BUMPER, Holden et al, GMDD
Response to A. Parnell (Referee)

Referee comments in black.
Author responses in red.
Manuscript changes in green.

We are grateful for this careful review, which has certainly improved the clarity of both the model description and presentation of results.

This paper introduces a Bayesian model (nicely acronymed 'BUMPER') to reconstruct palaeo-environmental variables from various proxies given modern training data and fossil samples. The model contains two neat ideas which I haven't seen before, namely that of using a mixture likelihood to model both abundance and presence/absence, and the idea of scoring training sets according to their richness and diversity.

I do have concerns about the mathematical model and the way it is described. As somebody who lives and breathes these types of models I found the mathematics confusing and my guess is that they will go straight over the head of the average reader of this journal.

We believe the confusion was most likely due to misleading notation in Eq 1 and Eq 2, and we are grateful to the referee for pointing this out. We have corrected this as detailed below.

- Starting with Equation 1, it seems to be missing a product term, which I think should appear after the proportionality constant. Either that or each SRC is being calculated for each taxa, species and site combination. This seems unlikely.

Our approach in Eq 1 was an attempt to describe how the SRC probabilities are progressively refined by sequential consideration of each training site. i.e. we were describing the calculation of prob(SRC_jk| y_ik, x_i), not prob(SRC_jk| Y_k, X). The accompanying text used to read "*Equation 1 is applied sequentially across all training-set sites. At each application, the posterior derived from the previous training-set site becomes the prior for application to the next site, so that the probabilities assigned to each SRC become progressively better defined.*"

However, we agree this has resulted in imprecise and confusing notation. As suggested by the referee, we have revised Eq 1 to express the product explicitly

$$prob(SRC_{jk}|Y_k, X) \propto prob(SRC_{jk}) \times \prod_i prob(y_{ik}|SRC_{jk}, x_i)$$

and have modified the text accordingly (the descriptive text in italics above is no longer required).

- Equation 2 seems to suggest that this is the normalising constant, but that can't be the case as the Bayes equation is p(SRC|y) = p(y|SRC)p(SRC)/p(y). It's p(y) that needs to be in the normalising constant.

Apologies, this is a consequence of the same confusing notation, i.e. using prob(SRC_jk) interchangeably for the prior and the posterior. We have corrected this to show that it is the posteriors that are normalised using the constraint:

$$\sum_j prob(SRC_{jk}|Y_k, X) = 1$$

After this point we revert to the prob(SRC_jk) notation, with clarification as suggested by the other referee, Cajo ter Braak. "From now on $prob(SRC_{jk})$ is the posterior probability, the probability of the SRC given the training data Y and X."

- Equations 3 and 4 suggest that the likelihoods are all only known up to proportionality and the proportionality component isn't mentioned. I think these should all be equals signs.

We have clarified this with additional text. We prefer to use normalised likelihood functions (and now name them accordingly) so that we can regard them as pdfs. In effect we are describing the reconstruction that would be derived from a single species with a uniform prior.

To reconstruct the environment from an observed fossil count y_{k0} of each taxon k within the fossil assemblage, the probability-weighted SRCs are used to derive likelihood functions for each taxon of the reconstructed variable x. Here we again use the Bayes relationship, this time stating that the probability that a reconstructed value is correct in the light of an observed species count is proportional to the probability that the species count would be observed in that environment. Considering a single observed species in isolation, Bayes' equation can be written:

$$prob(x|y_{k0}) \propto prob(y_{k0}|x) \times prob(x) \tag{3}$$

We derive a normalised likelihood function for the taxon, considering only the first term on the right hand side of Eq. 3. This allows us to treat the function as a probability distribution of the environmental variable, given no prior knowledge and a count of this single taxon in isolation. As we do not know the true SRC of the taxon with certainty (in general, the calibration will have resulted in many SRCs with non-zero probability), the likelihood function is derived from all significant SRCs, combined using the probability weights calculated in Section 2.1, as follows:

$$L_y(x|y_{k0}) \propto prob(y_{k0}|x) = \sum_j \{prob(SRC_{jk}) \times prob(y_{k0}|SRC_{jk}, x)\} \tag{4}$$

This expression is evaluated at 100 evenly spaced points across a range that comfortably spans the training set environmental range (see section 2.4), and is normalised to 1. It defines the probability density function of the environment, given no prior information and some observed count y_{k0} of species k.

And, at the end of section 2.4

"We note that the indicative tolerance is also used to define the range of environment considered in the reconstruction (Section 2.2), from $(x_{min} - 6t')$ to $(x_{max} + 6t')$. Significant probabilities beyond this range are unlikely given the constraints imposed upon the optima and tolerances. In any event, as with any transfer function, the model should not be applied under suspected extrapolation far beyond the training set environment."

- Equation 6 suggests that there is another Bayesian model being fitted. It's thus not clear whether there is one model being fitted (which is all that is required) or whether multiple Bayesian models are being stitched together.

Separate models are fitted for all species k (Eq 1). Equation 6 (now Eq 7) describes how the likelihood functions generated from all considered species are combined to with a prior for the environment to generate the reconstruction. Additional text to clarify

"The posterior probability distribution for the reconstructed variable is derived by combining any prior knowledge with the product of likelihood functions of all considered species in the assemblage, as follows:"

All this points to a more fundamental problem, namely that of the lack of a statistical collaborator. These authors are world-renowned experts in the field of collecting and understanding the nuances of proxy data and how it links with climate. There are statisticians and groups out there (for example the Past Earth Network) who can help.

Yes, we agree that an experienced Bayesian collaborator would have been useful. However, we hope we have convinced the reviewer that problems were only in the exposition, and have been satisfactorily addressed.

Models like these are now being studied by statisticians in collaboration with proxy experts. One that is not yet in the palaeoclimate literature (which is perhaps why the authors might have missed it) is that of Ilvonen et al which seems very similar to what the authors are trying to achieve here. A more flexible version can be found in Cahill et al which is in the palaeoclimate literature and uses multiple proxies (forams and d13C) in a Bayesian model for sea level reconstruction. A more recent model is my own Bclim (Parnell et al) which allows for joint inference (i.e. all fossil slices, all taxa, multiple climate variables) to be estimated together, with the aim of reducing uncertainty.

Thank you, we have added these references and accompanying text in the introduction:

"The field of Bayesian palaeoenvironmental statistics is rapidly developing. Recent work includes the development of a pollen-based multinomial regression model that assumes a Gaussian species response, with joint inference across core time-slices (Ilvonen et al 2016) and a foraminifera-based multinomial non-parametric response model that allows for multi-modal and non-Gaussian taxon response curves (Cahill et al 2016). Parnell et al (2016) have published an open source R package Bclim (Parnell et al 2015) that uses pollen response surfaces to generate a series of equally probable joint multivariate climate trajectories."

and

"though we note that Cahill et al (2016) incorporate a second proxy (δ^{13}C) through a prior, assuming a normal likelihood $N(\mu_i, \tau)$ with constant precision τ."

Lastly a note on the figures. Again I

Figure 1 has three lines on three panels

We do not plot individual SRCs, rather we

"Figure 1: Probability-weighted SRC parameters N_k (left-hand axis), p_k (left-hand axis) and P_k (right-hand axis) are plotted for all taxa. Three training-sets are considered: chironomid-based temperature (Matthews-Bird et al., 2016), diatom-based pH (Stevenson et al., 1991) and pollen-based temperature (Bush et al., in prep). The x-axes represent the distinct taxa in the training sets (59 chironomid taxa, 225 diatom taxa and 553 pollen taxa). For each of the three SRC parameters, probability-weighted values are derived for each taxon. These are ordered by increasing value and are plotted sequentially. Horizontal gridlines represent the discrete values allowed within individual SRCs."

Figure 2 contains lines, candlesticks, points, crosses, and three different colours. I've read the caption multiple times, and the text associated with it (which covers 3 different sections), and still cannot work out what's being learnt from this picture.

We have added some additional text that attempts to summarise the main lessons of the picture.

"In summary (see section 3.2), this plot demonstrates: 1) Reconstruction errors of WAPLS1 (circles) and BUMPER (crosses) are similar for all assemblages. 2) Increasing the sampling density of a training set reduces both the reconstruction errors (circles and crosses) and the BUMPER reconstruction uncertainty (solid lines). However, continued benefits beyond a sampling density of ~10 are modest. 3) Reconstructions from assemblages that benefit neither from high richness nor from low tolerances (high species turnover) are associated with significantly greater error and reconstruction uncertainty. We note that the overstatement of BUMPER uncertainty relative to the reconstruction error (solid

lines compared to crosses) is expected for this application to idealised data (see Section 3.2.3).
"

Additionally, we have expanded the text in Section 3.2 in an attempt to better explain these data in detail.

Figure 3 is much more useful, but seems to be hardly mentioned aside from the end of the last paragraph of Section 4.

We now refer to Fig 3 where relevant in the text (sections 4.1, 4.2 and 4.3) and have added some additional text in section 4:

"Notably, Figure 3 illustrates the reconstruction-specific uncertainty Δ ($\pm 2\Delta$ is plotted), which in general differs significantly from the training-set RMSEP that is usually assumed to describe the uncertainty of WA-PLS approaches."

It was a shame not to see any actual reconstructions of climate over time for any of the sites.

We decided not to include paleo-reconstructions, but rather to focus solely on the transfer function performance, using both idealised and real training sets. BUMPER has been previously applied (albeit without automated priors) to temporal reconstructions. These have been compared with alternative reconstruction methodologies (Holden et al 2008, Matthews-Bird et al 2016) and also with observational data (Holden et al 2008).

References used:

Ilvonen, L., Holmström, L., Seppä, H., and Veski, S. (2016) A Bayesian multinomial regression model for palaeoclimate reconstruction with time uncertainty. Environmetrics, 27: 409–422. doi: 10.1002/env.2393.

Cahill, N., Kemp, A. C., Horton, B. P., & Parnell, A. C. (2016). A Bayesian hierarchical model for reconstructing relative sea level: from raw data to rates of change. Climate of the Past, 12(2), 525-542.

Parnell, A. C., Haslett, J., Sweeney, J., Doan, T. K., Allen, J. R., & Huntley, B. (2016). Joint palaeoclimate reconstruction from pollen data via forward models and climate histories. Quaternary Science Reviews, 151, 111-126.