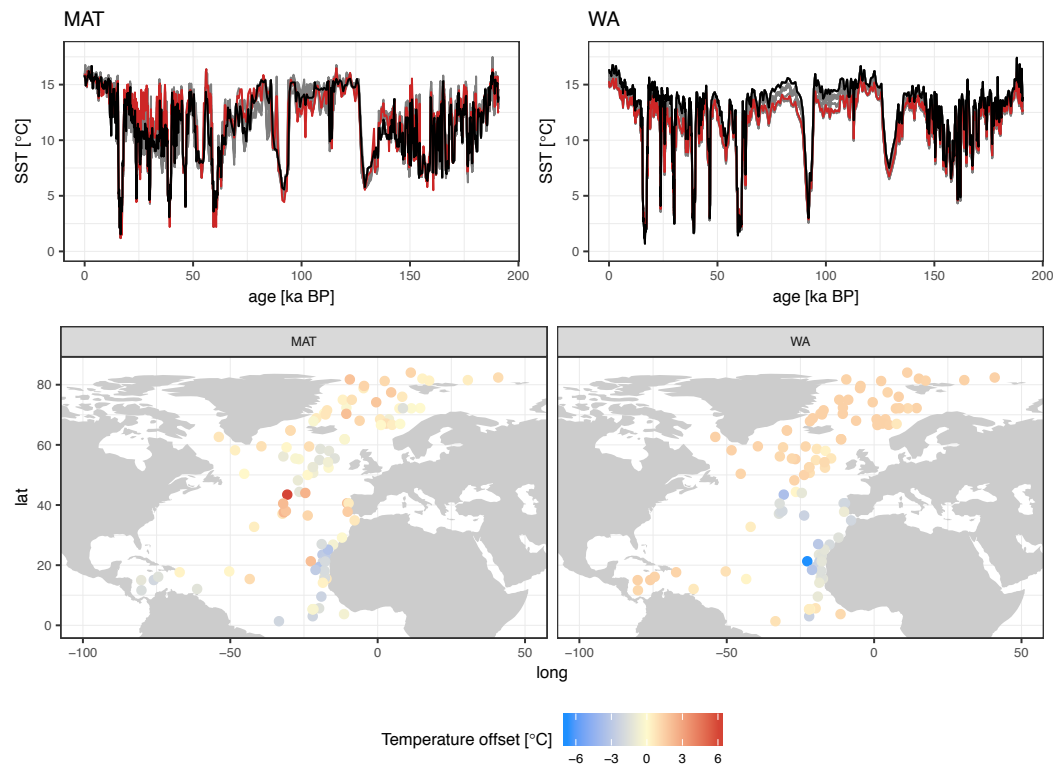
*We agree with the reviewer as was reflected by our statement “The analysis also reveals that there are many uninformative species, but very few – if any – real nuisance species...”. The reason why we assigned *G. glutinata* a possible nuisance status is for two reasons i) inclusion of this species in the calibration dataset leads to an increase in the prediction error (Fig. 1) and ii) inclusion of this species in the reconstructions leads to a jump in the difference in reconstructed temperature (Fig. 4). Our original submission already included this analysis, as reflected in the sentence “*G. glutinata* in the case of WA, which leads to an increase both in the prediction error and the difference of the reconstruction, supporting its potential rating as a nuisance species”.*

I am not sure we need to see the reconstructions with in 3A with fewer than the "minimum number" of taxa, as no one should be using these.

We agree. These will be removed from the figure.

The bias observed in for MAT in figure 3B looks much like what I would expect given *N. pachyderma* had been omitted.

This figure (now figure 4) will be changed to reflect the revised species importance ranking. The patterns for MAT and WA are now more similar:



It would be interesting (line 210) to compare the sensitivity of WA and MAT on the same range of models to identify the inherent variability of each.

The revised species importance ranking yields a top ten of most important species that overlap by 90 % (see figure above). The sensitivity of the results to the species importance ranking (the inherent variability) is therefore now (using the new ranking for NA) to a large extent assessed. We also note that the uncertainty in importance of the individual species are rather large, rendering the ranking of the unimportant species more uncertain. We see therefore little merit in repeating the analysis with 'swapped' species rankings.

Line 268. "that reconstructions that are highly sensitive to species pruning may indicate that the observed assemblage changes cannot be attributed solely to the environmental variable that is to be reconstructed"

This is an interesting idea, but I don't think this conclusion is justified given the results in the current version of the manuscript.

Our updated results still reveal the same patterns in the reconstructions and we find it hard to identify an alternative explanation, so we prefer to uphold the statement. We hope the revised analysis in the discussion will help to convince the reader that our conclusions are justified.

Whereas the manuscript demonstrates that only a few taxa are important for temperature reconstructions, I would hope that micropalaeontologists continue to count the full assemblage so that a range of questions can be addressed with the data. Only for routine analyses (for example water quality monitoring) is identifying only the important taxa justified.

We agree with the reviewer. Our intention was not to encourage counting fewer species, but to quantitatively investigate the effect of species pruning. This was motivated in part because some legacy datasets include only a limited number of species and we wanted to assess if these could be used. We will include an explicit statement in the revised version to encourage scientists to continue working with complete taxonomic resolution.

Minor points

Line 176. Celsius not centigrade

Addressed.

Capitalisation of axis labels in all figures needs to be checked.

Will do.

Figure 2. Might be better to use `scale_fill_continuous(trans = "log", breaks = ...)` than to log transform the percent.

OK.

The authors report (line 312) that the code is available on request. It would be much better to archive the code on, for example, github, or better still a permanent archive such as zenodo.org, ideally before review.

We agree with the reviewer. Our code is now available on:

https://github.com/lukasjonkers/species_selection and will be made available on zenodo upon acceptance of our manuscript.