Review: Combining a pollen synthesis and climate simulations for spatial reconstructions of European climate using Bayesian modelling

In general, the paper proposes a Bayesian filtering of climate model ensembles using pollen data. The ideas and presentation are interesting and useful, although there are some areas for improvement of both the statistical and computational methods. The filtering of climate to spatial locations with associated uncertainty is very important for secondary analyses (e.g. vegetative growth models, etc.) and provides a useful contribution to the literature. In what follows, I propose some refinements to the statistcial methods presented in the paper.

A few thoughts:

A) What is the meaning of the covariance $\mathbf{\Sigma}$? if it really is the inter-model variability, why mix over $\omega_k N(\boldsymbol{\mu}_k, \mathbf{\Sigma})$ instead of just using the distribution $N(\bar{\boldsymbol{\mu}}, \mathbf{\Sigma})$? What is the meaning of this distribution that your are mixing over? $\mathbf{\Sigma}$ is the covariance of the model ensemble, not a particular realization $\boldsymbol{\mu}_k$, thus the distribution you propose doesn't make much sense to me as a meaningful statsitical object. I'm open to being convinced that this idea makes sense, but right now I don't see the purpose.

   Also, using this mixing distribution is likely to eliminate the spatial autocorrelation in the $\boldsymbol{\mu}_k$ as neighboring locations are now no longer draws from the same climate model. Instead, it might be simpler and make more sense to mix over $\omega_k \boldsymbol{\mu}_k$. This mixing distribution is over the climate model ensemble and will preserve the spatial autocorrelations in the climate models. Ultimately, I'm not convinced that this covariance is what you want to model.

B) I would also suggest the title should be "... using Bayesian Filtering" rather than "Bayesian Modelling." You aren't modeling the climate, just assimilating the proxy data to filter the climate model ensemble. For instance, if the proxies suggest a temperature higher than all of the climate models, then the best your framework can do is the highest temperature from the climate model ensemble (plus a little error from $\mathbf{\Sigma}$). Hence, your models are only as reasonable as the climate models (which are likely not great estimates for a given time and location...).

C) The fit of the pollen data by a normal distribution is really poor. The fitted distributions in Figure 3 look nothing like the data distributions. Perhaps a better model is needed. In addition, if you are only using presences in the fossil pollen, your calibration is biased. It would be better to treat absences as a zero-inflated model where an absence could be a true absence or a missing presence due to non-climatic reasons. This is easily done by introducing a latent variable (like you did for the z). In the ecological literature on occupancy modeling, this is known as detection modeling (MacKenzie et. a. 2002)

D) Equation 13: If mixing over the z's is the problem, why not integrate/marginalize them out? You can always recover them using composition sampling later. It seems like an awfully complex computational framework (MC^3) for such a simple model framework that could be fixed simply by marginalization.

E) Maybe I missed it but what is the size of the model ensemble K and the number of calibration sites?

F) How to you evaluate the Brier score when you don't include the absences? This seems to introduce a bias and could make the Brier score improper (which limitis its usefulness in comparing models).

Journal Review points

1) Scientific significance: Does the manuscript represent a substantial contribution to scientific progress within the scope of Climate of the Past (substantial new concepts, ideas, methods, or data)?

Yes, the manuscript demonstrates the use of climate proxies in model ensemble filtering.

2) Scientific quality: Are the scientific approach and applied methods valid? Are the results discussed in an appropriate and balanced way (consideration of related work, including appropriate references)?

There are some questions about the implementation of the statistical model that are not completely resolved. (particularly the mixing distribution for climate doesn't make sense and the lack of absence data introduces bias in the estimates). The comments above can provide some guidance in resolving these issues.

3) Presentation quality: Are the scientific results and conclusions presented in a clear, concise, and well-structured way (number and quality of figures/tables, appropriate use of English language)?

The paper is reasonably well written from a technical perspective, although more motivation of why particular methods/equations are chosen would be useful. In other words, there is a lot written about *what* the methods are by not much about *why* the methods are chosen and what the ideas are trying to solve.

1) Does the paper address relevant scientific questions within the scope of CP?

Yes

2) Does the paper present novel concepts, ideas, tools, or data?

Yes

3) Are substantial conclusions reached?

Yes. A nice idea for Bayesian filtering is introduces

4) Are the scientific methods and assumptions valid and clearly outlined?

Not always (at least for the statistical methods).

5) Are the results sufficient to support the interpretations and conclusions?

Yes.

6) Is the description of experiments and calculations sufficiently complete and precise to allow their reproduction by fellow scientists (traceability of results)?

For the most part. A gitHub repository/code base would go a long way.

7) Do the authors give proper credit to related work and clearly indicate their own new/original contribution?

Yes.

8) Does the title clearly reflect the contents of the paper?

Yes, with a small change of emphasis

9) Does the abstract provide a concise and complete summary?

Yes.

10) Is the overall presentation well structured and clear?

Yes.

11) Is the language fluent and precise?

Yes.

12) Are mathematical formulae, symbols, abbreviations, and units correctly defined and used?

In general, yes.

13) Should any parts of the paper (text, formulae, figures, tables) be clarified, reduced, combined, or eliminated?

14) Are the number and quality of references appropriate?

Yes

15) Is the amount and quality of supplementary material appropriate?

I would like to see more done for reproducibility. The computational methods seem overly complex and making code availalbe for replication would be useful.

References

MacKenzie, D. I., Nichols, J. D., Lachman, G. B., Droege, S., Andrew Royle, J., & Langtimm, C. A. (2002). Estimating site occupancy rates when detection probabilities are less than one. Ecology, 83(8), 2248-2255.