

## General Comments

I very much appreciate the effort the author has made to share his code or should I say software with the community. Without having tested the R package by myself, I am confident to say that it constitutes a very valuable contribution for the community and that many researchers can benefit from it. The package is well described in the manuscript and it seems that the author achieved his goal of keeping the hurdles for application as low as possible while at the same time providing a reasonable degree of flexibility. From a user's perspective package's automated query of the databases is the optimal solution. Also, the diagnostic tools seem to be very helpful for assessing the quality of past climates reconstructions.

Overall, the manuscript conveys the functioning of the R package in a decent way. However, there are several inaccuracies in the language and sometimes it seems the paper has been written with a lack of thoroughness. For example, there are plots without axis labels. The figure and table captions are generally very short. I would suggest to give all of them a careful review. In principle, the figures as well as the tables should be understandable from their captions and the reader should not have to search in the main text for the relevant pieces of information. Additionally, formulas appear to be detached from the text and the definitions of the entities that appear in formulas are sometimes given way before the formula is presented. In general, any formula should be part of a sentence. The structure of the manuscript is well chosen. The introductory part could, however, do a better job in setting the scene. It took me a long time to understand, that we are dealing with proxy data on the one hand, with the abundance of a certain proxy being given in continuous units with respect to a depth (or time) axis and on the other hand with binary presence data from modern observations spread over the entire globe, which if supplement by the corresponding climate data serves for the calibration of the 'climate response function'. Maybe, this is clear to everybody working with palaeoecological proxy data, to me it was not. Another point that confused me while reading, is that in Section 2. there is a clear two-level structure comprised of taxa and the species of each taxon. Correspondingly, the way how PDF's of the different species are combined to the one of a taxon is introduced. Later in the manuscript, more levels added to this structure like families and genus. I do neither understand how PDFs are combined on these intermediate levels nor what is actually measured in the proxy data? I thought, one could either measure the abundance a taxon or of specific species in a sediment core? Again, maybe this is obvious to people who are used to work with palaeoecological proxy data.

My greatest concern relates to the mathematical presentation of how the package actually computes the past climate's reconstruction and with this to Section 2. of the manuscript. This part of the manuscript is definitely lacking the required accuracy and is in some sense misleading. The author calls the function  $\text{PDF}_{\text{reconstructn}}(c, z)$  a *posterior climate reconstruction* (compare l.110). This suggests, that a full Bayesian approach has been pursued. Instead, after having careful studied Section 2. I understand that the climate reconstruction presented by the author is in fact a maximum likelihood estimate of the past climate and that the uncertainties presented in Fig.9 correspond to the

percentiles of the corresponding likelihood function. I would really like to encourage the author to implement a full Bayesian approach in the R package, if not in this version, then in an updated version of the software.

Despite this major issue, I would like to emphasize that this manuscript is only the description of what has actually been the main part of the work, and that is the R package. I consider this package as a very valuable contribution and this paper is a decent description of the package and thus definitely merits publication.

## Detailed Comments on Section 2.

Section 2. starts with a derivation of what is called the 'climate response' or 'PDF' of a given species from the combined modern presence observations and climate data. First, I would recommend to give the object an unambiguous name. Since the term 'PDF' simply describes a certain class of functions, I strongly prefer 'climate response function' which could then be associated with some greek letter like  $\chi$  or  $\rho$  for example.

In fact, the starting point for the derivation of the climate response function is the joint probability distribution of two random variables, namely the climate  $C$  and the presence of a species  $S$  which belongs to some taxon  $t$ . While  $C$  is a continuous random variable, the random variable  $S$  is in fact a binary random variable that indicates the presence ( $S=1$ ) or the absence of the species ( $S=0$ ). Let the joint distribution be denoted as

$$\rho_{S,C}(s, c). \quad (1)$$

Given spatially extended observation data with  $N_{\text{obs}}$  observed tuples  $(s_i, c_i)$ , as depicted in Fig. 2(a) and (b), an empirical distribution can be defined as follows:

$$\rho_{S,C}^{\text{obs}}(s, c) = \frac{\pi^{\text{obs}}(s)}{N_{\text{obs}}} \sum_{i \in \mathcal{N}_s} \delta(c - c_i), \quad (2)$$

where let  $\mathcal{N}_s$  indicate set of indices with  $s_i = s$ , such that  $\sum_{i \in \mathcal{N}_s}$  indicates the sum over all  $c_i$  which have been observed in combination with a given value of  $s$ .  $\delta(x)$  is the  $\delta$ -distribution and  $\pi^{\text{obs}}(s) = \frac{N_s}{N_{\text{obs}}}$ , with  $N_s$  denoting the total number of observations where  $s_i = s$  (where the species is either present ( $N_1$ ) or absent ( $N_0$ )).

It seems that the authors in fact build what they term *climate response of a given species* ( $\text{PDF}_{\text{sp}}(c, s)$ ) on the probability distribution of  $S$  conditioned on  $C$ :

$$\rho_{S|C}(s|c) = \frac{\rho_{S,C}(s, c)}{\rho_C(c)}. \quad (3)$$

$\rho_{S|C}(s|c)$ , certainly is a function of  $s$  **and**  $c$ , however, it is important to note that it is not a probability density function with respect to  $c$ . In words, this function expresses how probable the species will be present, given that the climate has the value  $c$ . The fact that the marginal distribution of the climate variable is not uniform but instead follows some probability  $\rho_C(c)$  is already accounted for in Eq.(3), by the division by  $\rho_C(c)$ . That's why Eq.(3) is called a conditional probability. Compare line 82:

Each observation can also be weighted to account for the uneven distribution of modern climate (Kühl et al. (2002), Bray et al. (2006)). Extreme values are usually under-represented (see, for instance, the inset histogram on Fig. 2c), and this bias can push the estimation of the PDF  $\rho_{sp}(c, s)$  towards the mean climate observed across the study area (i.e. towards the center of the “climate space”).

The authors then impose  $S = 1$  and turn the Eq.(3) into a probability density with respect to  $C$  by introducing a convenient normalization. However, this is all done implicitly and can only be understood from spending some time looking at Eq.(1) and (3) from the manuscript, where the authors compute the expected value of this new probability density they call *climate response* with respect to  $C$ . However, they do not discuss why  $\rho_{S=1|C}(s = 1|c)$  can at all be interpreted as a measure for the *climate response* of a given species. As already mentioned, much of the mathematical considerations are omitted in the manuscript, which makes it very hard for the reader to find out what is actually going on ‘behind the scenes’.

As a next step, the expectation and variance of the *climate response* computed from the observations (Eqs.(1) and (2) from the manuscript) are used to define a continuous normal or log-normal pdf for a given species. Here, I wonder if this is the most convenient way to estimate  $\text{PDF}_{sp}(s, c)$ . Would a least square optimization of the functional form to the observational data yield different and potentially better results?

Finally, the  $\text{PDF}_{sp}(s, c)$  of different species of the same taxon are combined to yield one  $\text{PDF}_t(s, c)$ . It is, however, not explained how the quantity  $\text{PDF}_t(s, c)$  should be interpreted. Does it express some probability, and if yes, what probability? As a reader, one could assume that the taxon’s *climate response* should indicate the probability to find at least one species that is part of the taxon under given climatic conditions. However, this interpretation is inconsistent with the Eq.(6) of the manuscript, since this probability reads:

$$p(T = 1|C) = 1 - \prod_{sp} (1 - \rho_{sp, S=1, C}(1, C)), \quad (4)$$

Where  $T$  is the random variable associated with the presence  $T = 1$  or absence  $T = 0$  of a taxon, that is, at least one of it’s comprising species, and  $\rho_{sp, S=1, C}$  are the climate response functions of the different species comprised in the taxon. Thus,  $1 - \rho_{sp, S=1, C}(1, C)$  denotes the chances for the absence of the species  $sp$  and the product over all species gives the chance to find none of the species given that the climate is  $C$ .

There is a substantial deficit of justification for the way the taxon’s pdf is defined. I would highly recommend to elaborate more on this, since combining different probability densities is often not as trivial as it seems and subtle details are easily overseen. I would also like to encourage the authors to provide the mathematical interpretation of the quantity  $\text{PDF}_t(s, c)$ . What is the meaning of this function or probability measure?

From Eq.(7) the reader can finally understand, that the author follows a maximum likelihood approach for the reconstruction of past climate implicitly building upon Bayes

Theorem

$$p(c|t_1, t_2, \dots, t_n) = \frac{p(t_1, t_2, \dots, t_n|c) p(c)}{p(t_1, \dots, t_n)}, \quad (5)$$

where  $\text{PDF}_{reconstrctn}(s, z)$  serves as the likelihood-function. I think that introducing Bayes theorem at the beginning and then deriving step by step a convenient expression for the likelihood-function would make it much easier for the reader to follow the line of thought.

There are two further obvious questions: First, why do the authors not make use of the full Bayesian theorem and deduce posterior probability distributions  $p(c|t_1, \dots, t_n, z)$  and with this a rigorous quantification of uncertainty? Introduction of convenient priors could actually account for the fact, that also for past climate there is a statistical bias in the sense that finding a species or a taxon under climatic conditions which occur very often is more likely than finding the species or taxon under climate conditions which occur only rarely even if the latter conditions are substantially preferred by the taxon or species. The author compensates this effect in the derivation of the climate response from the calibration data, however, he does comment on the fact that the same effect existed in the past. Second: the implicit normalization of the species climate responses undermines their interpretation as building blocks of the likelihood function. The individual normalization of the  $\text{PDF}_{sp}(c, s)$  effectively introduces weights which do not have a mathematical meaning. As I already pointed out:  $p(s|c)$  is not a probability density function with respect to  $c$ !

## Specific comments:

1.9 I t

there is a space in the word.

1.22 Despite their conceptual simplicity and demonstrated capacity to reliably reconstruct climate from palaeoecological datasets, the limited availability of robust calibration datasets (i.e. regional collections of modern proxy samples) beyond the Northern Hemisphere extratropics has, however, hindered their application in these regions, despite the existence of suitable records from all environments worldwide (Chevalier et al., 2020b).

The reference of 'these regions' is unclear. Is 'these regions' regions beyond the Northern Hemisphere extratropics? Maybe 'outside this region'?

Should't extratropics be capitalized?

1.27 Built upon from the original work of Kühl et al. (2002) — who first proposed to replace the commonly-used modern proxy samples with modern proxy geolocalised occurrence data to estimate probabilistic proxy-climate relationships — CREST estimates and combines probability density functions (PDFs) to reconstruct climate parameters.

I am not a native speaker, but are you sure that 'built upon FROM' is correct?

Also, at this stage 'CREST estimates and combines probability density functions (PDFs) to reconstruct climate parameters' could mean anything.

Maybe the sentence becomes more meaningful if you said: 'CREST estimates and combines climate response functions for numerous species to reconstruct climate parameters from fossil occurrences of these species.'

1.37 In addition to its broad applicability, CREST also bears some fundamental statistical properties that make it well-adapted to the analysis of palaeoecological datasets (Chevalier et al., 2020b).

I would rather say: CREST is equipped with some fundamental statistical features.

1.39 ... CREST estimates, weights and propagates all the climate values that are compatible with the observed fossil data.

superfluous comma

1.45 analytical solution

This term could be misleading. Do you mean analytical in contrast to numerical?

1.52 In addition to its technical core, the package also contains an array of graphical diagnostic tools to represent the data at different pivotal steps of the reconstruction process and facilitate objective evaluations of the data and results.

It seems there is a 'to' missing in front of facilitate.

- 1.54 First, Section 2 summarises the mathematics and assumptions underpinning the approach and introduces the embedded calibration dataset.

I think underpinning should be replaced by underlying.

- Fig.1 Conceptual illustration of the differences between a modelling approach based on the estimation of the most likely climate where all the probabilities are concentrated around the mode (light grey)

What exactly does the word 'mode' mean in this context? Also, in order to be precise, you should add, that already in panel (a) the best estimate contains statistical uncertainties. Otherwise the corresponding pdf would be a delta peak.

The two types of approaches are illustrated in (b) and (c) with two theoretical fossil samples (in blue and purple) representing two independent reconstructions of the same climatic parameter for the same time period.

It seems like what you are actually illustrating are the consequences of the two different reconstruction approaches?

The same samples are reconstructed using an approach that estimates their complete uncertainty distributions.

The 'samples' cannot be reconstructed.

In this case, the response of the blue sample is broader, and the response of the purple sample becomes bimodal.

Would you really say 'response'? It seems that you are talking about a posterior probability distribution. Maybe, it is worth introducing the term 'climate response' at a very early stage in the manuscript.

- 1.60 As is standard with statistical climate reconstruction techniques, the core process of CREST can be decomposed into two major stages: 1) estimating the modern climatic responses of the proxies observed in the fossil sequence (Fig. 2a-c) and 2) combining these responses to reconstruct past climates (Fig. 2d). In the following sections, the main elements of these two stages are presented along with all the parameters and/or modelling assumptions that can be modified in crestr. For an in-depth description of the method and its assumptions, the reader is, however, referred to .

please delete two doubled paragraph

- Fig.2 Occurrences of four species part of the same pollen group exhibiting marked preferences for the lowest values of that climate (e.g. dark/cold values) cross the study area.

Shouldn't it be: the lowest value of that climate **variable across** the study area?

Maybe, it would be helpful if you pointed out, that the occurrence data is binary and does not account for the abundance of the species at a given location. For a better understanding, you could even indicate, that the species' PDFs are derived by binning the climate space and then compute the fraction of the climate variable's observation in a certain bin, that is accompanied by the presence of the species.

The histogram represents the proportion of the modern climate space (white) occupied by at least one of the four species (black), highlighting the higher chances of observing the taxon at the lower end of the climate gradient.

Please indicate that you are talking about the inset of panel (c).

Example of a posterior climate distribution resulting from the multiplication of PDFs and the type of synthetic statistics (e.g. optimum, mean, uncertainty range) one can derive from it.

This panel is very hard to understand within this Figure. I would suggest to divide the 4 species into two taxa such that panel c would show two different taxon's pdfs. Next, I would recommend supplementing the caption with the information that in a hypothetical proxy sample the two taxa were observed. This would allow the reader to understand much faster what is actually shown in panel (d). Please consider my objections presented in the comments on Section 2. with respect to the term 'posterior distribution' in this context.

- 1.72 In CREST, PDFs are used to transform the information contained in the modern observations of biological climate proxies into probabilistic climate responses. A PDF thus represents a weighted ensemble of all the conditions where the proxy is observed today. PDFs can be fitted in one or two steps depending on the nature and taxonomic resolution of the studied proxy. Climate responses are first fitted at the species level (hereafter PDF sp (c, s) with c representing the studied climate variable and s a species), and when necessary, these PDF sp (c, s) are then combined together to meet the taxonomic resolution of the fossil taxon (hereafter PDF tx (t, c) with t representing the observed taxon).

This sounds as if *a pdf* is something new.... The term simply denotes a specific class of functions.

- 1.72 In CREST, PDFs are used to transform the information contained in the modern observations of biological climate proxies into probabilistic climate responses.

I am not aware of the term 'climate response'. However, it might be that this is a typical term in the community that works with ecological proxy data, in that case, please ignore my comment. I would recommend to identify the climate response with the likelihood function in a Bayesian setting.

- 1.73 A PDF thus represents a weighted ensemble of all the conditions where the proxy is observed today.

hmm... There are two things which I find confusing about this sentence.

1) coming more from an ice perspective I am used to proxies whose value and not whose presence or absence can be used to for paleo climate reconstruction (most prominently levels of  $\delta^{18}O$  can be used to study past temperature) – it seems that this is different in this context. Maybe this could be clarified at an early stage.

2) A pdf usually characterized the probability for a random variable to assume a certain value in a random experiment. I assume, the random variable in this case is the climate variable  $x$  under study, conditioned on the presence of a given proxy  $y$ .

$$P(a < x < b | y) = \int_a^b \text{pdf}(x|y) dx$$

If this is what you aim to express, I don't think the above statement is very precise.

1.87 In CREST, this weighting can be accounted for by first sorting the  $N$  climate values that compose the climate space into bins of equal sizes (e.g.  $2^\circ C$  or 50 mm).

To me, it is unclear what 'the climate space' is. Is it a spatially extended region under study which is subdivided into  $N$  grid cells, each of which can be assigned a value in terms of a specific climate variable? Or is the 'the climate space' simply the value range which is covered by the climate variable globally (regionally)?

It seems, that  $N$  is the total number of observations and that each observation is associated with two variables, namely the climate variable – which is continuous – and the presence or absence of a certain species, which is a binary observation.

1.90 Please define the variables that you use in formulas. From the context, I could guess that  $m_{s,c}$  is the mean value for some climate variable  $c$ , from different observations  $c_i$  of this variable. Probably, conditioned on the presence of a certain proxy species  $s$ . Further I assume, that  $k(c_i)$  are the weights?

1.110 Finally, the PDF  $sp(s, c)$  of the  $S(t)$  species composing taxon  $t$  are linearly combined to create the climate response of taxon  $t$  to climate variable  $c$  (Eq. 6).

What exactly is meant by linearly combined? Does that mean  $\text{PDF}_{\{t,c\}} = \sum \alpha_i \text{PDF}_{\{s_i,c\}}$ , where  $\alpha$  are weights? how do you choose the weights?

The authors use the point estimates for mean and variance from the observed histograms to define a gaussian  $\text{pdf}_{\{s,c\}}$ . Is this equivalent to fitting a gaussian to the relative histogram in a least squared sense?

I must say, that I do not find the way how the species pdfs are combined to the taxon's pdf very convincing. To give a counter example: Let's say over a region as displayed in Fig.2 (a), there are  $k$  grid cells where the climate variable has the value  $c^*$ . Let half of the grid cells be populated by species  $s_1$  and let also half these grid cells be populated by species  $s_2$ , with an overlap such that in total  $\frac{3}{4}$  of these  $k$  grid cells are populated by either  $s_1$  or  $s_2$ .



Equation (6) now suggests, that the PDF<sub>tax</sub> will yield a probability of 50% that under climate condition  $c^*$  the taxon is present (given that the gaussian fit to for the species was fairly accurate). In fact, from the observations we know, that there is a 75% chance that under the climate conditions  $c^*$  the taxon can be found.

Please see my considerations in my detailed comment to Section 2.

- l.110 With the PDF  $tx(t, c)$  calibrated, posterior climate reconstructions can be estimated from their multiplication (Eq. 7, where  $z$  represents the age or depth of the sample to reconstruct, and Fig. 2d).

This is a very generic sentence, which is true only under very specific circumstances, namely, for a given time in the past, the presence of different taxons at the same location must be evident from proxy records. Then, the different PDF<sub>{tx}</sub> can be used to refine the past climate's reconstruction.

However, I believe, it would be worth explaining in one or two sentences, how past climate is reconstructed from a single taxon proxy record.

- l.114 As such, it is possible to select a subset of climatically sensitive taxa to reconstruct each climate variable and maximise the reconstruction signal (Chevalier and Chase, 2015), even if it is not always mandatory (Chevalier et al., 2021).

It is unclear what is meant by the 'maximisation of the reconstruction signal'. I assume the authors mean, that the reconstruction's uncertainty is minimized or even more precise, that the width of the posterior distribution of the reconstructed climate variable is minimized.

Another sentence shortly before this one also starts with 'as such'.

even if it is not always mandatory (Chevalier et al., 2021)

I do not really understand this comment. In my view, it is always desired to reconstruct the past climate as precisely as possible. Of course, it is not mandatory to reconstruct past temperatures in northern Europe to the precision of two digits behind the comma to deduce that there have been ice ages. So why do the authors add this comment here and even provide a reference for this statement?

- Eq.(7) Typically, equations are part of sentences. The normalization seems a bit odd:  $1^{\{1/\#\text{observed taxons}\}}$ ?

I have so far not encountered an exponential weighing scheme like the one used here, but that does not mean anything. It only seems a bit odd, that previously objects which are not pdf's have been normalized to one and now, the pdf<sub>restrctn</sub> is obviously not integrate to one anymore after introduction of the weighing. - Well, it's a likelihood function and not a pdf, so there is no need that the expression integrates to one.

- l.125 Maybe, before explaining the normalization of the data, you could say a few words about the structure of the data. How is the presence of a taxon measured? How can you compare the

abundance of different taxa with each other? Do you compare weights? Or maybe, this is clear to everybody from the community?

Eq.(8) Again, the formula should be embedded in a sentence. Also, it would be of help, if the variables used in the formula are defined close to the formula. Finally,  $z$  was defined as depth, which is a continuous variable – obviously, measurements are taken at discrete depths  $z_i$ . The sum should then correctly run over the index  $i$  and not over the continuous variable  $z$ .

l.138 The is a spacing missing above the section title

l.150 The coordinates of all the presence records of these six common palaeoecological fossil proxies were upscaled at a spatial 150 resolution of  $0.25 \times 0.25^\circ$ .

I assume the ‘150’ is a misprint? What means ‘upscaled at’ ? Do you mean binned into the grid? Or maybe ‘sorted into the grid’?

Fig.3 Data density of the six climate proxies available in the gbif4crest calibration database.

I assume that the density is defined by the number of presence observation of different species within a certain grid cell, divided by the specific surface of that cell? Please add a unit to the colorbar.

Table1 List of terrestrial variables available in the gbif4crest database. Each one can be selected in crestr using its associated code. List of abbreviations: (Temp.) Temperature, (Precip.) Precipitation.

Is this the ‘*climatological data*’ mentioned in line 165? And is this data available for each grid continental cell that? Why are the ‘*environmental*’ or ‘*geopolitical*’ variables not listed here?

Table2 See table 1.

Fig. 4 Maybe, you could in the DISTRIB table add one line – I believe that would help to understand the structure of these tables.

What means ‘type of observation: literature’ in the DISTRIB\_QDGC table?

The figure caption could be a lot longer and explain the purpose of the different tables.

l.204 inputs: contains the raw data (e.g. the counts/percentages, the ages of the samples or the names of the fossil taxa).

What are the *raw data* used as input for the crestObj? I may only guess, that this is proxy data the user has to provide?

l.211 – reconstructions: contains all the results (e.g. best estimates, synthetic error measurements as well as the full posterior distribution of the uncertainties).

From Section 2. it is unclear to me, how a full posterior distribution of the climate is obtained within the presented modeling approach. Please see my detailed comment on Section 2.

- 1.214 Five different input data files are compatible with crestr. However, most applications will only require two file (the df and PSE files, see below) to be created. More specific applications may require up to four of these files. All the files can be prepared outside the R environment and imported using standard R functions.

will only require two files

See 1.204 – it becomes clear now, that this input data is actually the proxy data users aim to build their climate reconstruction upon. I would propose to state this explicitly – it might be obvious to the author, though for the reader it is not.

- 1.223 The proxy-species equivalency (PSE) table

I am little confused by this section. In Section 2. it is explained how the climate responses of different species of a given taxon are combined to the taxon's climate response. Hence, there are two levels involved.

Now, it seems that the category 'taxon' was replaced by the category 'family' and on top of that a third level 'Genus' was added. So I wonder how, climate responses on the species level are first combined to a 'genus' level and then to a 'family' level?

Maybe this confusion of mine is simply due to my lack of knowledge in this field. If you think, Sec 3.3.2 will be understood correctly by the relevant audience, then please ignore this comment.

Also, I would find it helpful to understand, what type of proxy data for past climate can actually be observed? Species? Or only taxa? Why and how does this differ between situations?

- 1.316 – To estimate reliable PDFs, it is recommended to use at least 20 distinct occurrences for each species, but different values can be specified with the minGridCell parameter.

Here, I do not understand whether 'distinct' refers to the level of individual observations as stored in the DISTRIB table, or to the level of QDGC grouped observations?

- 1.362 4.3 Estimating the climate responses (the PDFs)

I recommend to call '*the PDF's*' consistently climate response functions throughout the entire manuscript. PDF is a specific type of function that fulfills certain requirements. As mentioned previously, the climate response functions are likelihood functions and if they weren't normalized they would not be pdfs. Also, the term 'climate response functions' describes more accurate the purpose of the functions.

- 1.373 set geoWeighting to TRUE if the species PDFs of the different composing species should be weighted according to the square-root of the extent of their modern distribution.

This refers to combining the species' pdfs to the taxa pdfs?

1.387 2) the climate values to reconstruct are likely to be in the study area (the reconstructions are bounded by the lowest and highest values observed in the modern climate space)

Can you think of a term other than ‘study area’? This sounds more like a geographical location.

Or maybe you actually mean the geographical region. In that case, the above sentence should be specified, e.g.: the climate values to reconstruct are likely to be covered by present day climate values in the study area.

1.393 In our case study, the spatial variability represent true patterns in regional species diversity with the presence of several biodiversity hotspots across eastern and southern Africa (Myers et al., 2000).

represents

1.397 This diagnostic figure is also very important to identify potential local or global correlations between different climate variables and assess the risks of confounding variables (Juggins (2013), Chevalier et al. (2020b)).

What is the ‘risk of confounding variables’? Please elaborate.

Fig.6 ‘Number of unique species occurrences’ in other words means ‘Number of different species observed at least once in a grid cell’ is that correct?

Please give the different panel labels (a), (b),... and make sure that all axis have labels.

Fig.7 In the top row, the map represents the density of unique species occurrences per grid cell and the time series represents the variability of the taxon against time or depth.

If I am not completely mistaken, the map shows data from the QDGC\_DISTRIB table, that is calibration data from present day observations, while the time series shows data from the proxy data from the sediment core. Please make sure to specify this unambiguously in the caption.

Also, make sure that all axis are labeled and provide labels for different panels of the plot.

The color bar is again labeled ‘Number of unique species occurrences’. I assume the difference to Fig.6 is that here, the number of observed species belonging to the taxon ‘Ericaceae’ within a grid cell is shown, while Fig.6 shows the total number of observed species under consideration. Please make sure this difference is specified either in the caption or in the label of the colorbar itself.

The left panel on the bottom row is not very useful, since the information on the presence of observations hides the information on the climate. Maybe, you could simply sketch the outline of those connected patches (with some minimum size) which are covered by the taxon.

1.429 (all the probabilities sum to 1)

Maybe better: the PDFs integrate to 1.

- 1.435 This type of representation can be particularly helpful to have objective interpretations of ecological changes from pollen diagrams (Chevalier et al., in press, Quick et al. (2021)).

I do not understand this sentence. Interpretations can never be objective.

- 1.452 Here for instance, both Aizoaceae and Chenopodiaceae/Amaranthaceae were excluded because they are not primarily sensitive to temperature in southern Africa.

To understand this, it would be nice to have them included in the violin plot (Fig.8).

- Fig.10 I can only guess, that the 64 in the upper right corner indicates the individual sample from the marine sediment core that is associated with a certain depth and that these kinds of plots can be obtained for all individual samples? If not, and the plot refers to the entire reconstruction, then I do not understand how the weights can be specified, since early in the manuscript the authors explained that the weights of the different taxa in the climate reconstruction can vary over the depth of the core.

The black curve represents the posterior MAT reconstruction, from which a 'best' climate estimate can be estimated from the maximum of the curve and uncertainties derived by calculating the area under the curve.

From what I understood from Section 2., the black curve does not show a posterior probability distribution for the past climate but instead the likelihood function.

The line thickness can hardly be discriminated by eye. Did you try plotting the different climate responses according to their weights as expressed in Eq.(7) of the manuscript?

- 1.501 The presence of multimodality can be the underlying cause of apparent noise in the reconstructions because minor changes in the taxa composition or percentages can force the system to oscillate between two maxima and thus 'appear' noisy, even if the background rate of change is minor.

We are not really confronted with an oscillating system here. Maybe rephrase:

The presence of multimodality in the climate reconstruction of subsequent samples can be the underlying cause of apparent noise in the reconstructions because minor changes in the taxa composition or percentages can easily switch the order of the two local maxima in terms of height. This results in a jumpy time series of the 'optimal climate reconstruction', even if the background rate of change is minor.

- Fig. 11 Why are all anomalies positive? Is this by coincidence and due to the choice of taxa presented in the plot?

Here, the results are only showed for a subset of the taxa observed in marine core MD96-2048 (only 20 out of the 171 available taxa are represented).

'Showed' should be shown.

The title of the plot 'mean annual temperature' is not very suited. As far as I understood, the plot shows the anomalies for reconstructions that are based on all but one taxa with respect to the reconstruction based on the full set of taxa.