

The comments of Reviewer 1 are in black, Reviewer 1's quotes of the original manuscript are in purple, and my responses in blue. Italicised texts are quotes from the revised manuscript.

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.

Thank you very much for this praise of the tool I developed. I understand that Reviewer 1 is not an expert on palaeoecological datasets, but the fact that Reviewer 1 did not try the package, considering it is the novelty presented in this paper, feels a bit like a missed opportunity. I am, however, very grateful for all the other comments and suggestions. Despite what I just said, I think Reviewer 1's "outside" perspective has also helped me clarify many details in the manuscript, especially things that are usually taken for granted but still need to be explained to reach a broad audience. I believe addressing Reviewer 1's comments has really helped improve the manuscript's clarity and accessibility, and I am very thankful for that.

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.

I considered the points raised by Reviewer 1 and tried to add as much clarity as possible across the manuscript. I agree that some elements might have been a bit too disconnected, indeed. Having taken some distance from the manuscript, I also realised that many comments came from an improper introduction of the studied objects and unclear areas of the manuscripts, also echoing some of the comments made by Reviewer 2. As such, I have largely modified the introduction and used it to better 'set the scene' and more clearly highlight the content of the paper. I also provided more details at the beginning of sections 2 and 3 to better introduce the type of data/proxies that I am discussing throughout the text. I think the level of details

was insufficient in places to reach a broad audience. I hope these changes will allow colleagues from other fields to understand the paper and the tool.

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 $PDF_{\text{reconstructn}}(c, z)$ a posterior climate reconstruction (compare l.110). This suggests, that a full Bayesian approach has been pursued. Instead, after having carefully 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.

As explained in length below, the use of 'posterior' was an inappropriate use of the language. The model is not Bayesian ... yet! Going full Bayesian is definitely an avenue of development that is being explored now that all the foundations (calibration data for different proxies, climate variables, automatic plots, etc.) for the global applicability have been created with this package. Still, I think the model works well in its current form, and it is hoped that this first non-Bayesian version of the package will be of interest. Subsequent versions of the package will definitely include a complexification of the model.

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 merited publication.

Thank you for the general positive feedback. I hope that my responses below and my manuscript modifications will address your concerns adequately.

Detailed Comments on Section 2.

[I could not copy / paste this bit of text from the pdf because it was full of mathematical terms.]

I greatly appreciate all the efforts Reviewer 1 made to suggest a full Bayesian version of the CREST algorithm. This is actually a work that is currently underway. However, progress at this stage is insufficient to be released, but we will look in detail at the suggestions that were made by Reviewer 1. However, the model of Reviewer 1 is based on an erroneous assumption about the available data type. The model proposed by Reviewer 1 is based on the existence of both presence and absence data of plant taxa. However, absence data are not available. Only presences are. While some absence might be inferred from regions with dense data networks (*e.g.* in Europe), sampling is incomplete in many other regions of the world (see Fig. 3 for instance), and many locations where a taxon is not observed could be an actual absence or a proof of insufficient sampling. This information is thus unreliable and should not be used. As a consequence, the CREST model is based on presence-only data, which explains some of the modelling choices that were made to accommodate data.

This 'mistake' of Reviewer 1 is likely caused by the lack of clarity of my introduction and the model's description. I hope the new version of the introduction and the new section at the beginning of section 2 better introduce these problems and better support the choices made to develop the CREST model almost a decade ago.

Specific comments:

I.9 It

there is a space in the word.

Corrected.

I.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'?

The introduction has been rewritten and the specific sentence highlighted here is rephrased as: *"However, the limited availability of the necessary calibration datasets beyond the Northern Hemisphere extratropics has often hindered their application in many environments and regions where quantified climate records are needed, despite the existence of suitable fossil records".*

Should't extratropics be capitalized?

I am not a native speaker either and have not seen it capitalised anywhere else.

I.27 Built upon from the original work of Köhl 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?

'From' was deleted.

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.'

Rephrased as follow: *"CREST estimates and combines probabilistic proxy-climate relationships to reconstruct past climate parameters from fossil proxy observations".*

I.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.

Corrected.

I.39 ... CREST estimates, weights and propagates all the climate values that are compatible with the observed fossil data.

superfluous comma I.45

The comma is necessary, but weighs (the verb) was misspelt as weights (the noun). Hence the confusion. The sentence has been simplified as follow: *“While techniques such as MAT or WA-PLS are primarily designed to associate modern proxy observations with their ‘most likely’ or ‘mean’ climate values only, **CREST estimates and weighs all the climate values that are compatible with the observed fossil data.** As such, the climate reconstructions obtained from CREST can be understood as an ensemble of all data-compatible climate values and not a simpler, less informative ‘most likely’ or ‘best’ climate estimate with statistical errors.”*

analytical solution

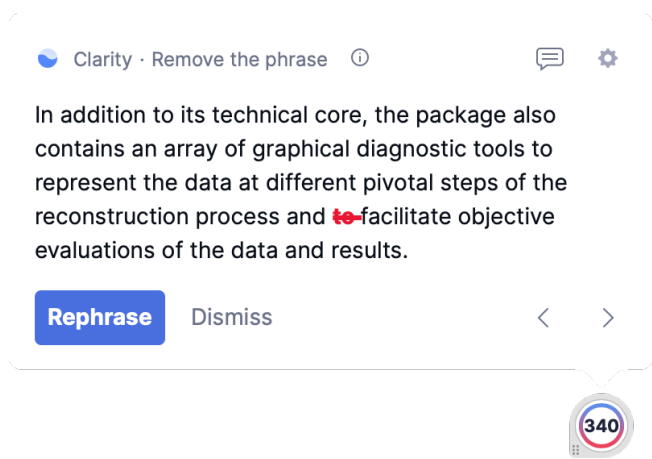
This term could be misleading. Do you mean analytical in contrast to numerical?

This term was confusing indeed. It is now removed.

I.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.

Grammatically, there is no need to repeat ‘to’ here. At least, according to my (British) English grammar corrector.



I.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.

These two words seem largely synonymous in this context, but 'underpinning' was 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.

I am not sure I understand why Reviewer 1 is confused about using this term. Here I used mode according to its definition, i.e. the value with the highest probability. It is also widely recognised that any parameter estimate has some associated errors. As such, I am not sure much emphasis is essential here. Maybe this is a difference between fields, but I am pretty sure nobody would expect reconstructions to be a single (delta?) peak. I have modified the caption as follow: "(a) Conceptual illustration of the differences between a modelling approach based on the estimation of the most likely climate value with small statistical errors surrounding it (light grey), and a modelling approach focused on the full spread of the data with the probabilities more spread along the climate gradient (dark grey)."

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?

Corrected as "The results of the two approaches are illustrated [...]"

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

The 'samples' cannot be reconstructed.

Replaced by 'the same fossil assemblages are analysed'

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.

I am not sure the term 'climate response' is more explicit at this point of the manuscript. I am now referring to those using a more generic 'reconstruction'.

Full caption with the changes suggested in the past 3 comments in bold: *“(a) Conceptual illustration of the differences between a modelling approach based on the estimation of the full spread of the data with the probabilities spread along the climate gradient (dark grey), and **a modelling approach focused on the estimation of the most likely climate value with small statistical errors surrounding it (light grey)**. In both cases, the area under the curve sums to one. The results of the two types of approaches are illustrated in (b) and (c) with **two theoretical fossil assemblages (in blue and purple) used to produce two independent reconstructions of the same climatic parameter for the same time period**. (b) The two reconstructions are derived from a method focused on only estimating the most likely climate value, resulting in 'apparently' incompatible reconstructions. (c) **The same fossil assemblages are analysed using an approach that estimates their complete uncertainty distributions**. In this case, **the blue reconstruction is broader, and the purple reconstruction becomes bimodal**. When the full spread of these uncertainties is considered, the two reconstructions are not incompatible anymore, and a joint climate estimate (gold) can be derived from their overlapping sections (hashed polygons).”*

I.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

The duplicated paragraph was removed.

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?

It is, indeed. Corrected.

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.

I have added a new section (2.1) that introduces the type of data required by the model. I think this will make the 'modelling' part that follows easier to follow and understand.

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).

Corrected as follow: "The **histogram (inset panel)** 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."

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.

I have repackaged the figure to simplify its meaning. The right-most curve does not represent a reconstruction anymore but the response of the pollen taxon (a zoom on the black curve presented in panel c). As explained above, all notions to a posterior reconstruction have been removed. The figure shows how to estimate the response of a pollen taxon using a given climatology and the distributions of four composing species.

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.

A pdf is not anything new, but it is probably not something commonly used by the target users of crestr? As such, I believe a definition of what it is and how it could be grasped in the context of CREST is not superfluous.

I.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.

The sentence was rephrased as follow: *"The transformation of the information contained in the modern observations of the biological climate proxies into probabilistic climate responses can be done in one or two steps depending on the nature and taxonomic resolution of the studied proxy."*. Climate response is a term that I have heard and been using for many years without any apparent issues.

I.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 d18o 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.

The aim of this sentence was to clarify things, but it seems to be doing the exact opposite. Since it was unclear to both reviewers, it was removed from the manuscript.

I.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 ° 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 cell, 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)?

What climate space means is not defined in the new section 2.1. "*Climatology(ies) of the variable(s) to reconstruct gridded at the same resolution as the modern occurrences. All the climate values observed in the study area define what is later referred to as the climate space.*"

It seems, that N is the total number of observations and that each observations 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.

Absence data are not available; this is now part of the new section 2.1 of the revised manuscript.

I.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?

All these variables are now clearly defined in the text (see section 2.2 of the revised manuscript).

I.110 Finally, the PDF $p(s, c)$ 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 $PDF_{\{t,c\}} = \sum \alpha_i PDF_{\{s_i,c\}}$, where α are weights? how do you choose the weights?

The model offers to weigh the pdfs differently based on how much we know about the taxa distributions. This option is not mandatory in the model. In many regions a trade-off between quality and quantity has to be made and to include more species, we need to lower the quality bar. This weighting option is just a little trick to rebalance things a bit.

*This linear combination ensures that all the climate values that support the presence of at least one species have a non-null probability in the $PDF_{tx}(t,c)$. Contrary to the previous step, no additional constraints are added here. The distribution of the $PDF_{tx}(t,c)$ can thus be asymmetrical and even multimodal if different (groups of) composing species exhibit distinct climate requirements. **An additional option is to weigh the different $PDF_{sp}(s,c)$ by the square root of the number of individual occurrences composing their distribution (N_s). Considering that it is more difficult to estimate robust parameters with few points, this weighting gives more importance to the species with more extensive geographical distributions today. Said differently, it gives more weight to the species whose climate responses can be the most reliably defined.***

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

Is it *strictly* equivalent? I cannot tell. But it would certainly lead to very similar results considering the nature of the estimated parameters.

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 $3/4$ of these k grid cells are populated by either s_1 or s_2 .

Equation (6) now suggests, that the PDF_{tax} 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.

I think it is a very interesting way of looking at the data that deserve further thinking. This is very important and I will try to take it into account to improve the algorithm and make better reconstructions, indeed. However, this paper is about the implementation of the algorithm presented in my 2014 paper, and I want to keep it in that way. The model has proven to be able to produce reliable reconstructions as it is. I have, however, added a paragraph of discussion on how the CREST algorithm could be improved in the future, and it definitely includes some of the considerations Reviewer 1 made in its detailed comments of section 2 (including a transition to a fully Bayesian model).

As I mentioned earlier, developments towards a more Bayesian model are currently underway. But progress and validation are unfortunately not advanced enough to be published yet.

I.110 With the PDF $p_{tx}(t, c)$ 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_{tx} 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.

I am afraid I am not understanding what Reviewer 1 is asking here. Single taxon reconstructions are not a thing with palaeoecological datasets. The value of these datasets is their multitude! Reconstruction techniques tap into the co-variations of taxa to infer climate.

I.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.

This phrasing was unclear and has now been clarified, as follow: "In this approach, the presence of each taxon in a sample is considered independent from the others. It is thus possible to select a subset of climatically sensitive taxa. **In some cases, identifying a subgroup of climate-sensitive taxa can help disentangle the different climate signals represented by the palaeoecological data and improve the quality of the reconstructions [CC15], even if it is not always necessary [Chevalier2021Limpopo].** In practice, these choices should be dictated by the data themselves and the users' understanding of the studied proxy system. `_crestr_` provides graphical tools to help identify the possible climate sensitivities of the studied taxa [Chevalier2021Atlas]."

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

Corrected.

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?

This part of the sentence referred to the beginning 'selecting a subset of sensitive taxa', which is not always required depending on the data. This was rephrased as follow: "*In this approach, the presence of each taxon in a sample is considered independent from the others. **It is thus possible to select a subset of climatically sensitive taxa. In some cases, identifying a subgroup of climate-sensitive taxa can help disentangle the different climate signals represented by the palaeoecological data and improve the quality of the reconstructions [CC15], even if it is not always necessary [Chevalier2021Limpopo].** In practice, these choices should be dictated by the data themselves and the users' understanding of the studied proxy system. `_crestr_` provides graphical tools to help identify the possible climate sensitivities of the studied taxa [Chevalier2021Atlas].*"

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

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_restrctn 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.

I think Reviewer 1 misunderstood how $\omega(t,z)$ was calculated, which created some confusion. I have now better explained the different options and I hope this expanded text clarified things. I also agree that the pdf_rcnstrctn is not a pdf, indeed. The object has been renamed RECON(c, z) and all the equations have been included in a sentence.

I.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?

I have explained the structure of the data in section 2.1, and I think that the questions asked here by Reviewer 1 wouldn't be asked by colleagues from the field that are familiar with palaeoecological datasets. For instance, a presence is simply determined by the observation of the taxon in the sample (which can subsequently be truncated if it doesn't reach a sufficient percentage), the abundances are never compared directly, only the trends. It is actually to overcome this lack of comparability that I propose the normalisation of the percentages.

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.

All formulas and chunks of code are now embedded in the text. I believe that indexing z would only make things unnecessarily complicated.

I.138 There is a spacing missing above the section title

This is a problem generated by the Rmarkdown -> markdown -> latex scripts. Fixing this is beyond my skills and I have no doubt this will be fixed at the typesetting stage.

I.150 The coordinates of all the presence records of these six common palaeoecological fossil proxies were upscaled at a spatial 150 resolution of 0.25×0.25 °

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

150 was a typo and is now fixed. 'Upscaled' was replaced by 'mapped onto'.

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.

The title of the figure is 'Number of occurrence data per grid cell'. I believe this states what the plots are about. I have, however, replaced 'Data density' with 'Distribution and grid cell density of the six climate proxies available [...]' to avoid further miscomprehension.

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.

The non-climatic environmental and geopolitical variables are not meant to be reconstructed. These tables are about the variables that can be reconstructed. This is now clarified in the text (see below).

*"In the _gbif4crest_ database, all the QDGC grid cells were associated with a collection of terrestrial and oceanic environmental variables that can be reconstructed (@Fick_Hijmans_2017, @Zomer2008, @WOA_2018_temp, @WOA_2018_salinity, @WOA_2018_oxy, @WOA_2018_nutrients, @Reynolds2007, see details in Tables \ref{table:variables-terr} and \ref{table:variables-mari}). Despite the diversity of variables available, it is recommended to avoid serial reconstructions and, on the contrary, to identify the few important variables for the studied palaeoecological datasets *_a priori_*. **The grid cells were also associated with 'non-reconstructible' environmental and geographical descriptors that serve to tailor the calibration dataset to the users' needs. These include the coordinates, the elevation and elevation variability within the grid cell (Amante and Eakins, 2009), the country (<https://www.naturalearthdata.com>) or ocean (<https://www.marineregions.org>) names, as well as different levels of ecological classification for the terrestrial (Olson et al., 2001) and marine (Costello et al., 2017) realms."***

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

I did not modify the figure, but I added some additional explanations in the text regarding the structure of the tables. However, I also need to point out that understanding the tables is not necessary to use the package as the interface with the database is done 'under the hood'.

What means 'type of observation: literature' in the DISTRIB_QDGC table?

I have expanded on that in the main text and added a link to a webpage that explains how GBIF classifies its data.

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

While I generally agree that figures should be understandable on their own, in this case, it would only lead to a repetition of the text, which is quite long.

I.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?

Yes, indeed. This has been clarified.

I.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.

See my response to that point there as well.

I.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

Corrected.

See I.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.

Now clarified, see above.

I.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.

I believe that palynologists in general would understand the distinction here. There are two levels of identification that matter for CREST: at the species level (to fit species pdfs) or not at the species level (to fit taxa pdfs). However, Reviewer 2 has also requested that I clarify some of these details at the beginning of section 2 where I present the species and taxa PDFs. I have explained this in greater detail there, and I hope the PSE table can now be better understood with these additional details.

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?

The level of identification of fossil proxies depends on many factors, including the type of proxies. For example, vegetation macro remains can most often be identified at the species level, while pollen grain cannot. Tree pollen grains can often be identified at the genus level, while grass pollen grains are most often identified at the family level. Other factors also influence this, including the palynologist (experience increases resolution), fossil preservation (the more damaged the fossil the harder it is to specifically identify it), and many more. Lots could be written on the topic and I believe this is standard knowledge for people using this type of data. As such, I will not expand on this in the manuscript.

I.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?

To fit reliable *species* PDFs. This is now clarified.

I.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 accurately the purpose of the functions.

What Reviewer 1 is referring to here is the reconstruction, which is not what this section is presenting. Here, we are presenting how the proxy-climate responses are fitted. The term PDF for this type of reconstruction was not pinned down by me, and I am simply using terms that are used in the community.

I.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?

Absolutely. I have now added a pointer to Eq. 6 to make sure readers will be able to make the connection as well.

I.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.

geographical location is exactly what it is. The term study area is now defined at the beginning of section 2 (along with presenting the types of data required), and I have fixed the sentence based on Reviewer 1's suggestion.

I.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

Corrected.

I.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.

Now defined as follow: "*assessing the risks of confounding variables (i.e. variables correlated with important variables but do not directly impact the studied proxies; @Juggins2013, @Chevalier_etal_2020)*"

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?

Yes, it is. After consideration, the title was not changed.

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

Thank you for noticing these missing labels. I have been so used to seeing these figures (and I have generated so many versions of them...) that I became a bit blind to some of these obvious issues. Labels are now added to the different axes of figures 6 and 7 (and the package updated accordingly).

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.

Corrected.

'taxaCharacteristics' graphical diagnostic tool to assess the sensitivity of taxa to climate. In the top row, the map represents the density of unique species occurrences per grid cell **derived from the modern calibration dataset** and the **time series represents the variability of the taxon against time or depth**. The bottom row is specific to each selected variable (only one row in this case), where the map represents the geographical distribution of the taxon (in white) against the climate background, the histogram represents the climate space (in colour) and how this climate is sampled by the taxon (in black) and finally the right plot represents the climate response of the taxon (in black) as well as the response of all the composing species (in grey). This diagram shows that Ericaceae is preferentially associated with colder temperatures in the study area. The red diamond indicates the location of the studied record.

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

Corrected. See above.

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.

I do not think users generating the plots with their own R code would be confused. Figure 7 is specifically generated for 'Ericaceae' (see code for plot_taxaCharacteristics), while figure 6 is designed for the calibration dataset as a whole.

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.

I disagree with this statement. This type of figure is probably the one that I use the most to better understand the sensitivity of my taxa. I also think that all my figures are complementary, and this one should be used in conjunction with the climate space plot (fig. 6). In addition, this plot becomes also quite interesting when more than one variable is selected (not shown here) because it allows seeing where the distribution is in relation to the underlying climate gradients. While trying to plot an outline of the distribution might seem like a good idea, in practice the data are often too disconnected to lead to anything visually convincing. It would only add more complexity, if not noise.

I.429 (all the probabilities sum to 1)

Maybe better: the PDFs integrate to 1.

While this is more correct, I prefer to keep the wording a bit more accessible to my target audience, which is not necessarily expected to be well-versed in the statistical language.

I.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.

'Objective interpretation' was a poor choice of words, indeed. Replaced with 'more informed interpretations'.

I.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).

This exclusion is based on ecological considerations of the taxa, *i.e.* what is known about them. It is independent of the shape or look of the associated *pdf*. I think it might in fact be more misleading to plot them because they might not appear widely different from the others.

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.

In this example, the '64' indicates the age of the sample. It is derived from the *df* dataset. I have now clarified in the caption what it means.

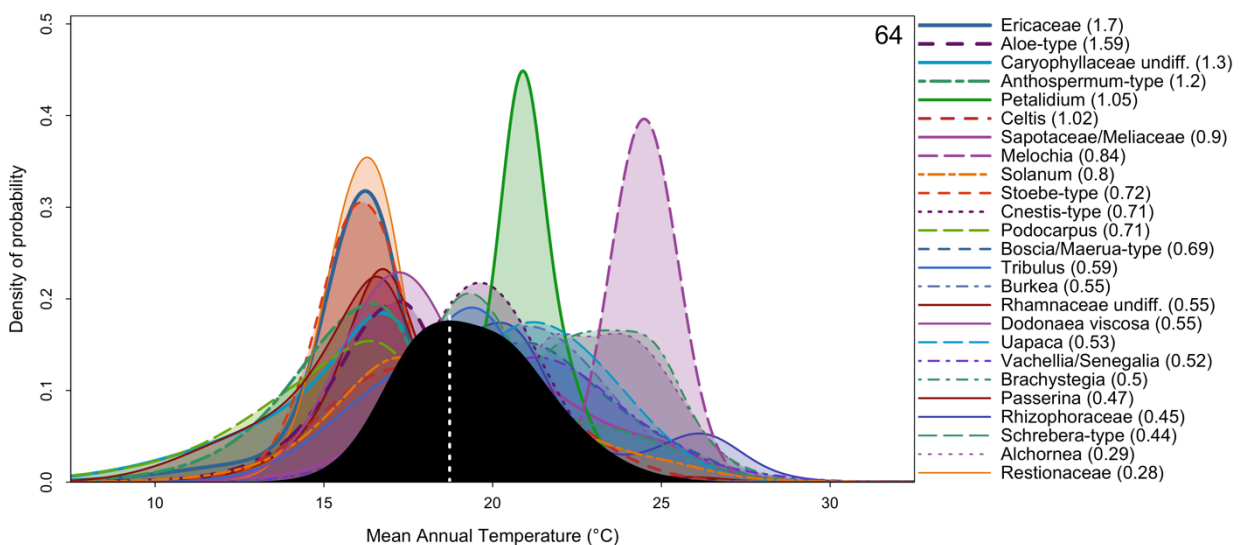
The black curve represents the posterior MAT reconstruction, from which a 'best' climate estimate can be estimated from the maximum of the curve and uncertainties derived by calculating the area under the curve.

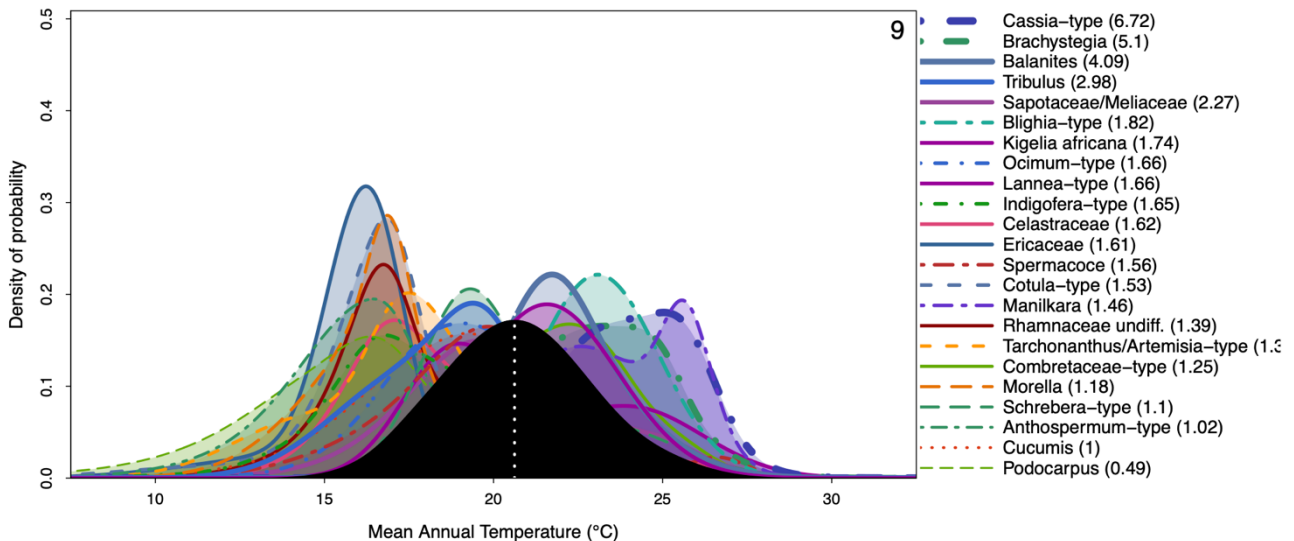
From what I understood from Section 2., the black curve does not show a posterior probability distribution for the past climate but instead the likelihood function.

This has been corrected.

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?

They were actually plotted according to the log of suggested weight, which I agree did not lead to the best results. I chose this option because using the absolute weights leads to unpleasant and very crowded graphics, as the lines can get very thick in some cases. I have tried different options and I am happy with $1.5 \cdot \sqrt{\text{weight}}$. It gives a reasonable thickness range without overfilling the graphics. Considering that the taxon name information is also colour coded and included in the type of dashed line, I think the results are satisfying. See some examples below (age 64kyr and 9kyr)





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. “

Rephrased as follow: “In this plot (Fig. \ref{fig:samplePDFs}), the PDFs of all the taxa present in the sample and selected to reconstruct the climate variable are represented along with the reconstruction. This type of plot can help identify if a particular PDF is at odds with the general assemblage, which usually indicate the possible presence of a confounding factor. It is also helpful to visualise the full spread of the uncertainties and, by extension, highlight that reconstructions can be multimodal. **While it is not the case in the example here, multimodality can be the underlying cause of apparent noise in the reconstructions with minor changes in the taxa composition or percentages forcing the system to oscillate between two maxima and thus 'appear' noisy.** This effect could also be seen from the reconstruction plot where the full uncertainties would be represented (`simplify=FALSE`).”

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

This is due to a lack of clarity of the caption. The plot represents the absolute anomalies with the bars, and the anomalies' sign is colour-coded. This was designed to save vertical space. As a result, the plot represents both positive and negative LOO anomalies. I have now modified the figure with a legend that specifies what the colour means. I have also clarified it in the caption.

Leave-one-out (LOO) graphical diagnostic tool to illustrate the influence of different taxa on the reconstructions. Here, the results are only shown for a subset of the taxa observed in marine core MD96-2048 (only 20 out of the 171 available taxa are represented). The height of each bar represents the absolute effect (in °C) of removing the taxon from the reconstruction, and the sign of this effect (increase or decrease of the reconstructed temperature) is colour-coded. Here, blue (cf. Ericaceae) and red (cf. Combretaceae-type) taxa are cold and warm indicators, respectively.

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.

Corrected.

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.

I have adapted the package so that the name of the plots will now be 'Leave-one-out anomalies for [studied variable (variable unit)]'.