Dear Alberto Reyes, dear reviewers,

In addition to the more general Author Reply we submitted earlier in this discussion, please find a detailed point-by-point rebuttal to the comments raised by the reviewers below. We hope that our replies to the reviewers' concerns and the revisions we propose will be adequate to merit publication of our manuscript in Climate of the Past. In the text below, we indicate review comments in red and our replies in **black**.

This is a valid suggestion by the reviewer, and we agree that a combination of the low-precision clumped isotope method with (potentially) lower accuracy $\delta^{18}O_c$ measurements could yield an optimal compromise between the pros and cons of these methods. In fact, our "optimization" approach is based on this idea. However, the approach of combining bulk clumped isotope measurements and high-resolution $\delta^{18}O_c$ measurements has disadvantages, especially in cases where the $\delta^{18}O_{sw}$

seasonality is in antiphase with temperature seasonality (see comment on line 272 below). In addition, larger bulk measurements for clumped isotope analyses may significantly average temperature and $\delta^{18}O_{sw}$ reconstructions of summer and winter seasons compared to microsampled transects. We will clarify the advantages and disadvantages of different methods for combining clumped and $\delta^{18}O_c$ measurements in seasonality studies by introducing the approach by Keating-Bitonti et al. (2011) in the introduction and comparing it with our approaches in the discussion (sections 5.1).

(2) Though beyond the scope of the variables considered here, there are two additional complicating factors in real-world sclerochronologic data that may be worth mentioning at some point: (2a) Unequal sample time-averaging as a function of growth rate - in practical application (versus virtual subsampling), the number of days (or weeks or months) averaged into a single sample will vary; when the organism is growing quickly the sample will represent comparatively less time than when it is growing slowly, using the same diameter drill bit. If growth slowdowns/hiatuses, e.g., correspond with winter extremes, this results in not only fewer but also more time averaged (and thus dampened) estimates of winter temperature. (2b) Uncertainty in the seasonal phasing of SSTs in paleoclimate studies – even in instances were samples from a fossil specimen can be aligned along some reliable age model (be it via growth band counting or a statistical model), assigning those data points a calendar date is ambiguous at best. The timing of the date of maximum and minimum SST can vary considerably based on latitude, environment, and other local factors. This is a particularly important source of uncertainty to consider when binning the data by month – shifting the calendar date assumptions by, e.g., 15 days, would result in a whole new grouping of monthly data points and could significantly alter the results.

These are very good suggestions and we gladly incorporate these considerations in our revised discussion. Since both these complications relate to the timing of growth of sampled intervals, we will incorporate them into the section discussing the effects of uncertainties in the age model (section 5.2.2).

(3) Regarding the figures, I really appreciate the consistent color theme throughout, but for accessibility purposes you may want to double check that the color scheme is colorblind safe. Additionally, there is a lot of really useful information that is not always easily extracted from the figures – I'd encourage you to think about creative ways of graphically presenting the information that will allow the reader to quickly glean the most important aspects. For example, you could include a heatmap showing how accuracy and precision vary by case and climatological parameters (similar to the colored conditional formatting of S12), which may be easier to interpret than the line plots.

We appreciate the suggestion by the reviewer and paid special attention to the colour scheme of our figures, which was also a concern raised by Referee #2. We now use the RdYlBu 4-colour scheme provided by Colorbrewer 2.0 to make our figures colour-blind- and print-friendly (see https://colorbrewer2.org/?type=diverging&scheme=RdYlBu&n=4). To be honest, the multi-dimensional character of our results caused us some difficulties in coming up with a proper presentation method for aggregating all results in overview figures (Figures 6 and 7). We appreciate the reviewer's suggestion for revising our presentation method for these figures and agree that a colormap may be a more intuitive way to visualize this data. We will experiment with this idea and try to come up with a more reader-friendly visualization of the data now shown in figures 6 and 7.

(4) I'm not convinced that Section 5.4 (Implications for other sample size problems) C3 sufficiently contributes to the overall content of the manuscript to warrant inclusion. This section reads like a grant proposal rather than a discussion and could easily be condensed into a single one-paragraph section briefly outlining potential additional applications of the approach.

Both reviewers expressed doubts about the added value of section 5.4 in its current form. We therefore decided to shorten this section into one paragraph, as suggested by Referee #1 as opposed to deleting the entire section following Referee #2.

(5) Prior to the final submission, the manuscript, supplement, and figures all require a thorough read through to ensure consistency of terminology, case numbers, and color scheme. For example, high precision/accuracy are at times conflated with high reproducibility error/offset in the text. Case numbers are not always consistent (both in the text and supplements), and it appears as if the colors have been switched in at least one figure. I've tried to point out examples of these in the line-by-line comments, but I'm sure that I've missed some. The supplemental files are hard to navigate - it would be helpful to include a 'Read Me' tab in applicable excel files defining all acronyms and abbreviations and descriptions of the information provided in each subsequent tab.

This is a very valid concern, and we apologize for any mistakes that may have crept in our terminology or colour scheme in the initial version of the manuscript. For the revision, we will go through the manuscript, figures and supplements in detail and make sure all references and colours are consistent. In response to this comment and related comments by Referee #2, we will also more clearly define our use of terminology for describing reliability of reconstructions ("accuracy", "precision", "offset", etc.) and provide more detail in the supplements explaining the data tables. In addition, we have now compiled the R scripts of our reconstruction approaches into an R package. This package, including documentation of all scripts, is now freely available online through the CRAN database (https://cran.r-project.org/web/packages/seasonalclumped) to ensure easy access to the reader. We will cite the package in the revised version of the manuscript.

Specific Comments.

*NOTE: line numbers restart after L347; second set of line numbers indicated by LN

Our apologies for the break in the line numbering, we will amend this in the revised version.

Text:

L12: The term 'events' is a bit ambiguous here, particularly since the manuscript focuses on recovering climatological parameters (i.e., multi-year averages that smooth 'events')

We acknowledge that this statement may be ambiguous because the length of an "event" may depend on its (broad) definition. To keep referring to the compromise between achieving accurate climate information by combining measurements on the one hand, and retaining a high temporal resolution in reconstructions on the other hand, we propose to rephrase as follows:

"The challenge is to isolate meaningful information on climate variability from these records by reducing measurement uncertainty through a combination of proxy data while retaining the temporal resolution needed to assess the timing and duration of variations in climate parameters."

L54: 'the seasonal cycle is the most important cycle in Earth's climate.' This is a bold and rather subjective statement – I can think of many who might argue the carbon cycle is equally important. I'd suggest changing 'most' to 'one of the most' or omitting the statement all together.

We understand this comment and will rephrase according to the reviewer's suggestion. The point we wanted to make is that the annual cycle is the most dominant cycle in climate variability when compared to cycles on other timescales (e.g. diurnal, orbital or tectonic timescales), but we realize that this is quite technical and distracts from the main point made in this sentence. The suggested rephrasing by the reviewer is the more elegant solution.

L91: Colon missing after 'Optimization'

We will add the colon here

L120: Consider changing 'depth domain' to 'sampling domain' (here and in all subsequent references); ontogenetic trajectories aren't necessarily depths

Good point, we will rephrase this here and throughout the manuscript.

L121: Delete 'this' L139: There's an error in this equation ($\delta18O_{sw,freshwater}$ is repeated twice). The mass balance should read: $\delta18O_{sw} = f* \delta18O_{sw,freshwater} + (1-f)*\delta18O_{sw,ocean}$

Very good, we apologize for the mistake and thank the reviewer for pointing it out. We will rephase the second $\delta^{18}O$-term to $\delta^{18}O_{w,\,ocean}$. In addition, we will rephrase all mentions of $\delta^{18}O_{sw}$ ("$\delta^{18}O$ of the seawater") to $\delta^{18}O_w$ ("$\delta^{18}O$ of the water") to avoid the internal inconsistency of labelling $\delta^{18}O_w$-values with "sw, freshwater".

L142: Space between the per mille symbol and VSMOW; this information is repeated on L145

We will add the space between the symbol and "VSMOW" here and throughout the manuscript and remove the duplication in L145 as follows:

"…in the nearby Elbe and Weser rivers (see Ullmann et al., 2010)."

In addition, we will mention here that all $\delta^{18}O_{sw}$ values are given in ‰VSMOW throughout the manuscript to remove the mentions of VSMOW further in the text (see comment to L159)

L157: Change 'is' to 'are'

We will rephrase this accordingly

L159, etc.: I'd suggest stating upfront that all references to seawater composition are in reference to VSMOW; the VSMOW reference after each composition is a bit clunky

A good suggestion, and we will define $\delta^{18}O_{sw}$ values as given in per mille against VSMOW on first mention (L142)

L217: Should be case 31 not 30, I believe

Correct, we will scrutinize our references to case numbers throughout the manuscript in an attempt to catch similar mistakes.

L272: Was the seasonal seawater composition range calculated based on the difference between warmest and coldest month (as described in the text) or between the most depleted and enriched $\delta18O_{sw}$ seawater months? There are many examples where the SST and seawater cycles are out of phase (e.g., springtime snow melt flux driving a $\delta18O_{seawater}$ depletion extreme prior to the summer temperature extreme). The difference between peak summer and winter seawater composition is an important variable to constrain, but it does not necessarily always equate to the seasonal range in $\delta18O_{sw}$.

This is a fair point which actually highlights a typographic error. We compare our results against the difference between the most depleted and enriched monthly $\delta^{18}O_{sw}$ values, which do not necessarily equate to summer and winter months (see calculations in **S5**). We will rephrase this accordingly. Our Δ_{47}-based reconstruction approaches aim to disentangle the effects of $\delta^{18}O_{sw}$ and temperature on carbonate chemistry. The possibility of disentangling these two cycles, which are potentially out of phase, is one of the major advantages of Δ_{47}-based seasonality reconstructions using microsampled carbonate records.

L285: Is it fair to assume a normal marine in this environment? I suspect that this was done to illustrate a point and while I don't disagree that constraining mean annual (let alone seasonal variability) in δ18Osw in deep time adds huge uncertainty to conventional oxygen isotope interpretations, we can often make somewhat more realistic estimates of the (mean annual) value based on latitude and environment than just the global normal marine value.

This is a fair point, but we chose to assume one $\delta^{18}O_{sw}$ value for all our $\delta^{18}O_c$-based reconstructions to mimic the process underlying most seasonality reconstructions based $\delta^{18}O_c$ measurements in real (fossil) carbonate records. We agree that in some fossil cases $\delta^{18}O_{sw}$ values are assigned with consideration of the depositional environment, but very often a fixed (non-seasonal) value is assigned based on general assumptions about the global ocean during the period under study. A good example is the often-cited ice-free marine $\delta^{18}O_{sw}$ value of -1‰VSMOW (after Shackleton, 1986). Alternatively, the chosen $\delta^{18}O_{sw}$ value is sometimes adjusted based on the $\delta^{18}O_c$-based temperature reconstructions, which risks circular reasoning. Our point is that, without independent evidence for the $\delta^{18}O_{sw}$ value, seasonality reconstructions based on $\delta^{18}O_c$ data rely on estimates of this value which may be very incorrect, especially in environments such as the one in the oyster study by Ullmann et al. (2010). We therefore think that the assumption of a $\delta^{18}O_{sw}$ value of 0‰ is not unrealistic for these cases. In addition, nearly all of our virtual datasets are based on a mean $\delta^{18}O_{sw}$ value of 0‰, so our assumption of this value as the "true" marine $\delta^{18}O_{sw}$ value in all cases might even underestimate the inaccuracy of $\delta^{18}O_c$-based reconstructions overall considering the large changes in global (let alone local or seasonal) $\delta^{18}O_{sw}$ that likely occurred in the geological past (see e.g. Veizer and Prokoph, 2015). To make this point more clearly and to motivate our decision, we will introduce the problem of uncertainty in the $\delta^{18}O_{sw}$ value in the Introduction.

L294: Lower accuracy or higher offset

We rephrase this to "lower accuracy". Note that we define accuracy as the agreement between reconstruction and "true" value, so within this definition high offset and low accuracy are synonymous. To avoid confusion, we will refrain from using "offset" to describe "accuracy" throughout the manuscript after giving our definition of accuracy (lines 98-99).

L294: Change 'on' to 'of'

We will rephrase this accordingly.

LN14: Accuracy improves, offset decrease

Correct, we will rephrase to: "Accuracy also increases with sampling resolution"

LN14: Change 'samples year' to 'samples/year'

We will rephrase this accordingly.

We will insert the space here.

Correct, this should read "high precision" and we will rephrase this accordingly.

Agreed, we will remove "event or spike".

We will rephrase this accordingly.

We will rephrase to: "In cases in which multi-annual variability in…"

We will rephrase this accordingly.

These are typographic errors and we will remove one of the hyphens.

Correct, this should read "2-10 ppm" and we will rephrase accordingly.

Figures:

Figure 2: Consider adding heading labels for the blocks of virtual cases (e.g., a 'Sensitivity cases' header for 1-13, 'Natural cases' for 14-18, etc.). It will help focus the reader's eye to the differences between cases and will serve as a more useful reference throughout the results and discussion.

An excellent suggestion, and we will add these headers.

Figure 7: Color scheme is off in this figure – it looks like the smoothing and binning data may be reversed in the box plots.

Correct, colours for these approaches have been accidentally reversed. We apologize for the mistake and will correct it in the revision whilst adapting the figure to the more accessible colour scheme.

We will add a definition to the caption.

We thank the reviewer for pointing out this omission and will add data on case 18 to the table in the revision.

We thank the reviewer for their comments and will pay special attention to the clarity of our terminology and explanation of our methodologies in the revised version.

In response to this comment, we will clarify in greater detail how the three reconstruction approaches we test in this study are carried out and which data are used in each of these approaches. We will revise section 3.4 of our manuscript and Figure 4 (see comment below) to provide more detail on the workflow of these approaches. In addition, we have compiled the R scripts used to do the calculations into an R package, which is now available through the open-source online R database CRAN (https://cran.r-project.org/web/packages/seasonalclumped). In this package, the three Δ_{47}-based approaches and the $\delta^{18}O_c$-based approach are explained in the package documentation and accompanied with a data on all virtual case studies used in this work for easy reference. We will refer to this package in the revised version of the manuscript, but we will also make sure that the explanations needed for the reader to understand and distinguish between the reconstruction approaches are present in the main manuscript text.

Why was mean annual temperature used as a target? When isotope records are developed from species that have a known growth hiatus, for example during winter, it is unlikely that mean annual temperature would be targeted. An alternative target would be used, for example mean summer or a growing season mean instead. Targeting annual mean would make the record look worse than it actually is.

We agree with the reviewer that many studies refer to growth temperatures in describing their reconstructions to acknowledge the fact that most carbonate records do not grow throughout the year. A similar case can be made for the use of seasonality defined as the difference between warmest/most $\delta^{18}O_{sw}$-depleted and coldest/most $\delta^{18}O_{sw}$-enriched month, which would also be underestimated in archives in which the monthly extremes are not recorded due to growth hiatuses. However, the goal of climate reconstruction in general is to arrive at (estimates of) climate parameters which can be compared with results from climate models or previous reconstructions independent from the growth season of the archive. We believe that the conversion from mean growth season temperature to mean annual temperature eventually needs to be made. In our comparison between reconstruction approaches, we wanted to consider any uncertainty or bias associated with variability in growth rate of the archive because it can bias real climate reconstructions (from fossil archives) as well. Therefore, we opted to compare our reconstruction results with climate parameters as they are officially defined (USGS; O'Donnell et al., 2012; see discussion on lines 175-184 on page 31-32, with our apologies for the discontinuity in the line numbering).

To further clarify this point, we briefly motivate our decision for choosing these benchmarks for comparison in the revised methods section where we give our definitions for accuracy and reproducibility.

Pg 5 ln 89-91: The explanations for each of the methods are insufficient.

We regret that the explanation of the approaches was not clear and will attempt to clarify them in the revised version (see reply to major comment).

Line 195: "not exactly normally distributed" remove the word exactly.

We will remove "exactly".

Pg 29 ln 119: missing "to" between "equivalent the"

Correct, we will insert "to" here.

Pg 30 ln 142: 144 – "Carbonate records from suitable environments include, for example, the A. islandica bivalves from considerable depth (30-50m) in the open marine Northern Atlantic (e.g. Schöne et al., 2005, on which case 33 is based)." It would b inaccurate to assume that there is no variability in ï¿Ad'18Osw in NE Iceland. Variability in ï¿Ad'18Osw can also derive from changes in water mass that are bathing the shells during carbonate precipitation.

This is a fair point, and we will rephrase this statement to reflect that we cannot exclude any seasonal variability in $\delta^{18}O_w$ in this setting:

"Carbonate records from environments with more stable $\delta^{18}O_w$ conditions include, for example, the *A. islandica* bivalves from considerable depth (30-50m) in the open marine Northern Atlantic (e.g. Schöne et al., 2005, on which case 33 is based), although even here variability in $\delta^{18}O_{sw}$ due to, for example, shifting influence of different bottom water masses cannot be fully excluded."

Pg 30 ln 146: "the processes affecting δ18Osw on smaller scales" such as?

We will elaborate a bit more on these processes by adding: ", such as local evaporation and freshwater influx from nearby rivers (e.g. Petersen et al., 2016)"

Pg 30: 151-152: "While variability in δ18Osw compromises accurate δ18O-based seasonality reconstructions, the compilation in Fig. 3 shows that its influence on the δ18O records is too small to affect the shape of the record to such a degree that seasonality is fully obscured."

It is not clear whether the reviewer would like to suggest any changes to this statement.

Pg 36 ln 270 "While hiatuses encompassing half of the seasonal cycle are uncommon" I would not say that they are uncommon. There are many examples of sub-tropical to polar marine bivalve species which exhibit a cessation in growth during certain months of the year.

We agree that severe slowing or completely cessation of growth during certain months of the year are common, but we retain that hiatuses masking half a year are very uncommon. The point we would like to make here is that in most natural cases and all tested cases except for case 18 archives record at least half of the seasonal cycle. This is important because most approaches to extract seasonal variability from carbonate archives become very inaccurate if more than half of the seasonal cycle is missing (see description in section 3.2). With less than half of the seasonal cycle present it becomes nearly impossible to recognize the season of growth and therefore to estimate seasonality or mean annual temperature in any meaningful way (see reply to general comment above). We will add a sentence after this statement to clarify this point and refer to the description in section 3.2.

Pg 36 ln 280 missing "in" between "simulated the"

We will insert "in" here.

Sections 5.4.1 to 5.4.5 should be deleted. This is already a very long manuscript and this extra detail does not add anything to the focus of this manuscript.

Both reviewers questioned the relevance of section 5.4 of this manuscript and we agree that this section makes the manuscript longer than necessary. However, we retain that a (short) discussion of the implications of this work beyond climate reconstructions is useful. Therefore, instead of deleting the entire section, we decided to summarize its content into one paragraph, following the suggestion of Referee #1.

All figures: Colours used in Figures need to be changed to be accessible. The website https://colorbrewer2.org/ provides a useful free resource to check colour choices.

We thank the reviewer for their suggestion and adapted all our figures according to the RdYlBu 4-colour scheme provided by Colorbrewer 2.0 (https://colorbrewer2.org/?type=diverging&scheme=RdYlBu&n=4) in order to make them more accessible to colour-blind readers.

Fig 2: Whilst it is good to see a visual representation of each of the cases, the size of the panels makes it difficult to actually see what the purpose of each case is. It would be helped to add a schematic or table highlighting the purpose of each case.

This is a good point, and echoes a comment by Referee #1. In reply to this comment, we will add headers to indicate the various categories of cases (e.g. "sensitivity tests", "natural cases", "Real SST/SSS data" etc.; see **S1**). In addition, we will try to include an abridge version of table **S1** to the main text to make this figure, and the difference between cases in general, easier to interpret by the reader.

Fig 4A. Whilst a schematic would be a helpful figure to help the reader visualise what is a very complex methodology, the schematic presented in Fig 4A doesn't help. Currently this schematic does not help to clarify the methodology. I would suggest that the example provided in Fig 4V is removed and more space provided for panel A. This would provide space to add detail along with the arrows.

This is a good suggestion and in the revised version we will limit Figure 4 to include only panel A. This would allow us to add more information next to the arrows detailing the workflow of creating virtual data and doing reconstructions. We would like to keep the example now included in panel B of Figure 4 in the main text and will include it as a separate figure.

Fig 6. Axis text size needs to be increased.

We agree that this figure is relatively dense and in response to a comment by Referee #1, we will try to represent the results of individual cases in a different way to make it easier to grasp the information. The present Figures 6 and 7 will then likely move to the supplement, where we will include them with larger text sizes.

Fig 12: Change the colours on the plot, you can't tell the difference between the d18O and the binning results.

We apologize for the fact that our choice of colours for the plots throughout the manuscript has compromised the clarity of our figures. As suggested, we used the Colorbrewer tool to pick more accessible colours and revise all our figures accordingly.

References

Keating-Bitonti, C. R., Ivany, L. C., Affek, H. P., Douglas, P. and Samson, S. D.: Warm, not super-hot, temperatures in the early Eocene subtropics, Geology, 39(8), 771–774, https://doi.org/10.1130/G32054.1, 2011.

Shackleton, N. J.: Paleogene stable isotope events, Palaeogeography, Palaeoclimatology, Palaeoecology, 57(1), 91–102, 1986.

Ullmann, C. V., Wiechert, U. and Korte, C.: Oxygen isotope fluctuations in a modern North Sea oyster (Crassostrea gigas) compared with annual variations in seawater temperature: Implications for palaeoclimate studies, Chemical Geology, 277(1), 160–166, 2010.

Veizer, J. and Prokoph, A.: Temperatures and oxygen isotopic composition of Phanerozoic oceans, Earth-Science Reviews, 146, 92–104, https://doi.org/10.1016/j.earscirev.2015.03.008, 2015.