Quantitative climate reconstruction from sedimentary ancient DNA: framework, validation and application

Response to comments of Anonymous Referee #1

Reviewer comment: (1) Pollen preserved is lake sediments have been used to generate quantitative reconstructions of past climate for the last 50 years, but two inherent problems severely limit the precision possible. First, species with ecologically distinct niches can be difficult or impossible to separate palynologically, secondly, some wind dispersed species produce vast amounts of well dispersed pollen that can blow into lakes far beyond the ecological limits of the species. This manuscript tests the potential of ancient DNA, which has high taxonomic resolution and limited dispersion, as an alternative to pollen for reconstructions. The manuscript tests different reconstruction methods, including traditional transfer functions that use a calibration set and an alternative method that uses presence-only data.

The methods used are:

- CREST based on GBIF occurrence data
- CREST based on GBIF occurrence data with MaxEnt preprocessing
- WAPLS using a calibration set
- MAT using a calibration set.

The first two methods are just applied to the aDNA data, the latter two are applied to both aDNA and pollen data. This may be the first time a head-to-head comparison of CREST and transfer functions has been done. It might be worth extending the analysis to run CREST with the pollen data as well for a more complete comparison.

Our response: We have now implemented the CREST analyses using the standard pipeline using the data sources in the crestr R package. Since this is not the focus and scope of the paper we included the results in the supplements.

New text: For completeness, we also ran CREST on pollen data using CREST’s built-in standard pipeline (see Appendix Fig. A3). The pollen-based CREST reconstruction shows similar long-term trends but less variance compared to the sedaDNA-based reconstruction.

New Appendix text:

Methods:

Using the CREST's built-in standard pipeline requires a proxy-species equivalency (PSE) table as fossil taxa are most often identified at a lower taxonomic resolution. Hence, all individual species available in the CREST database that are likely to have produced a certain fossil taxon need to be associated with their corresponding fossil taxon (Chevalier, 2022). We therefore determined the taxonomic rank for all harmonized fossil pollen taxa and indicated their respective level (1 = family level, 2 = genus level) alongside the taxonomic lineage in the PSE table. The reconstruction was conducted in the same way as with the plantDNA_PDF/SDM and plantDNA_PDF/GBIF methods, except that we used the proxy-species equivalency table instead of a customized distribution table.

Results:

The pollen-based CREST reconstruction shows similar long-term trends but less variance compared to the sedaDNA-based reconstruction. The assessment and potential optimization of the pipeline was not in the focus of this manuscript.

new Appendix Figures:

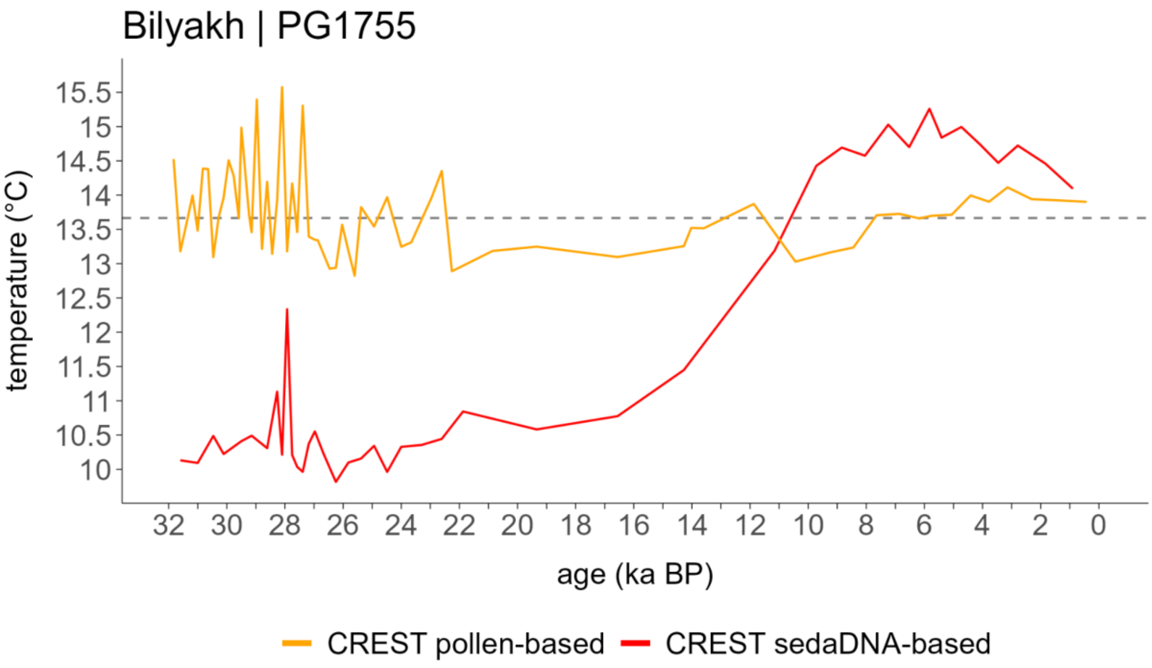

Fig. A2: comparison between sedaDNA- and pollen-based summer temperature reconstruction using the CREST framework.

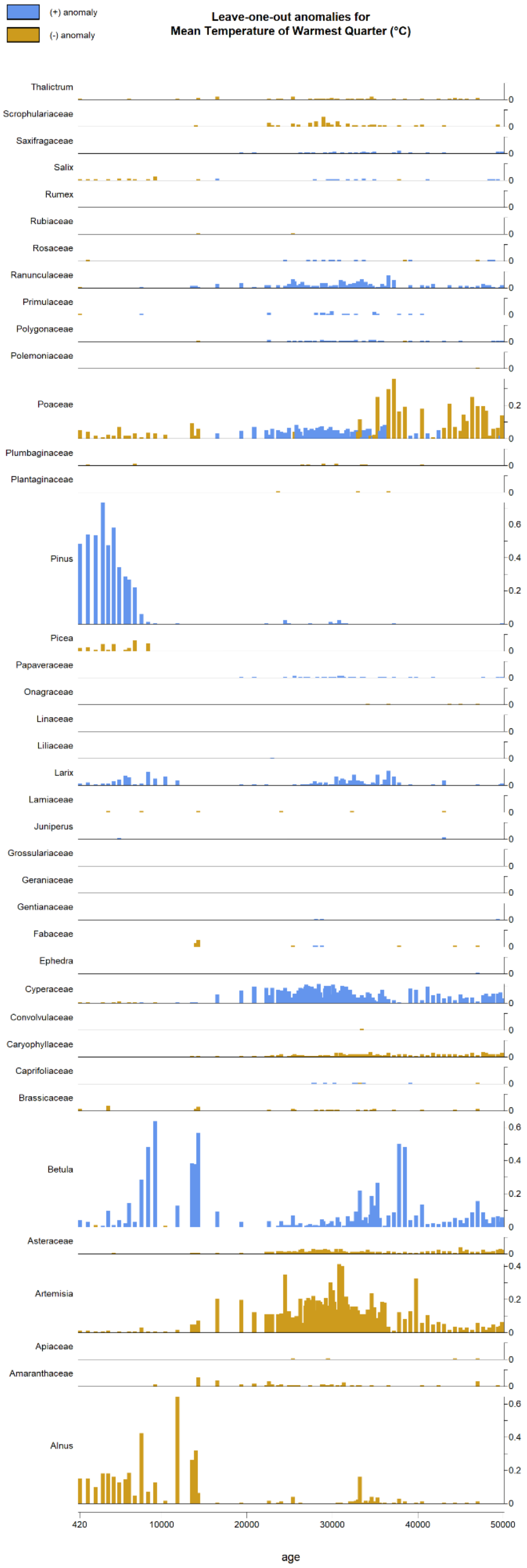

Fig. A3: influence on the summer temperature reconstruction using the CREST framework with harmonized pollen data inferred from leave-one-out analyses in the crestr R-package (Chevalier, 2022).

Remark to the reviewer on Fig. A3: we will revise the Figure and bring it to the same layout and color scheme as e.g. Fig. A1 when uploading the revised manuscript.

Reviewer comment: (2) The justification for using CREST with MaxEnt preprocessing is "that several taxa do not equally cover the temperature range". This is a rather vague justification, and I'm not exactly sure what is meant by it. The GBIF-MaxEnt-CREST pipeline is novel and rather involved. I recommend starting this section with a short paragraph that outlines the process so that the details are easier to follow.

Our response: We agree that it can be justified better i.e. revising the text to make clear that SDMs (using the MaxEnt model) help to provide a more extensive model prediction bases - basis for the CREST analysis. We revised the text accordingly.

New text: GBIF occurrences are especially biased in remote northern regions, providing highly valuable observations but limited to certain areas that are more easily accessible. Therefore, we developed species distribution models (SDMs) that predict the potential spatial range of the species and took a random sample within the range to derive a more robust temperature niche estimate (Appendix Fig. A1).

new Appendix Figure:

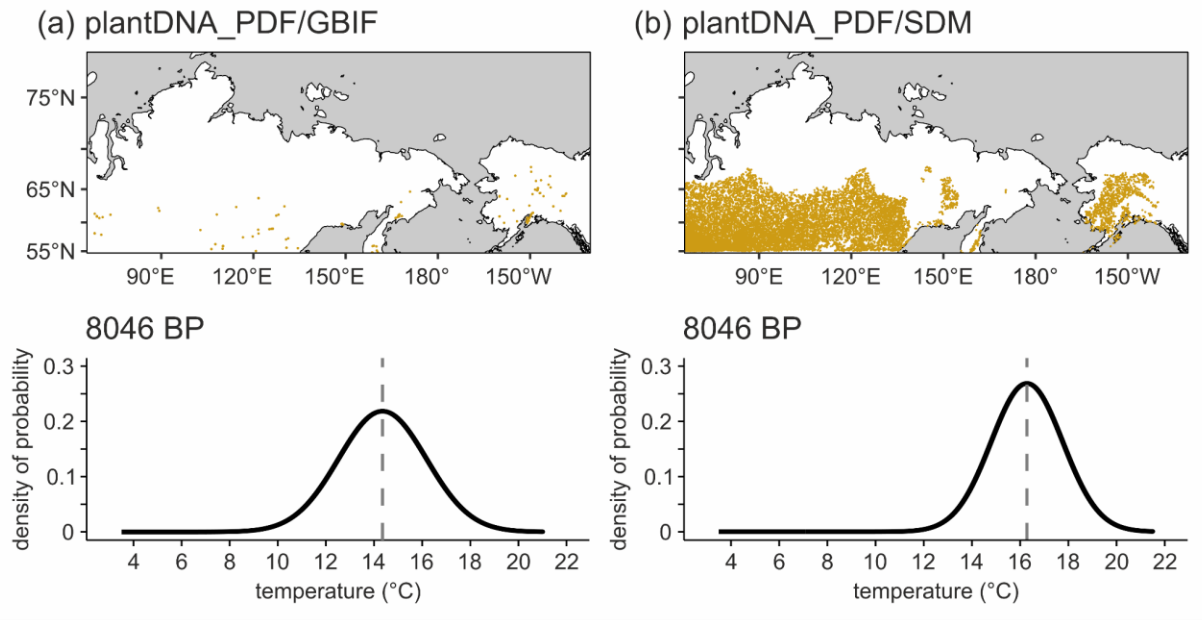

Fig. A1: PDFs fitted to *Alnus* occurrences when using (a) the original GBIF occurrences and (b) modeled occurrences derived from SDM to determine temperature niche estimates.

Reviewer comment: (3) Pre-processing with MaxEnt seems to improve the performance of CREST, but I do not understanding what exactly how MaxEnt helps CREST perform better. I'm dubious of the claim that it "enhance[s] the point density in the occurrence data" as it is not possible for the method to create data. Maybe a plot comparing the niches estimated by both methods would help explain what is happening.

<u>Our response:</u> The justification was indeed too vague. MaxEnt predicts continuous distributions across space. In short, the model is fitted by the climate niche (n-dimensional) derived from observations, enhanced by random background samples from regions of unlikely occurrences and predicts the climate suitability for the entire study area. This approach helps to reduce the sampling bias (due to human observed occurrences), fills gaps notably in under-represented parts of the climate gradient, leading to more realistic and robust PDFs. In addition to the better description (first paragraph of 2.2.1 - see also comment above), we added a new Appendix Fig. A1 alongside a description to the Appendix.

<u>New text:</u> GBIF occurrences are especially biased in remote northern regions, providing highly valuable observations but limited to certain areas that are more easily accessible. Therefore, we developed species distribution models (SDMs) that predict the potential spatial range of the species and took a random sample within the range to derive a more robust temperature niche estimate (Appendix Fig. A1). SDMs were implemented using the MaxEnt model (Phillips et al., 2006) from the R-package dismo (Hijmans et al., 2024).

<u>new Appendix Figure:</u>

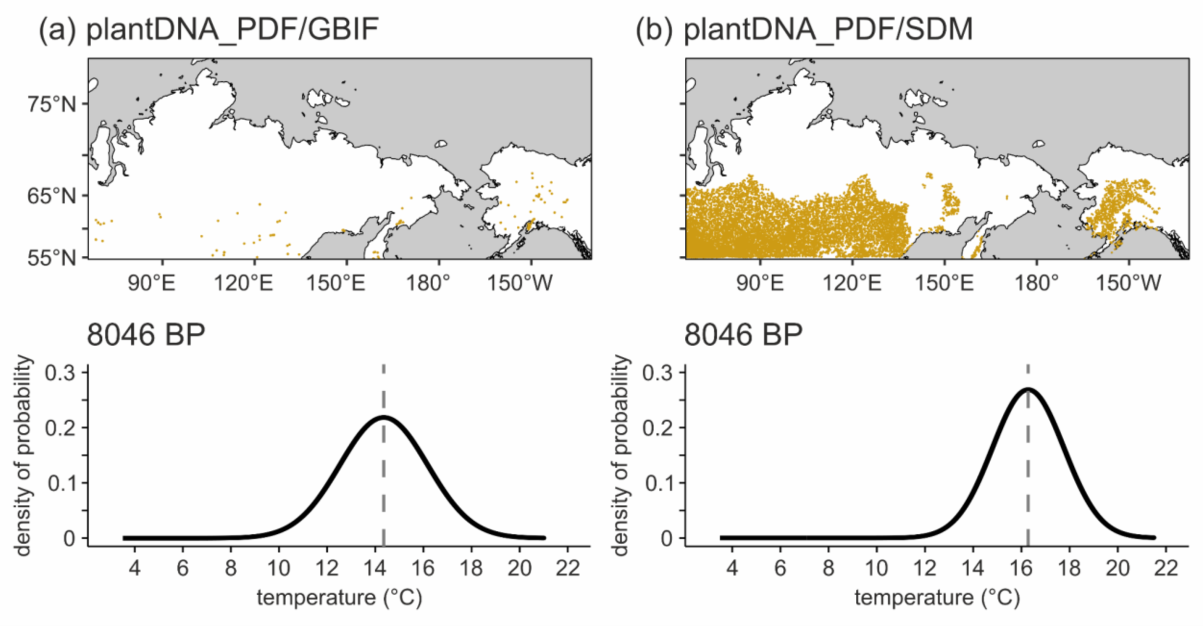

Fig. A1: PDFs fitted to *Alnus* occurrences when using (a) the original GBIF occurrences and (b) modeled occurrences derived from SDM to determine temperature niche estimates.

Reviewer comment: (4) I can easily imagine that the CREST is too constrained in the niche shape they can fit, and it might be profitable to allow more than normal or log-normal PDFs. So rather than pre-processing the data with MaxEnt, the first step of CREST is replaced by MaxEnt (or another flexible model). Of course, the penalty for using more flexible models is that they are prone to over-fitting.

Our response: We totally agree, different and maybe even more flexible models or distributions might offer advantages, yet with the potential for over-fitting. We added a sentence in the discussion to make readers aware of potential future developments and pitfalls.

New text: While our pipeline uses MaxEnt as preprocessing for CREST PDFs, an alternative would be to replace the PDF step entirely by a more flexible niche model (e.g. Maxent). This could further improve the niche shapes but may also lead to overfitting.

Reviewer comment: (5) There is an issue with the cross-validation of the transfer function models. The ms reports that the uncertainties on the reconstruction are calculated using bootstrapping, but it is unclear what cross-validation scheme is use to estimate the models' performance.

Our response: Added.

New text: Informed by results from the leave-one-out cross-validation and randomisation t-test, we decided to use results from component 1.

Reviewer comment: (6) One widely used cross-validation scheme is leave-one-out cross-validation. Somewhat confusingly, this ms, following Chevalier (2022), uses the term leave-one-out to refer to a type of sensitivity analysis in which taxa are left out sequentially. It would be better to call this step a sensitivity analysis.

Our response: We agree.

New text: These tools include graphical overviews of climate sensitivities for taxa, plots depicting individual sample PDFs and their reconstruction outcomes, and a sensitivity analysis (originally termed leave-one-out analysis in Chevalier (2022)).

Reviewer comment: (7) The ms emphasises the median bias of the reconstructed temperatures as a metric of transfer function performance. I have not seen this metric used before. Mean bias is sometimes reported, but not prominently as it can be zero even if the transfer function has no skill. Median bias will have the same unfortunate property. Maximum bias is more useful.

Our response: We actually used median absolute bias. We now clarify this and also report maximum absolute bias in Tab. 1.

New text: We report both median absolute bias and maximum absolute bias to better capture systematic offsets in the reconstructions.

Reviewer comment: (8) The pollen-MAT error for the Billyakh core top is very low. Is this lake part of the calibration set? If so, such a low error is not surprising, and it might be worth removing before reconstructing the core top it to get an unbiased estimate. The same applies to the WAPLS model, but the effect will be much weaker and could probably be ignored (unless the assemblage is distinctive), but treat the two transfer function methods the same way.

Our response: Surface samples or core-top samples from Billyakh were not part of the pollen calibration set. We double-checked, it seems that it is just good luck that the transfer function works so well for the Billyakh core top using MAT.

New text: Samples from the lake were not included in the calibration set.

Reviewer comment: (9) If Billyakh is one of the calibration set sites, it could be marked on the panels on the left of figure 2.

Our response: Does not apply because it is not part of the calibration set.

Reviewer comment: (10) How was it decided to use seven analogues in the MAT models? (seven should be written in words, as should any other small integers).

Our response: We wanted to stay consistent with the approach in Herzschuh et al. (2023b)

New text: We used a square-rooted chord distance dissimilarity looking for seven close matches in the modern data following Herzschuh et al (2023b).

Reviewer comment: (11) I don't know if it needs to be stated in the ms, but it may be worth reminding readers that there is a risk of circular reasoning when interpreting assemblage changes due to climate when that assemblage has been used to reconstruct the climate.

Our response: We agree. Added.

New text: Finally, we note the risk of circularity when inferring drivers of vegetation change from the same assemblages used to reconstruct climate; therefore, our reconstructions should not be used to attribute drivers of vegetation change.