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

Response to comments of Referee #2 (Charline Giguet-Covex)

Reviewer comment: (1) This manuscript aims to reconstruct past summer temperatures in Siberia over the last 32,000 years. Four approaches are developed based on modern datasets (observations and fossil records) and applied to a sediment archive (a core from Lake Billyakh). Each method entails specific benefits and drawbacks, which makes the approaches highly complementary and strengthens the robustness and quality of the study. In particular, the approaches combining "raw" GBIF data or species distribution models derived from these data (a method mostly used in ecology) with taxon-specific probability density functions to link taxa to climatic conditions are very original in palaeoenvironmental contexts.

Our response: Thank you for this comment.

Reviewer comment: (2) It is also noteworthy that using plant sedaDNA communities in combination with several statistical approaches significantly improved predictive accuracy (see median bias and root-mean-squared error of prediction), compared to other proxies (e.g. pollen, chironomids). Another strength of the study is the availability of a large modern surface-sample dataset (203 sites), which was used for training (WA-PLS, MAT) and for independent validation of the methods (GBIF-SPD or SDM-SPD).

Our response: Thank you for this comment

Reviewer comment: (3) My main concern relates to the impact on reconstructions that incorporate the SPDs approach of the relative contributions of different taxa estimated by sedaDNA, which most likely do not reflect their actual contributions in the landscape. Since this information is used to weight the SPDs, I think the potential impact should be discussed. This could be addressed in section 4.3. At present, the only mention is in the outlook section: "Furthermore, integrating absolute quantification methods (e.g. DNA-based biomass estimates) could reduce PCR-related biases, enhancing proxy accuracy (Ushio et al., 2018)."

Our response: We addressed this comment by adding some discussion.

New text: In our PDF-based reconstructions, the joint climate signal is weighted by the relative read abundances of taxa to reduce stochastic noise compared to unweighted presence–absence approaches. However, sedaDNA read counts are influenced, among others, by PCR bias, marker specificity, and taphonomic processes, and thus do not necessarily reflect the true abundance of taxa in the landscape (Elbrecht and Leese, 2015; Giguet-Covex et al., 2019).

Reviewer comment: (4) In addition, in my view, the manuscript would benefit from a section addressing other climate proxies beyond vegetation assemblage reconstructions (e.g. BrGDGTs, chironomids, δ18O from diatoms or ostracods).

Our response: We now addressed this in the discussion.

New text: Co-located multi-proxy studies pairing sedaDNA with independent temperature proxies are required, because these proxies differ in seasonality, source area, and taphonomy. Candidates include brGDGT palaeothermometry (De Jonge et al., 2014), chironomid head-capsule transfer functions for lake-summer temperature (Eggermont and Heiri, 2012), and lacustrine $\delta^{18}O$ in carbonates or diatom silica (Leng and Marshall, 2004; Meister et al., 2024).

Reviewer comment: (5) Just one more question that I forgot to include in the review: for the WA-PLS you perform 2 transformations (Hellinger- and then square-root transformed) and for the MAT you only do the Hellinger transformation). Can you explain why?

Our response: Actually, it is similar between the two approaches, i.e. we used Hellinger- and square-root transformation. Because in MAT the additional square-root transformation is already included in the distance metrics sq.chord (rioja R package) which we used for implementing the reconstruction.

Specific comments

Reviewer comment: (6) L27: I would add *"in appropriate contexts, i.e. where vegetation composition is primarily driven by climatic conditions."*

Our response: We did not include it to keep the abstract short. But we explicitly point to this prominently in the third sentence in the introduction.

New text: These approaches build upon the assumption that compositional changes within biotic communities are predominantly driven by changing climatic conditions (Birks et al., 2010).

Reviewer comment: (7) L50: Perhaps mention *topography* as well?

Our response: We added "topography" to the list of information that pollen signals integrate over larger spatial scales.

Reviewer comment: (8) L285–286: The sentence is awkward: *"However, some taxa occur with high abundance that occur today under warm conditions including Crepinidae_01 and Asteraceae_03."* → Suggested: *"However, some highly abundant taxa also occur today under warm conditions (e.g. Crepinidae_01 and Asteraceae_03)."*

Our response: Thank you, we rephrased the sentence according to your suggestion.

New text: However, some highly abundant taxa also occur today under warm conditions (e.g. Crepinidae_01 and Asteraceae_03).

Reviewer comment: (9) L294: For *"outlier taxa"*, please provide a definition: e.g. taxa with warm ecological preferences found during glacial periods, or the opposite.

Our response: We specified "outlier taxa" to "outlier taxa with warm ecological preferences".

Reviewer comment: (10) L339: The word *"quantitatively"* seems unnecessary — temperature is inherently quantitative.

Our response: in this context, "quantitatively" refers to reconstruction, not to temperature.

Reviewer comment: (11) L349–350: Specify that you are referring to *chironomid head capsules*, not chironomid DNA.

Our response: we changed “chironomid-based WA-PLS reconstructions” to “chironomid-based WA-PLS reconstructions from head capsules”

Reviewer comment: (12) L359–361: Runoff refers only to the flow of water over the land surface (rain, snowmelt). Strictly speaking, this does not necessarily include erosion, which is the key process in Giguet-Covex et al. (2019). Suggested formulation: *“SedaDNA mainly originates from the lake’s immediate catchment, since erosion of the surrounding slopes is the dominant transport process.”*

Our response: thank you, we rephrased the sentence according to your suggestion.

New text: SedaDNA mainly originates from the lake’s immediate catchment, since erosion of the surrounding slopes is the dominant transport process (Giguet-Covex et al., 2019).

Reviewer comment: (13) L375: With *Courtin et al., 2021*, you may also cite *Giguet-Covex et al., 2019*.

Our response: We added *Giguet-Covex et al., 2019* to the citation.

Reviewer comment: (14) L380: I would precise "...of direct and indirect climate drivers..."

Our response: We changed “...of multiple climate drivers on community composition” to “...of direct and indirect climate drivers on community composition”.

Reviewer comment: (15) L391: Figure 6 is not placed correctly in the sentence.

Our response: thank you for the notice. We placed *(Fig. 6)* correctly now.

Reviewer comment: (16) L396–401: The explanation could be clarified. Current text: *“The plant metabarcoding assemblages from Lake Billyakh used for reconstruction are substantially more diverse* *than the pollen data (73 vs. 41 taxa), likely contributing to a more stable reconstruction. In contrast, the pollen-based reconstruction displays greater variability (Fig. 4), possibly due to a lower signal-to-noise ratio stemming from fewer taxa contributing to the reconstruction. This interpretation aligns with findings by Heiri and Lotter (2010), who demonstrated that lower taxonomic resolution in chironomid-based temperature reconstructions decreases sensitivity in detecting subtle climate variations."*

→ Suggestion: Make the link between diversity and stability explicit. For example:

"A higher number of taxa increases the likelihood of detecting species with narrow ecological niches, which may strengthen the climate signal and improve the signal-to-noise ratio. This mechanism, rather than diversity per se, may explain the stability of the sedaDNA-based reconstruction." Also, clarify the distinction between *diversity* and *taxonomic resolution* to avoid confusion (you first speak about diversity and at the end about taxonomic resolution).

Our response: Thank you. We implemented this suggestion

New text: The plant metabarcoding assemblages from Lake Billyakh used for reconstruction are substantially more diverse than the pollen data (73 vs. 41 taxa) likely due to the better taxonomic resolution in the sedaDNA signal. A higher number of taxa increases the likelihood of detecting taxa with narrow ecological niches, which may strengthen the climate signal and improve the signal-to-noise ratio.

Reviewer comment: (17) L407: I assume you mean Figure 4.

Our response: yes, Fig. 4 is the correct reference. We corrected it in the text.

Reviewer comment: (18) L423: Please provide the RMSEP value for the plantDNA_PDF/SDM approach.

Our response: We added the RMSEP of 1.9°C for plantDNA_PDF/SDM to the text.

Reviewer comment: (19) L425–427: Clarify whether you are referring to all approaches or only the PDF-based ones. The results for WA-PLS and MAT in Fig. 2 seem less clear.

Our response: We clarified.

New text: The RMSEP of plantDNA_PDF/SDM (1.9°C) is roughly similar to leveraging traditional calibration approaches using modern proxy-based assemblage datasets, which yield 1.1°C for plantDNA_WAPLS and 1.2°C for plantDNA_MAT (Fig. 2). However, in contrast to the PDF-based approaches, they yield a high bias for the Lake Billyakh core top reconstruction of >3°C (Table 1).

Figure improvements

Reviewer comment: (20) Fig. 1: Add a blue border to the text boxes associated with sedaDNA (field sampling, laboratory analyses, bioinformatics). In the bioinformatics part, remove animals (since only plant data were used here).

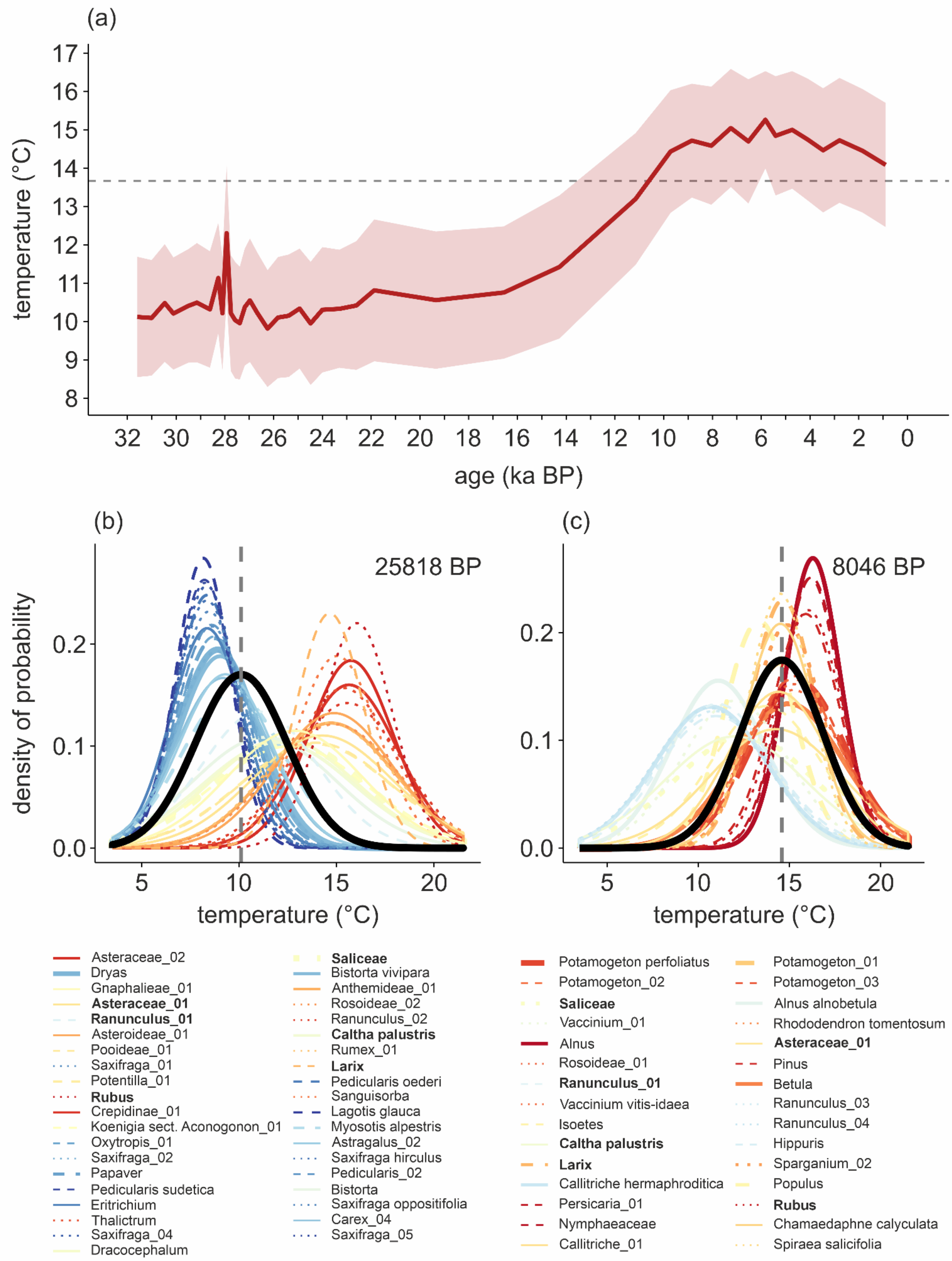
Our response: Thank you. We implemented the suggestions.

Reviewer comment: (21) Fig. 2: The ΔT for Billyakh is not shown — could you add its color code? Also, include the number of modern lake sediment sites used.

Our response: We prefer to not indicate this in the figure. This may be misleading as Billyakh is not part of the calibration set.

Reviewer comment: (22) Fig. 4: Highlight (e.g. in bold) taxa detected in both periods.

Our response: Thank you. We implemented the suggestions (see below).


(a)
temperature (°C)
17
16
15
14
13
12
11
10
9
8
32
30
28
26
24
22
20
18
16
14
12
10
8
6
4
2
0
age (ka BP)
(b)
25818 BP
density of probability
0.2
0.1
0.0
5
10
15
20
temperature (°C)
(c)
8046 BP
0.2
0.1
0.0
5
10
15
20
temperature (°C)
Asteraceae_02
Dryas
Gnaphalieae_01
Asteraceae_01
Ranunculus_01
Asteroideae_01
Pooideae_01
Saxifraga_01
Potentilla_01
Rubus
Crepidinae_01
Koenigia sect. Aconogonon_01
Oxytropis_01
Saxifraga_02
Papaver
Pedicularis sudetica
Eritrichium
Thalictrum
Saxifraga_04
Dracocephalum
Saliceae
Bistorta vivipara
Anthemideae_01
Rosoideae_02
Ranunculus_02
Caltha palustris
Rumex_01
Larix
Pedicularis oederi
Sanguisorba
Lagotis glauca
Myosotis alpestris
Astragalus_02
Saxifraga hirculus
Pedicularis_02
Bistorta
Saxifraga oppositifolia
Carex_04
Saxifraga_05
Potamogeton perfoliatus
Potamogeton_02
Saliceae
Vaccinium_01
Alnus
Rosoideae_01
Ranunculus_01
Vaccinium vitis-idaea
Isoetes
Caltha palustris
Larix
Callitriche hermaphroditica
Persicaria_01
Nymphaeaceae
Callitriche_01
Potamogeton_01
Potamogeton_03
Alnus alnobetula
Rhododendron tomentosum
Asteraceae_01
Pinus
Betula
Ranunculus_03
Ranunculus_04
Hippuris
Sparganium_02
Populus
Rubus
Chamaedaphne calyculata
Spiraea salicifolia