I am really grateful that Reviewer 1 looked at my manuscript carefully for the second time and agreed with the proposed rewritten sections. The new suggestions were also quite valuable. However, he/she is expressing some concerns regarding the 'understandability' of the manuscript in places (e.g. the data used). I believe these concerns will not directly apply to colleagues from the palaeoecological community. I understand that Reviewer 1 is from another field, which probably explains some misunderstanding. Lots of field-related 'known unknowns' are associated with palaeoecological datasets, and these can be hard to grasp for colleagues from different fields. This paper cannot, however, start explaining these elements. Otherwise, the size and focus of this manuscript would change too much. But I regularly refer to a recent and exhaustive review I wrote on the topic (Chevalier et al. 2020b), which I believe to be sufficient. Any element I may add to this manuscript to 'explain' palaeoecological datasets would only be a repetition and/or simplification of what is in the review.

I hear his/her concerns regarding the possible mathematical inaccuracies or rather shortcuts. But as he/she mentioned, he/she was able to derive the same results from a complex mathematical suite of Bayesian equations. This is thus a strong validation of the simple equations presented. The goal of this paper is not to be technical but rather to make the use of the technical method appealing to a broad range of potential users, as illustrated by the relative length of Sections 3 to 5 compared to Section 2. Section 2 was only meant to summarise an existing technique and put the package's different parameters in context. By no means did I intend to make a detailed theoretical description. I also want to emphasise that this paper is about the R package and that all the other elements are already exhaustively described in other articles: the method itself (Chevalier et al. 2014), the calibration dataset (Chevalier 2019), the characterisation of pollen taxa climate sensitivities (Chevalier et al. 2021b), and different papers with direct applications (Chevalier and Chase 2015, 2016, and a few more by myself and independent research groups). I hear Reviewer 1's concerns, and I acknowledge that the model presented here could be improved (in many ways). All models can. I am also grateful for all the efforts to reframe my model on more grounded foundations. These results and, more broadly, this discussion through the review process is something that I will consider for future versions of the model or package (as suggested by Reviewer 1). But at this stage, I prefer finalising the package part of the 'crest project', which consists in creating a tool for the broader community to more easily apply an accepted and already used method. Changing the methodology, even by a little, would imply starting a complete phase of development and testing phase to be validated. I believe this to be the work of an entirely new paper.

The comments made by Reviewer 1 are in black and purple, and my responses are in blue.

General comments:

The author has presented a revised version of the manuscript with substantially improved readability and many inaccuracies being cleared out with respect to the original version. It would like to point out again, that probably the lion's share of the work has gone into the code of the R-package and the curation of the database and that this work is of highest value for the community.

On the other hand, I must admit that I am still not convinced by the presentation of the mathematics forming the bases of the climate reconstruction method. This criticism comprises twofold: First, the introduction of Equations (1-5) and (7) suffers from several mathematical inaccuracies – however, I noticed that these in parts trace back to the article presented by Kühl et al. (2002). Since the derivationof these equations is by no means the central aim of this paper, these inaccuracies might be acceptable, given that these equations appear to be correct. Also, it seems that some terms have established as domain specific language and thus may be clearer to the target audience than they are to me. Second, I believe that Equation (6) is in fact not correct, even though its use might generate reasonable results. I have expressed my concerns about Eq. (6) already in my previous review.

If not for this manuscript, I am convinced that putting the probabilistic climate reconstruction method on solid ground mathematically would be a beneficial task for the future. I have attached a pdf that outlines the derivation of Eq.(5) starting from Bayes theorem.

The remainder of the mansucript gives the reader a good overview of the *crest* R package, in terms ofits capabilities, requirements and usage. It is well structured and the final example of application really takes the reader / user by the hand.

Specific comments:

l.2 In particular, the methods based on probability density functions (or PDFs) can be used invarious environments and with different climate proxies because they rely on elementary calibration data (i.e. modern geolocalised presence data).

I would replace 'the methods' by 'methods'. Maybe 'methods based on probability density functions' are just 'probabilistic methods'.

I believe 'the' to be correct because I want to specify a subset of methods from the broad range of existing techniques. Both are grammatically correct, but 'the' includes a nuance that I want to have. And I will keep on referring to 'methods based on probability density functions' because this is how the community knows them.

l.14 It is hoped that crestr will be used to produce the much-needed quantified records from the many regions where climate reconstructions are currently lacking, despite the availability of suitable fossil records.

What is meant 'quantified records'? In my understanding a record is a directly measured time series – so a data processing software could not be used to 'produce a record'? Do you mean 'reconstruction'?

Corrected as follow: "*It is hoped that crestr will be used to produce the much-needed quantified climate reconstructions from the many regions where they are currently lacking, despite the availability of suitable fossil records.*"

l.15 no paragraph in the abstract

Paragraph removed, and sentence linked as follow: "*It is hoped that crestr will be used to produce the much-needed quantified climate reconstructions from the many regions where they are currently lacking, despite the availability of suitable fossil records. To support this development, the use of the package is illustrated with a step-by-step replication of a 790,000 year long mean annual temperature reconstruction based on a pollen record from southeastern Africa*."

l.19 Over the years, numerous techniques of increasing complexity have been proposed, each one based on a unique set of assumptions regarding the modelling of ecological datasets and their translation into climate reconstructions (e.g. Birks et al. (2010), Chevalier et al. (2020b)).

I assume that with 'ecological datasets' your refer to observations. In that case I would say that observations are not being modeled. Of course, sometimes one uses models to draw inference from datasets and probably that is what you mean?

I do not understand the nuance Reviewer 1 is making here. Any statistical model that exists is derived from observations based on certain assumptions about these observations and is then followed by the modelling of said observations. I did not modify that sentence because its meaning will be clear to the community.

l.26 ...and their accessibility with multiple software solutions.

Saying that an analysis technique is 'accessible with multiple software solutions' sounds strange to me. Do you mean, there exist relatively simple software implementations of the techniques?

Replaced 'with multiple software solutions' with 'via multiple software tools'.

l.26 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 (Chevalier et al., 2020b).

Again, what are quantified climate records? It seems you mean reconstruction – in my understanding records are not reconstructions.

'Records' replaced by 'reconstructions'. Here and elsewhere.

l.32 Because modern occurrence data are generally easier to obtain than modern proxy assemblages, this fundamental difference implies that Indicator species methods can contribute to filling in the reconstruction gaps that exist at the global scale.

Grammar: Modern occurrence data are generally easier to obtain than modern proxy assemblages. This fundamental difference implies that Indicator species methods can contributeto filling in the reconstruction gaps that exist at the global scale.
For the non-paleoecologists: what's the difference between proxy assemblages and proxy occurrence data?

The community will understand proxy assemblages, and the meaning of occurrence data is explained in the previous sentence. I did not modify the sentence.

l.36 Derived from the original work of Kühl et al. (2002) —– whoIt

seems there is an extra hyphen in the pdf.

Corrected.

l.37 CREST estimates and combines probabilistic proxy-climate relationships to reconstruct past climate parameters from fossil proxy observations.

Maybe you could add: CREST estimates and combines probabilistic proxy-climate relationships **from modern occurrence data** to reconstruct past climate **variables** from fossil proxy observations.

This sentence is self-explanatory within its paragraph. The definition of the relationships from modern occurrence data is defined in the first part of the sentence (not reported by Reviewer 1, but see original or revised text).

l.45 However, the complexity of collating and formatting the thousands of distinct occurrencesrequired to estimate reliable PDFs limited its practical use.

I understand, that in your context the term PDF carries a very specific meaning. However, in general, this is not the case and a reader that is not used to the specific use of the term PDF as a synonym to your 'climate response functions' will probably struggle to understand the above statement. Also, you introduced the abbreviation PDF only in the abstract, but up to this pointnot in the main text.

I actually wanted to remove the use of the acronym 'PDF' before the methods section – thank you for noticing this last occurrence. I have replaced it with 'proxy-climate relationships' to keep the naming coherent across the text.

l.48 ‘climate records’

Corrected.

l.50 This paper thus introduces ~~a~~ the new multi-platform R package crestr designed to replace the original interface.

Keeping ‘a’.

l.51 crestr includes the global calibration dataset

It is linked to the dataset but does not include it – strictly spoken.

‘includes’ replaced with ‘integrates’. The nuance Reviewer 1 wants to make is correct, but I think it is also important to highlight in a simple way that the package comes with calibration data. Practically, many tools that exist are difficult to use because of this absence of calibration data. I do not want to lose this important point to indicate early on that the calibration is in a cloud-based database and not with the package itself. The information is available in the relevant section.

l.59 As such, the climate reconstructions obtained from CREST can be understood as an ensembleof all data-compatible climate values

Maybe ‘As such, the application of CREST yields a probabilistic quantification of the past climate in view of the data under study as opposed to simpler, less informative ‘most likely’ or ‘best’ climate estimates. While the latter may only capture statistical uncertainties, the former rigorously takes into account the large quantitative uncertainties inherent to analysis of this kind.’

Just to highlight a little bit the fact, that in terms of uncertainty propagation *crest* performs a lot better than the mentioned ‘best-estimate’ methods.

I like the first sentence suggested by Reviewer 1. The second one is also very interesting from a technical perspective but I fear that it would raise more questions. I have thus rephrased the sentence as follow: “*As such, the application of CREST yields a probabilistic quantification of all the climate values that are compatible with the studied data instead of simpler, less informative ‘most likely’ or ‘best’ climate estimates. While the 'best estimate' approach might be optimal when a fossil assemblage is analysed in complete isolation, the presence of independent - local or regional - information (_e.g._ other reconstructions from the same core or independent records) usually provides additional information that may not always be consistent with best estimate reconstructions.*”

fig.1 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 (e.g. CREST; dark grey), and a modelling approach focused on the estimation of the ’most likely’ or ’best’ climate value with small statistical errors surrounding it (e.g. MAT or WA-PLS; lightgrey).

I would say the probabilistic approach is based on the ‘full spread of the data’ but not on the ‘estimation of the full spread of the data’.

Corrected.

l.96 The **individual?** climate responses of all the species identified are estimated as univariate probability density functions (PDFs) for every climate variable.

Yes, corrected.

l.99 The individual species'

Corrected as 'The species' individual climate responses are' for clarity.

l.99 estimation of the empirical mean (m s,c)

An empirical mean is not estimated, but computed from the data. Just delete 'the estimation of' .

Corrected, indeeed.

l.107 Here, the weights are calculated by first sorting the N climate values (all the c i) that composethe modern climate space into bins of equal width (e.g. 2 ◦ C or 50 mm). Then, each climatevalue c i is given a weight k(c i) defined as the inverse of the relative size of the bin c i itbelongs to:

1. As far as I understand N is the number of gridded observations of the modern climatevariable c, while N_s is the number occurrences of the species S. According to line 100, the c_i are only those climate values, that coincide with the occurrence of the species. However, for the weights k(c_i), the entire climate space should be taken into account. So the addition (all the c_i) is not correct. Maybe, the difference between the c_i and **all** climate values form the study area can be clarified somehow.

I have tried to improve the labelling of all the quantities used in this modelling. For example, I defined Cs,i as the climate values where species s is observed (i varies between 1 and Ns).

2. The first sentence defines all bins such that they have equal width. The second sentence refers to different 'sizes' of the bins. I assume that the 'size' of a bin here means the number of observations falling into one bin. However, typically one would use 'size' and 'width' interchangeably, so 'size' might be misleading, here.

Rephrased as: "*defined as the inverse of the number of values that belong to same bin $\text{bin}_{c_{j}}$:*".

l.117 Once estimated, m s,c and s 2 s,c are used to define a regular, unimodal distribution for the PDF sp (s, c) of species s for climate variable c. Here, we assume that the shape of these species responses should be unimodal and can be either normal:

I know it is a detail, but the term 'distribution' is slightly misused in this context. The distribution of a random variable in statistics defines the the probability for the random variableto assume a certain value. Hence, here the PDF **is** a distribution.

Maybe: 'Assuming unimodality and either normality or log-normality, the estimated m_(s,c) and s^2_(s,c) are used to define the species climate response PDF's as follows: …'

Rephrased as: "*Once estimated $m_{s,c}$ and $s_{s,c}^2$ are used to define a regular, unimodal $\text{PDF}_{\text{sp}}(s,c)$ for species s and climate variable c.*"

l.120 delete paragraph

No.

l.125 I have expressed my doubts about this equation already in my previous review. After going through the math once again, I still belief this equation is not correct – even though it might lead to reasonable results. Also in the cited literature, I could not find a convenient derivation of Eq.(6).

I have acknowledged Reviewer 1's concerns in my previous response, and this is currently under investigation. This model can probably be improved – every model can – but this is beyond the scope of the present paper, which consists in presenting an implementation of a model that has been successfully used for many years now.

l.126 delete paragraph

No. I really do not understand why Reviewer 1 would suggest this deletion. Defining all the parameters and transformations is important. In addition, I believe this paragraph (and the ones below that Reviewer 1 also suggests to delete) to be very important for non-statisticians to understand what things are, and also what they are not.

l.138 Climate c is reconstructed from fossil sample z (z can be an age, depth or any identifyer) by multiplying the PDF tx (t, c) of the T (z) selected taxa:

- Maybe 'Past climate c that corresponds to a specific age (or depth) z from which a fossil sample is available can be reconstructed…'

This suggested phrasing will appear convoluted to colleagues with palaeoecological datasets. The working unit is the sample, and it is more often than not irregularly sampled along with age and depth. I am not convinced that indexing on depth would make sense, especially since many depths will have no data associated. The idea here is only to indicate that it is applied for each sample. I believe that too much mathematical formalisation can sometimes add unnecessary complexity, especially for non-experts, which ultimately masks the intended message.

- you first assign z := the fossil sample and then state that either z:= age, or z:= depth.
  - what exactly is 'the fossil sample'? A dataset comprised of abundance data ofdifferent taxa on a depth or age axis? This is a little inaccurate.

Rephased as: "*Climate c is reconstructed from fossil sample z (itself associated with a unique age, depth or any other identifier) by multiplying the $\text{PDF}_{\text{tx}}(t,c)$ of the $T(z)$ selected taxa*:"

l.141 delete paragraph

No.

l.151 The summation index should be 'i' or sth else and then the depths / ages should be discretizedas z_i, but the summation index cannot be a continuous variable. The inner parenthesis in the denominator are not required.

'z' indicates the samples across the paper. It is definitely not a continuous variable (see response aboce). I do not want to add another layer of variable indexing – this would be too complex for something that is fairly simple: sum across all samples.

I have actually removed the outer parentheses. I prefer keeping the inner ones for clarity even if they are not mandatory.

l.152 delete paragraph

No.

l.181 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 and Hijmans (2017), Zomer et al. (2008), Locarnini et al. (2019), Zweng et al. (2018), Garcia et al. (2019a), Garcia et al. (2019b), Reynolds et al. (2007), see details in Tables 1 and 2).

Does this mean, that all variables listed in Tab.1 and Tab.2 are reconstructable? If so, I would suggest to express this a bit more clearly and also emphasize this in the table captions.

Clarified in the captions as: "*List of terrestrial/marine variables available for reconstruction in the \textit{gbif4crest} database. Each one can be selected in \textit{crestr} using its associated code. List of abbreviations: (Temp.) Temperature, (Precip.) Precipitation*.".

l.193 For example, the first version of the gbif4crest dataset released in 2018 contained about 17.5 million QDGC entries, while the new version contains approximately 25.3 million entries (~44% increase).

Why QDGC entries and not only entries? The second is a duplication of the first sentence of the paragraph and can be deleted, except the (~44% increase).

Only the QDGC data are used for the constructions. The sentence was corrected as suggested.

l.230 Maybe, it would be helpful to explain, which function call initializes the crestObj in first place already at this stage?

Sentence corrected as follow: "*In crestr, all the CREST-related data are stored within a single S3 object of the class crestObj that is first initialised by either crest.get_modern_data or crest.set_modern_data (see section 5.2 for details)*.".

l.257 Does the df have as many columns (+1 for the depth/ age) as fossil taxa are considered for the climate reconstruction?

Absolutely. See text with "with either the age, depth, or sample ID as the first column and the fossil data in the subsequent columns." The package also contains an example dataset, and there is the Limpopo dataset also shared as supplementary material. There is also online documentation to help people build their datasets https://mchevalier2.github.io/crestr/articles/get-started.html.

l.280 In a second time

In a second step

Replaced with 'in subsequent steps'.

l.281 Mabe here, a sentence like

The pdf for the fossil taxon Stoebe-type is thus comprised of the linear combination of species pdfs according to equation (6) associated with all species that fall into the geni Stoebe and Elytropappus.

would be helpful at this stage, provided that my interpretation is correct.

This is a good suggestion to make this step more relatable to the data modelling. I have clarified this at the

beginning of the PSE section.
"The PSE data frame is required to use the gbif4crest calibration dataset. It is used to associate individual species available in the TAXA table with their corresponding fossil taxon. This step is important to estimate the species responses (PDFsp(s,c)) and taxon responses (PDF_{tx}(t,c)) described in Section 2.2."

l.284 Additional taxa can also be added to the PSE file to exclude species known not to be part of a group. For instance, this 'trick' could have been used to simplify the climate response of the 'Asteraceae undiff.' group by excluding more species from it, even if the pollen grains corresponding to these species have not been observed.

I would suggest to move the 'even if… ' to the first sentence.

Corrected.

l.290 If I understand correctly, if I provide a 'distributions' table as an input to the get_modern_data() function, then I do not need the PSE is that correct? Why is the 'distributions' table not listed in Fig. 5 in the 'input' category? If users decide to use the 'distributions' table as input, then they have to provide climate_space data frame as well, is that correct?

It would be an input to 'set_modern_data()' but that is correct. The distribution data go into the 'modelling' part of the object – same as the other ones – because they are used to infer the climate responses. This is imperfect, but the possibility of using other datasets was a late addition to the package. Users will have to add a climate_space dataset if they want to use the weighting option and plot maps. I would certainly recommend using it, but it is not mandatory. I have made this distinction more obvious.
"Including a `climate_space` dataset is recommended, even if it is not mandatory."

l.392 here called rcnstrctn; Fig. 5

here called rcnstrctn and whose structure is displayed in Fig. 5

Corrected.

l.393 Alternatively, the function crest.set_modern_data() could be called instead of crest.get_modern_data() to use personal calibration data instead of the gbif4crest database.

This could already be mentioned in 4.3.3 and 4.3.4.

I tried to avoid having too many cross-references across sections. In my opinion, it would only complexify the first reading of the paper. And in practice, knowing which functions will be used is not necessary to build the correct datasets. In addition, there is online documentation with further details. An R package is a complex tool and accounting for all possibilities is difficult and probably not recommended. I present a way of running such analyses, and if people follow each step one by one they will obtain results.

l.449 Ideally, the climate sampling should be as homogeneous as possible to ensure proper samplingof all the possible climate values, even if the extreme climate values will always be under- represented compared to the median ones. However, deviations from a theoretical one-to-one (or at least proportional) equivalence between climate and occurrence data abundance are not necessarily a bad characteristic.

I know this has not changed much with respect to the previous version of the manuscript, yet I must admit I do not fully understand these sentences.

What exactly is 'the climate sampling'? If c is the climate variable to be reconstructed, Iassume, the term refers to the entirety of values for this variable present in the study area, irrespective of the presence or absence of species used for the reconstruction. This relates to my comment with respect to line 107 – a clear distinction between 'the climate space' that comprises all N climate values and the 'the climate sampling' which is comprised only of theN_s climate values accompanied by species occurences (?) would be helpful.

A definition of what is meant by climate space is already provided in section 2.1. Then N and Ns are defined in section 2.2 with N corresponding to the size of the climate space and Ns being the climate values where species s is observed.

If that interpretation holds true, does the climate sampling contain multiple instances of a climate value c_i if the corresponding grid cell contains multiple independent occurrences of the species?

By construction, every species is only observed once in a grid cell. So the Ci are unique. However, if several species part of the same pollen taxon are observed in the grid cell, the Ci will be counted once for each.

What is meant by a 'one-to-one equivalence between climate and occurrence data abundance'? Would that be something like, the warmer the clime, the more occurrence data there is in the study region?

The paragraph was rephrased as follow to make it simpler. "*Ideally, the climate values sampled by the calibration data should be as homogeneous as possible to ensure proper representation of all the possible climate values, even if the extreme climate values will always be under-represented compared to the median ones. However, deviations from a theoretical equivalence between the observed climate distribution and the climate values sampled by the calibration data are not necessarily a bad characteristic. In our case study, the variability of the sampling density represents actual patterns in regional species diversity with the presence of several biodiversity hotspots across the mountainous regions of eastern and southern Africa [@Myers_etal_2000]. This higher diversity in the colder areas explains why the black histogram (_i.e._ the climate values associated with occurrence data) on Fig. \ref{fig:climatespace} is skewed towards the left compared to the grey histogram (_i.e._ the distribution of the climate space in the study area). All these elements should be checked and accounted for while designing the final calibration dataset.*"

l.451 In our case study, the spatial variability represents actual patterns in regional species diversity with the presence of several biodiversity hotspots across the mountainous regions of eastern and southern Africa.

Spatial variability of what?

Corrected as "the variability of the sampling density". See above.

l.457 (i.e. variables correlated with important variables but do not directly impact the studied proxies; Juggins (2013), Chevalier et al. (2020b))

i.e. variables which are correlated …

Corrected.

Fig.6 number of unique species occurrences

I understand that the author does not want to change the title of the right subplot, however, I would appreciated a lot a clarifying not in the caption, that unambiguously defines the term ‘number of unique species occurrences’.

The term is first defined in the introduction, further explained in section 2.1, and these data represent the core of the entire approach. I am confused by the fact that Reviewer 1 is still unclear about what these data are about. He/She seems to have understood them correctly though, especially since he/she has generously offered a Bayesian translation of my model.

l.508 Here for instance, both Aizoaceae and Chenopodiaceae/Amaranthaceae were excluded because they are not primarily sensitive to temperature in southern Africa.

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

Answer by the author: 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 itmight in fact be more misleading to plot them because they might not appear widely different from the others.

This does not make sense to me. If these two taxa are not sensitive to temperature, this shouldbe reflected in the data and this be visible in their climate response (or PDF), which in the violin plot should appear more stretched along the y-axis. In fact, the violin plot is advertised a tool to assess the climate sensitivity of the different taxa – the authors answer is not consistent with this role of the violin plot.

This is exactly what confounding factors are. Due to modern correlations between unrelated parameters, they can both “look like” they are good predictors. But independent ecological knowledge tells us that their presence is related to other factors, e.g. evaporative conditions and relatively saline environments. The violin tool is advertised as one of the tools that can be used to make such an assessment. I also strongly refer to Fig. 7 to help identify these sensitivities (what does the complete distribution look like? When is the taxon present/absent/abundant? Are different variables correlated? etc). Finally, the reference Chevalier et al. 2021b, which is mentioned several times in the paper, is entirely focused on this specific aspect.

Pollen data are really complex data and this type of assessment is commonly made by palynologists. It is probably the limited inside knowledge that makes this surprising or confusing to Reviewer 1.

Fig.11 Couldn’t the strong Loo values for the Ericaceae already be interpreted as some sort of bias inthe sense of what you say in line 583?

Large LOO values can arise when the PDFs are biased by unaccounted factors and are, as a result, at odds with the rest of the PDFs.

It could in theory, but it doesn’t in this case. The ecology of Ericaceae is well-known and it is a taxon that thrives in the tropics during cold periods. This type of analysis is really an interplay between statistics and ecology, and both should always be accounted for to define the assumptions and interpret the results.