RC1:

Review of manuscript submitted to Climate of the Past by Philip Meister and colleagues: A global compilation of diatom silica oxygen isotope records from lake sediment – trends, and implications for climate reconstruction

The temperature dependence of oxygen isotope fractionation makes them a widely applied (paleo)climatology tool. Oxygen isotope ratios (expressed as $\delta^{18}O$) in diatom biogenic silica represent a valuable archive, but their interpretation and model-archive comparison can be complicated, particularly in lakes. Here, Meister and colleagues compile published lake diatom $\delta^{18}O$ records to address the extent to which a common signal can be observed. They find 54 lakes in total, from 71 publications, that have available data. Compiling, binning and filtering the data, they observe commonalities in lake diatom $\delta^{18}O$ records over the common era and a consistent, decreasing trend over the Holocene, which is broadly consistent with existing palaeoclimate (particularly temperature) records and understanding of the $\delta^{18}O$ proxy.

In general, this is a useful compilation that seems comprehensive and thorough and achieved in a sensible way, and is being made available via an appropriate repository (Pangaea.de). The introduction largely sets out the state-of-the-art (and Fig. 1 is particularly useful, I feel). The compilation and standardisation seems like a large effort the authors should be commended for. The topic of the manuscript falls within the scope of Climate of the Past, and the conclusions are largely supported by the data. The manuscript is generally well written (although with scope for tightening the language and a few typological/grammatical errors that could be caught with a through proofread) and the figures are clear. Overall, I think this is a manuscript and a data compilation that deserves to be published, and will hopefully stimulate more work in the field (especially given their Fig. A1B which shows declining production of lake diatom $\delta^{18}O$ data).

My major criticism is that the discussion of the compilation is rather descriptive and qualitative. There is very little in the way of statistical analysis, which might be useful in parts of the discussion related to the magnitude of $\delta^{18}O$ trends, regional differences in the timing of minima/maxima, and the presence or not of periods of stasis. To what extent can they be demonstrated to reflect 'real' underlying phenomena, vs. just being an artefact of low and noisy data availability? The comparison of the new NH compiled record to existing proxy data and insolation curves is also rather qualitative. Beyond this, the discussion focuses almost entirely on sites from >45 deg N, which is motivated only on L249 (unless I miss it elsewhere). Given the general paucity of data, it seems a shame not to exploit the compilation as much as possible. While other regions might be too data scarce to do the binning/filtering steps in e.g. Figs 5 and 6, do they contain useful information on the spatial pattens which could complement the discussion in e.g. section 4.2 and Fig. 7? Fig 5G and 6G each contain only two sites that meet the quality control criteria – is it really the case that there are not even two sites from e.g. South America or tropical Africa sites that are useful?

Answer: The aim of this manuscript is to compile all diatom isotope data and not per se a climate reconstruction. The $\delta^{18}O_{BSi}$ proxy is highly variable as also demonstrated in the large variability between individual records, but also in the compilation itself. It is astonishing though, that despite all the different influence factors on $\delta^{18}O_{BSi}$ summarized in Fig 1 and Section 1.2, so many well-known trends (i.e. Holocene cooling), and minima/maxima (e.g. LIA) are visible pointing to an overall connection of this proxy to climate and especially air temperature (T_{air}).

For a meaningful climate reconstruction, it has shown to be absolutely mandatory to constrain the selected records (latitude, hydrological setting, T_{res}) and to bring them to a comparable temporal resolution (binning) to meaningfully (based on statistical analyses) compare the datasets and derive common aspects and trends. This leads to a reduction of jointly interpretable datasets, which unfortunately excludes some regions (such as South America and Africa). However, following the reviewer suggestion, we have added a short discussion of the records of excluded regions in the respective discussion sections. However, this interpretation is based on observations, and not on a statistical assessment.

Moreover, in line with comments of Rev #3, we have added the more quantitative aspects to the discussion by including slopes of trends and rates of changes in the 12K and 2K compilations to the discussion. See more details below.

See more details below.

Minor comments

Fig 1/main text L105: Could diagenesis also be included in this figure? And/or more detail given in the main text? Presumably older samples are more susceptible to diagenetic overprinting? How can we be sure this is not a major driver of the observations?

Answer: This is a valid and very useful comment. Early diagenetic processes are likely affecting diatom samples. Most of the discussion of diagenetic effects on $\delta^{18}O_{BSi}$ deals with post-mortem alteration of diatoms (Moschen et al., 2006, Tyler et al., 2008) that deal with the (isotopic) difference between recent/living and fossil diatoms. Different preparation techniques have been developed to get rid of the loosely-bound oxygen (i.e. from organic matter or hydroxyl groups) to overcome this effect. Since the time series considered in this manuscript are all from fossil diatoms, post-mortem alteration is less important for this discussion. This subject has been tackled in specialized manuscripts.

Diagenetic effects could influence the oxygen isotope composition also on longer time scales by preferentially incorporating ^{18}O from exchangeable Si-OH bonds into the Si-O-Si structure which would result in higher $\delta^{18}O$ values (Brandriss et al., 1998; Schmidt et al., 2001; Moschen et al., 2006, Akse et al., 2022). Chapligin et al. (2012) have tested long-term diagenetic effects at Lake El'gygytgyn, one of the oldest lake records of this data compilation, and despite the reduction of silanol bonds with time, found no significant effect on $\delta^{18}O_{BSi}$ for the last 250kyrs.

This information has been added to Section 1.2: Controls on $\delta^{18}O_{BSi}$

Within the lake itself, long-term diagenetic effects on diatoms include recrystallisation and incorporation of heavier ^{18}O from silanol bonds into the Si-O-Si crystal lattice – a process that may increase the original $\delta^{18}O_{BSi}$ signal (cf. Akse et al., 2022; Fig. 1). However, despite notable recrystallisation, Chapligin et al. (2012b) found no significant diagenetic effect in records spanning the past 250 kyr. (L143-147)

We have decided to add the following sentence about diagenetic alterations of the proxy to the discussion:

Long-term diagenetic effects have shown to be of little influence on $\delta^{18}O_{BSi}$, at least for the last 250kyrs (Chapligin et al., 2012b) (L701-702)

Moreover, we have added diagenesis as a post-depositional effect to Figure 1.

L210: I would suggest a brief summary of the hydroLakes database is warranted, since L219 refers to a 'geostatistical approach', implying that values are not specific to a given lake, and many readers (including myself) would not be familiar with it. Would it not make sense to preferentially use the parameters as given in the original publications, supplementing them with hydroLakes only when necessary?

Answer: Values of the Hydrolakes database are in fact specific to a given lake. The advantage of the Hydrolakes database is that they apply a uniform method for each lake, whereas original publications often even vary in the approach for obtaining lake basin parameters. The approach of the Hydrolakes database uses digital elevation models (DEMs) of the terrain surrounding the lakes in question and extrapolates this terrain into the lake basin. This is unfortunately not available for all lakes and the individual publications often do not offer this information. We consider consistency as the most important issue in this study, and, therefore, to our opinion, a subset of studies is sufficient for our purpose.

The word "geo-statistical" is used in the title of the Messager et al. publication (see reference list) to describe their approach.

Fig 4: excludes 33 (of the 49 extant) lakes because the HydroLakes dataset doesn't include their catchment area. I would suggest this is something that is relatively easy to define in a consistent way given a digital elevation model, with readily available topographic analysis tools, e.g. the 'TopoToolbox' for Matlab, or similar capabilities within ArcGIS or qGIS.

Answer: Recalculation of catchment sizes is beyond the scope of this manuscript. As outlined above, the catchment sizes are not essential to the discussion of the $\delta^{18}O_{BSi}$ in this manuscript (not a selection criterium), and to our opinion, the selected subset of data is sufficient for Fig. 4A and the information drawn from this

We have added the following sentence to the methods section in order to clarify this issue:

Therefore, and because some parameters (i.e. catchment size) are not essential to the discussion in this manuscript, we have chosen this approach favouring consistency over more data points.(L226-227)

L233 "presumably" instead of "supposedly"?

Answer: changed

L261: Does this approach of subtracting the mean of a record only work when every record covers the full timespan under consideration? Otherwise a record that covers only e.g. the mid to late Holocene would bias that period (because the mean subtracted from that record would be smaller than a record with an identical gradient/trend but longer coverage).

Answer: Yes, this is correct. This is why we decided not to include records with too short a temporal coverage (for the 12K compilation a minimum of 10 (out of 12) bins, for the 2K compilation 7 out of 10 bins were needed). After removal of the mean, the trends remain valid but the records might be slightly offset (due to shifted mean values). We have added the following sentence to the Methods section:

Obviously, the missing bins in both Holocene and Common Era compilations might have an effect on the mean value of the individual records, but not on the overall trend. (L266-267)

L270: It's a bit unclear how many sites/records are actually used. L271 states 64 sites; L283 is states both 54 and 56 (i.e. 7 + 49); Table A1 has 53 lakes. Can this be clarified?

Answer: The reviewer is right, all these numbers were confusing and we are grateful for this hint. This was because sometimes, there is more than one record from one site, in other cases subsets of records have been part of different publications. So, we have identified 71 publications from 64 different sites. These have been combined to 53 records (as presented in the Tab A1, now Tab A2). The number of the lakes and paleo lakes with and without chronology have been adjusted to the information given here.

L275: Table A2 is mentioned before Table A1. Is A2 complete? There doesn't seem to be enough lakes here to match the numbers given in the main text. Also, it might be helpful to include the number #X in both tables A1 and A2 somehow to allow cross referencing.

Answer: We agree with the reviewer and we thank for this remark. We exchanged the sequence of Table A1 and A2 for consistency. Moreover, the information of the record number #X was also added to former Table A2 (now Tab A1). In new Tab. A1, all publications and all records are mentioned to allow following the selection and merging to final the datasets. Sometimes, more than one dataset is in one publication, sometimes there are several different records from the same lake merged into one record number #X.

L285: "extant"? (rather than 'still existing')

Answer: changed, also in the Discussion

L310: A reference to Downing and Duarte (2009) (their Fig 5) or similar might be useful here.

Answer: Thanks for this suggestion. We included the reference and one sentence.

Moreover, high northern latitudes have by far the highest abundance of total area of water bodies (Downing and Duarte, 2009). (L318-319)

L351: 'may correspond to more than one record'

Answer: changed

L379: There can be some buffering even when t_res is less than sampling frequency, meaning the "full" amplitude is not necessarily displayed See e.g. Richter and Turekian (1993).

Answer: Correct. We have added two sentences and the corresponding reference to the text.

When t_{res} is lower than the sampling frequency, the hydroclimatic amplitude is also not fully captured (Richter and Turekian, 1993). It should be noted that temporal resolution is non-uniform across the records, with generally higher resolution for more recent time intervals (Figs. 2 and 5A) that – in addition to a sampling bias – may, in part, reflect increasing lake sediment compaction with depth and time. (L390-393)

Fig 5: Presumably the axis labels shouldn't read 'kyr' but 'yr'?

Answer: The reviewer is right. This has been changed.

L435: it seems like more could be done here with the records from other regions of the world.

Answer: This is in line with a comment of this reviewer further up. We tried to improve the text by elaborating on other regions and time periods, even if these records did not fulfil the selection criteria for statistical assessment. Due to the scarcity of records from other regions and overlapping time periods, it is not possible to depart from case studies. These have already been discussed in the individual publications.

To clarify, we have added the following paragraph to the Discussion, section 4.2

*The African (N=7) and South American records (N=3) do not fulfill the criteria set for the NH and Eurasian stacks, either due to hydrological or temporal constraints (e. g. too few datapoints). However, all three South American records (*see Tab. A2; #18, 28, 32) *show an early Holocene maximum around 10 kyr BP. Two out of three South American $\delta^{18}O_{BSi}$ records show a decrease over the Holocene, whereas the third record displays no clear trend.*
Due to differences in hydrology, an African $\delta^{18}O_{BSi}$ stack has not been calculated either. Instead, individual records (N=5) have been compiled and binned for the African continent (Figs. A5A, A5B). The largest $\delta^{18}O_{BSi}$ variability within one individual record ($\Delta^{18}O_{BSi}$ spread of ca. 20‰; #21) as well as for a continent (with values up to +45‰, #33) have been observed. This is related to the very different settings both in altitude (Tab A2), but also in hydrological characteristics (Fig. A5D). Overall, the first half of the Holocene is characterized by slightly lower $\Delta^{18}O_{BSi}$ values, than the second half. As this observation is less obvious for open lakes (Fig. A5D) a bias linked to widely different hydrological settings of the respective lakes (see Tab. A2) has to be assumed, which renders it difficult disentangling the drivers of $\delta^{18}O_{BSi}$. This further underlines the importance of hydrologically-constrained records to infer a common (climate) signal (see Fig. 5A, 6A). L(553-565)

L465: This seems like repetition/overlap with paragraphs starting L486 below. Is a ref here to fig 7 appropriate?

Answer: In L465, Figure 6 is interpreted showing time-series of individual records displaying min and max at different times. Figure 7 (correctly linked to at L486) brings this aspect to the spatial domain. Both say that the timing of Holocene minima and maxima is different (so yes, there is a redundancy, but a desired one). Thus, we decided to not make changes in the text.

L473-5: An example of where more statistical rigor might help: how robust is this particular interpretations relative to a simpler view of consistently decreasing $\delta^{18}O$? (a straight line could be drawn through the ±1sd shading).

Answer: Following the reviewer suggestion, we calculated the slopes for the summary compilation for both the 2K subsets and for NH, Eurasia and North America respectively. These parameters are given in a new paragraph in Chapter 4.1.

The trends calculated for the last 2 krs (2K) for the NH (Eurasian) $\delta^{18}O_{BSi}$ stacks are high with -0.64‰/kyr (-0.85‰/kyr) for the 10 bins, with even higher slopes of -1.6‰/kyr (-1.7‰/kyr) for bins 6-10 (1 to 2 kyrs), and of -1.4‰/kyr (-1.8‰/kyr) for bins 1-5 (corresponding to the last millennium). The negative slope is interrupted by three consecutive bins (4-6; 0.6-1.2 kyr BP)

and bin 8 (1.4-1.6 kyr BP) with higher $\delta^{18}O_{BSi}$ than the previous bin. For the last 2K, the other 5 bins (1-3, 7 and 9) show a negative sign for the NH and Eurasian $\delta^{18}O_{BSi}$ 2K reconstructions. North American records (N=2) do not show a consistently decreasing trend (Fig. 5G), but slightly higher values at 1700 and 1900 yrs BP compared to the most recent bins (100 and 300 yrs BP, respectively). They do, however, show lower $\delta^{18}O_{BSi}$ values between 900 and 1300 yrs BP, followed by a $\delta^{18}O_{BSi}$ maximum at 500 yrs BP. As there are only two records available after filtering, caution has to be applied in interpreting this pattern.
For the North American $\delta^{18}O_{BSi}$ 2K reconstruction, a lower overall gradient is observed (-0.12‰/kyr), which shows also a steep slope for bins 6-10 with -1.5‰/kyr. In contrast to the NH and Eurasian $\delta^{18}O_{BSi}$ stack, bins 1-5 show a slight decrease of -0.4‰/kyr, only. This leads to slightly shifted minima and maxima between $\delta^{18}O_{BSi}$ reconstructions for the last two millennia: NH and Eurasian reconstructions have their absolute maxima in bin 10 or at 1.8-2.0 kyrs (absolute minima: bin 1; 0-0.2 kyrs) with intermediate minima at 1.2-1.4 kyrs BP (bin 7) and maxima between 0.6-1.0 kyr BP (bins 4 and 5). In contrast, for the North American reconstruction, the absolute $\delta^{18}O_{BSi}$ minimum (maximum) is at 1.2-1.4 kyrs BP or bin 7 (0.4-0.8 kyr BP (bins 3 and 4), whereas the early maximum (bin 10) and late minimum (bin 1) are less pronounced. In summary, we observe an overall decreasing trend in $\delta^{18}O_{BSi}$ for the last two millennia, for NH, Eurasian and North American stacks, which is accelerated in the first millennium. (L439-456)

Moreover, a similar paragraph has been added to the following chapter 4.2 about the 12K $\delta^{18}O_{BSi}$ reconstructions:

The Holocene trend calculated for the NH (Eurasian) $\delta^{18}O_{BSi}$ stack is -0.19‰/kyr (-0.21‰/ kyr) for all 12 bins, with a lower slope of -0.11‰/kyr (-0.10‰/kyr) for bins 7-12, and a much higher slope of -0.39‰/kyr (-0.36‰/kyr) for bins 1-6 (corresponding to 0-6 kyrs BP).
Throughout the Holocene, 10 out of 12 bins are lower than the previous one (negative sign) for the NH $\delta^{18}O_{BSi}$ reconstruction, except for bins 7 (6-7 kyrs) and 9 (8-9 kyrs), which show an increase. For the Eurasian $\delta^{18}O_{BSi}$ stack all bins are lower in $\delta^{18}O_{BSi}$ than the preceding one (except for bin 7, or 6-7 kyrs, which shows an increase of +0.2‰). The highest negative slopes exceeding -0.55‰ per kyr are reached in bin 2 (1-2 kyr) for both, the NH and the Eurasian compilation. For the North American $\delta^{18}O_{BSi}$ reconstruction, a slightly lower gradient is observed for the Holocene (-0.13‰/kyr; based on just 10 bins), which also shows a steeper slope for bins 1-5 with -0.22‰/kyr. In contrast to the NH and Eurasian $\delta^{18}O_{BSi}$ stack, bin 6-10 show a slight increase of +0.10‰/kyr, likely linked to the bins 6 and 9, which are the only ones showing a positive sign, whereas all 7 other bins show a decrease compared to the preceding one. (L512-524)

L489: ‘show a tendency’ (also L503)

Answer: changed in both cases

Fig 7: Even if not discussed (see above) could the minima/maxima of records outside the northern hemisphere high latitudes be displayed here?

Answer: we set the selection criteria for the records to be included to the analyses as explained in the Methods. As records outside Eurasia and North America do not fulfil the criteria set, and e.g. strongly vary in their hydrological characteristics, they have to be excluded from the statistical evaluation.

Rev#1 proposed a better exploitation of the dataset. This is why we added a paragraph to section 4.2 about what can be learned from combining the African and South American records, even if they do not meet the criteria for the statistical evaluation (see above). In this paragraph, we emphasise the importance of hydrological filtering when climate reconstructions are a primary goal.

L505: Again, without more statistical rigor, from e.g. Fig. 6E/G I would be cautious about over interpreting these differences.

Answer: we have changed the phrasing here to comply with the reviewers comment to:

Despite the scarcity of the records fulfilling the selection criteria, this suggests a different behaviour of the regions (North America and Eurasia), though we also acknowledge that one of the two North American sites (#35, Heart Lake) is located in the central North Pacific Ocean and has a different climatic history. (L549-552)

Moreover, as outlined above, we have now added a more quantitative approach (i.e., slopes and rates of change between bins) for 12K and 2K reconstructions for NH, Eurasia and N America – see new paragraphs highlighted above.

L520: presumably also because most biogenic production is in summer (particularly relevant for short residence time systems).

Answer: This information has been included into the text.

Fig 8 caption: subscripts and superscripts not displayed correctly.

Answer: this has been changed accordingly.

L558: Fig. 8D is labelled June, here it says July.

Answer: The information in the text is correct. The figure has been adjusted (to July). Thanks for this.

L564: How far behind?

Answer: A lag of several thousand years is possible over longer time scales. This information (the word "millennia") has been added to the text

Section 4.4: given some of the discussion previously I would have expected some stronger/more explicit recommendations in this section, for how diatom $\delta^{18}O$ can become a more useful proxy (beyond the rather generic 'further research is needed…').

Answer: The dataset needs to be enlarged and preferably from lakes meeting the selection criteria outlined in the discussion. The best option would be to have studies from sites with rather uniform spatio-temporal characteristics. We added this "wishlist" to the Section 4.4. The following sentence was added to the text:

We recommend studies with uniform spatial and temporal coverage, i.e. hydrologically open lakes with a long, continuous sedimentation history, as our synthesis indicates these to be *most promising for generating comparable binned time series to extend climate reconstructions further into the past.* (L711-714)

L595: 'would be consistent with' or similar?

Answer: Following the reviewer suggestion, the sentence has been changed to:

This is notable as glacial and interglacial periods are characterized by different environments, atmospheric circulation patterns and likely hydrological settings (e.g. formation or closure of outflows from lakes, lake level fluctuations). (L678-679)

References

Downing, J.A., Duarte, C.M., 2009. Abundance and size distribution of lakes, ponds and impoundments, in: Likens, G.E. (Ed.), Encyclopedia of Inland Waters. Elsevier, Oxford, UK, pp. 469-478.

Richter, F.M., Turekian, K.K., 1993. Simple models for the geochemical response of the ocean to climatic and tectonic forcing. Earth and Planetary Science Letters 119, 121-131, doi:

Answer: both included

RC2: 'Comment on cp-2022-96', Witold Bagniewski, 23 May 2023 reply

This study compiles 71 δ18OBSi records from lake sediments covering various time periods and locations, predominantly >45° N during Holocene. Despite originating from different geographic locations, the records feature common patterns that correspond to known climate events of the Holocene and the Common Era. The collection of δ18OBSi records, binned to a common temporal resolution and complemented with metadata on hydrological parameters, is a valuable contribution that will facilitate climate reconstructions and proxy–model comparisons.

This manuscript is very well written and the records compiled for this study are well presented. Although the statistical analysis presented here is rather simple, this is probably the correct approach considering the large differences in temporal coverage and sampling frequency between the records, as well as the sparse spatial coverage. This work is valuable for the paleoclimate community and clearly deserves to be published in Climate of the Past. However, I found several issues that should be addressed before publication.

Answer: We highly acknowledge the remarks from Rev#2. We agree that the statistical analysis was rather simple, but it is already a quite dense manuscript, which, we sense, would become less readable when a detailed statistical exploitation of the data would be implemented. As this comment contrasts with comments of Rev#1, we added paragraphs about the slopes and rates of changes of for the different compilations (12K and 2K). We think that these statistical analyses allow a better comparison of the data compilations on both, temporal and spatial scales, being aware that these sections are rather technical.

See above

My only major criticism has to do with the analysis in Section 4.4, which, in my opinion, is unconvincing and requires revision. Please see my comment below.

Answer: This is in line with Rev #1. We answer below where in Section 4.4 is discussed in detail.

Information provided in the two tables is useful and well presented. However, these tables and the associated PANGAEA dataset would benefit from including additional information discussed in the text, such as the dating method, temporal coverage, and temporal resolution. These details could be incorporated into Table 1A or shown in a separate table.

Answer: Thanks for these suggestions. We welcome the comments of all 3 reviewers that suggest more information be included for the different sites, in our tables. However, on balance we believe that adding in all these extra details would overwhelm the readers with too much information; the information already given has been carefully selected. Adding this information to table A1 as part of the manuscript would also make these tables pretty much unreadable. However, all this requested information (dating methods, temporal coverage, and temporal resolution) will be part of the dataset submitted to Pangaea, and will therefore be available for further studies.

The Introduction is comprehensive and well-written; Fig. 1 provides a helpful overview of the processes shaping the δ18OBSi signal. However, it does not become clear until Section 1.4 why this study focuses on diatoms but not other sources of d18O data. It would help the reader understand the novelty and purpose of this study if this was mentioned in the Abstract and/or earlier in the Introduction.

Answer: Following the reviewer suggestion, we have added the following sentence to the abstract:

This study provides a comprehensive compilation and combined statistical evaluation of the existing lake sediment $\delta^{18}O_{BSi}$ records, largely missing in other summary publications (i.e. PAGES network). (L43-45)

L54: This sentence could be made more clear. Does the phrase "common δ18OBSi patterns" describe a comparison between the different lake records or a comparison between the lake records and previously known climate events?

Answer: changed. A common signal between the different lake δ18OBSi patterns first and then prominent min/max a correspondence to known climate episodes. This has been clarified in the text

L58: Typo "extratopic"

Answer: changed

L182: "Such compilations, however, generally do not include δ18OBSi–records." Why not?

Answer: Good question. We do not really know. The parameter is more difficult to measure in silica than in carbonates, and being tendentially more abundant in high latitudes and altitudes. There are relatively few records compared to carbonates or other natural climate archives. One aim of our paper is to make this to our opinion powerful proxy more visible. As reviewer #3 summarizes: the resolution, the number and the spatial representativeness of the data are relatively limited.

L232-233: "the effect of different 14C–calibrations and different age model approaches is supposedly minor." Could you please elaborate on that?

Answer: What we meant is that the 14C curve is updated regularly and older papers might be dated with e.g. INTCal13, whereas newer publications should already include INTCAL20. We did not want to enter into a dating discussion or to recalibrate all ages (which are by the way not always available), but just use the age models from the original publications. This sentence is showing the awareness that minor effects might be related to this (i.e. one sample shifted to another bin, or changing the mean value of two bins slightly)

Section 3: The authors have grouped the records according to Marine Isotope Stages. It would be helpful to define the temporal boundaries of the MISs within the text.

Answer: The boundaries between the MIS are already given in Figure 2. These are defined in Lisiecki and Raymo, 2005 as: MIS 1-2: 14kyr; MIS 2-3: 29kyr, MIS 3-4: 57 kyr, MIS4-5: 71 kyr, MIS 5-6: 130 kyr. We added this information to the figure captions and the text. The beginning of MIS 1 deviates substantially from the Holocene. Hence, for our terrestrial biogenic silica isotope compilation, we have considered the Holocene (0-11.7 kyrs (12 bins) similar to approaches by the PAGES community. We have made this more clear in the text that our 12K reconstruction relates to the Holocene (and not to MIS 1).

L276-280: Can these dating methods be included in Table A1?

Answer: As outlined above, adding this information to table A1 as part of the manuscript would make these tables pretty much unreadable. However, all this requested information (dating method, temporal coverage, and temporal resolution) is part of the dataset submitted to Pangaea, and, by this, available for further studies.

Meister, Philip; Alexandre, Anne; Bailey, Hannah; Barker, Philip; Biskaborn, Boris K; Broadman, Ellie; Cartier, Rosine; Chapligin, Bernhard; Couapel, Martine JJ; Dean, Jonathan R; Diekmann, Bernhard; Harding, Poppy; Henderson, Andrew; Hernandez, Armand; Herzschuh, Ulrike; Kostrova, Svetlana S; Lacey, Jack H; Leng, Melanie J; Lücke, Andreas; Mackay, Anson W; Magyari, Eniko Katalin; Narancic, Biljana; Porchier, Cécile; Rosqvist, Gunhild C; Shemesh, Aldo; Sonzogni, Corinne; Swann, George E A; Sylvestre, Florence; Meyer, Hanno: A global compilation of diatom silica oxygen isotope records from lake sediment. *PANGAEA*, https://doi.pangaea.de/10.1594/PANGAEA.957160

This information has been added to the text and reference list. A temporary link has been generated for the Pangaea review process

https://www.pangaea.de/tok/006d146aaa06808639b016084bc7684fc466d705

L288-297: Could the temporal coverage information be included in Table A1?

Answer: As outlined above, adding this information to table A1 as part of the manuscript would make these tables pretty much unreadable. However, all this requested information (dating method, temporal coverage, and temporal resolution) will be part of the dataset submitted to Pangaea, and by this available for further studies. See previous answer.

Fig. 2B: It appears that there is missing information for some records. Please provide an explanation.

There are 40 records included in Fig 2B (out of 53 records in total). This largely excludes paleo lakes as the figure 2B ends at ca. 130 kyrs. From the 47 remaining records, 5 are without any chronological control (see text), others do not give sufficient chronological constraints to display them in Fig. 2B.

Section 3.4: Section title is "Temporal coverage and resolution of combined records" but is "temporal coverage" discussed here?

Answer: rephrased to “Temporal resolution of combined records”

Section 3.4: Can the temporal resolution be included in Table A1?

Answer: As outlined above, adding this information to table A1 as part of the manuscript would make these tables pretty much unreadable. However, all this requested information (dating method, temporal coverage, and temporal resolution) will be part of the dataset submitted to Pangaea, and by this available for further studies. See above.

Section 3.4: Is "sampling resolution" the same as "temporal resolution"? Additionally, it should be noted that temporal resolution is nonuniform across some (all?) of the records, with generally higher resolution for more recent time intervals (e.g. see Fig. 5A). This impacts the comparison shown in Fig. 4B.

Answer: As outlined in Section 3.4, the temporal resolution depends on many aspects: Temporal resolution of the records and the resulting signal properties are determined by both the lake basin itself (i.e. accumulation rates and preservation of diatom silica) and the sampling routine applied to the sediment core (i.e. the sampled intervals as well as the thickness of individual samples).

To clarify this issue raised by the reviewer, we have added the following sentence to chapter 3.4

It should be noted that temporal resolution is non-uniform across the records, with generally higher resolution for more recent time intervals (Figs. 2 and 5A) that – in addition to a sampling bias – may, in part, reflect increasing lake sediment compaction with depth and time. (L391-393)

L488: The authors state that Eastern Eurasian sites feature a Holocene maximum at 12 kyr BP, but it seems that this is because the start of the Holocene has been defined as 12 kyr BP. As can be seen in Fig. 9A, the actual maximum occurs earlier in some records, around 13-14 kyr BP. Therefore, it might be more appropriate to show 13-14 kyr BP as the Holocene max. in Fig. 7A for these records.

Answer: The reviewer’s observation is correct. Some records have their maximum before the onset of the Holocene. We acknowledge that the marine isotope stages which are used in the manuscript would allow to go further back in time (i.e. to 14 kyrs b2K). The chronostratigraphic boundary of the Holocene is set to 11.7 cal kyr b2K (see Walker et al 2008, JQS). And this is the obvious reason, why the bin 11-12 kyrs is the last one considered in the Holocene reconstruction. We have changed the text in the Holocene section (using the term Holocene rather than MIS-1 wherever appropriate) Moreover, we have added the following sentence to the section 4.2 (Holocene…) to account for the remark of Reviewer #2

To facilitate comparison with other Holocene reconstructions, for deriving Holocene trends only 12 kyr (in 1 kyr-bins) are considered and the boundary between MIS 1 and MIS 2 is set to 12 kyr. Thus, the Holocene NH subset of records (covering the past ca. 12 kyr)…(L469-471)

Section 4.4: The comparison of MIS 1 and MIS 2 means, in my opinion, is very misleading. As shown in Fig. 9A, some records only cover the very end of MIS 2 when d18O is near maximum. Thus, these records have higher mean MIS 2 d18O values compared to records that span the entire MIS 2 period. Fig. 9B shows large differences between the records, which can be wrongly interpreted as differences in the climate signal, when they are likely the result of the records covering different time intervals. In fact, the records in Fig. 9A appear to be in a good agreement, except for one outlier. The authors acknowledge their concern in L582 but unfortunately I feel that there is no benefit in the anaysis presented in Fig. 9B. I recommend either removing the lower panel of Fig. 9 or replacing this analysis with a different one. Section 4.4 should be revised accordingly.

Answer: We understand and agree with this valuable point that reviewer #2 addresses here. We have removed Fig. 9B from the discussion. This is, however, in contrast with a comment from Rev #1 who asks for exactly this info for the sites not included in the time series. Thus, we have taken the decision of following the argumentation of reviewer #2 here, as we agree the benefit of figure 9B is minimal. Thus, we have based our discussion only on Fig. 9A. We have revised Fig. 9A to allow for distinguishing between individual records. We have also added the following sentence to chapter 4.4 and revised the chapter slightly.

However, it must be stressed that the lack of records covering the complete MIS 2 complicates a robust statistical comparison between MIS 1 and 2. (L681-682)

Fig. 9A shows NH records, but Fig. 9B shows all individual records. It would be more consistent to present the same set of records in both panels.

Answer: As pointed out in the previous answer, we have removed Fig. 9B, and reduced the discussion chapter slightly excluding details about geographical differences.

Fig. 9B: It is unclear whether what is shown is the difference MIS 1 - MIS 2 or MIS 2 - MIS 1.

Answer: MIS 1 – MIS 2. This information is now given in the figure captions of Fig. 9.

EC1: 'Comment on cp-2022-96', Denis-Didier Rousseau, 06 Jul 2023 reply

Dear author,

As the discussion phase of your manuscript is still ongoing, pending a third review, could you please post a short response to the two reviews already submitted? In this way, you could initiate a discussion that might clarify the reviewers' point of view. It's simply a matter of seizing the opportunity of this particular phase of the review process within the CP.

All the very best

denis-didier Rousseau

CP co-editor in chief

Review of the paper by Meister et al. entitled "A global compilation of diatom silica oxygen isotope records from lake sediment – trends, and implications for climate reconstruction"

General comments

In this study, the oxygen isotopes in biogenic silica ($\delta^{18}O_{BSi}$) of 71 down-core records published to date were analysed and interpreted with respect to climate change for different regions and time periods (Common Era, Holocene, MIS 2). The focus is mainly on the correlation of $\delta^{18}O_{BSi}$ values to temperature. As a specialist in climate dynamics and meteorology, I am not able to assess the geochemical sections of the publication.

The strength of the paper is that the existing data and publications dealing with $\delta^{18}O$ in diatom silica are collected, archived, documented and interpreted together. This must be seen as an important step, because the dynamics of $\delta^{18}O$ are a substantial component of the dynamical climate system. Therefore, $\delta^{18}O$ also forms an important component of modern climate models. For these reasons, the publication of this contribution is desirable and highly recommended.

Answer: We thank the reviewer #3 for this positive attitude towards our manuscript.

The paper has several weaknesses. On the one hand, the relevant processes, as excellently illustrated in Figure 1, should be given more consideration in the interpretation of the data. Processes such as evaporation and rainfall, transport distance and season of precipitation, continentality, freezing and melting, etc. play a crucial role if they are related to temperature or precipitation. On the other hand, both the resolution, the number and the spatial representativeness of the data are relatively limited. Overall, the statements made primarily concern the Northern Hemisphere.

Answer: We agree with reviewer #3 that the processes influencing the $\delta^{18}O_{BSi}$ signals are highly complex and sensitive to different climatic and hydrological factors. However, with the data at hand, it is beyond the scope of this manuscript to go beyond the analyses shown in the manuscript. This is due to both the scarcity of records and the different hydrological parameters of the records available which limits their like-for-like comparability. The main goal of this study is to investigate if there are common patterns in the $\delta^{18}O_{BSi}$ signals of different lakes, not to disentangle all the factors influencing it. These factors have been addressed in the original publications as well as being summarised in section #1.2. Individual records have been interpreted differently in original publications (with different drivers impacting on $\delta^{18}O_{BSi}$), but in the end this can generally be reduced to either changes in $\delta^{18}O_{lake}$ or temperature. More aspects outlined in the following comment related to Fig. 5.

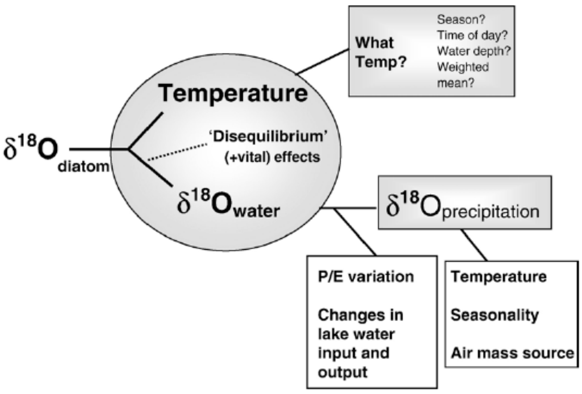
For the reasons mentioned, I recommend acceptance of the publication after major revision.

Specific comments

Figure 5: The variability of the data is extremely high. Perhaps instead of plotting maxima and minima (Fig. 7), individual curves should be assessed in relation to the processes shown in Figure 1. In general, it must be acknowledged that in Figure 5 F the two main cooling periods (Dark Ages Cooling or LALIA and Little Ice Age) are quite clearly indicated. Büntgen et al.

(2016) have recently characterized the Dark Ages Cooling and Wanner et al. (2022) diagnosed the Little Ice Age cooling (Wanner et al., 2022).

Answer: The general strategy was to compile all existing $\delta^{18}O_{BSi}$ records after given criteria. The variability is high and this is why we present the 1SD and 2SD in the NH and Eurasian reconstructions both for the 2K and 12K time slices. The individual records have been interpreted differently in individual publications (with different drivers impacting on $\delta^{18}O_{BSi}$), but in the end this can generally be reduced to either changes in $\delta^{18}O_{lake}$ or temperature. This is nicely summarized in Leng & Barker, ESR, 2006 (see below). The publication and this specific information needed for further interpretation has been included to the manuscript.



Leng and Barker, Earth Sci. Rev., 2006

We have included the mentioned publications as citations in the text at the relevant paragraph in section 4.1. Buentgen et al. 2016 and the reference to the Dark Ages Cooling was included in the first submission already.

The Holocene subsets clearly indicate a negative temperature trend. This should be commented on in more depth because it is still not clear whether temperatures increased or decreased in the late Holocene (Liu et al., 2014; Wanner, 2021). The most recent reference on Holocene climate should be commented: Kaufman and Broadman, 2023.

Answer: We agree with the reviewer but it is worthy to note that this paper was submitted in December 2022, and the Kaufman and Broadman 2023 paper was not published at the time of submission. We acknowledge that this paper revisits the Holocene "conundrum" and adds to the overall discussion of Holocene trends and a cooler or warmer Late Holocene. We revisited the text in the discussion in section 4.3 and added the following information (including the suggested reference).

A significant discrepancy between Holocene cooling deduced from proxy reconstructions (e.g. Marcott et al., 2013; Kaufman et al., 2020; Figs. 8B and 8E) and Holocene warming simulated in climate models has been called the Holocene "temperature conundrum" (Liu et al., 2014), revisited in Wanner et al. (2021), Kaufman et al. (2020) and Kaufman and Broadman (2023). This discrepancy has been attributed to uncertainties in both proxy reconstructions and climate models. One major aspect is the seasonal bias in organic-based proxy records (such as pollen, diatoms etc.) towards summer (Liu et al, 2014). As the Holocene displays opposite trends in summer and winter insolation at 60°N (Fig. 8E, Laskar et al., 2004), high latitudes provide an optimal setting for testing the seasonality aspect in Holocene temperatures. Permafrost ice wedge $\delta^{18}O$*, a clear winter-season proxy, shows a continuous warming trend in the last 7 kyr*

BP (Meyer et al., 2015b) that supports the hypothesis that seasonality in proxy-based records is one key variable to be considered. Pollen-based reconstructions for the Holocene show a clear temperature decrease since an early to mid-Holocene temperature optimum (Fig. 8D; Herzschuh et al. 2021, 2022, 2023), not only valid for the summer season, but, though less pronounced, also for annual reconstructions. Proxy-based reconstructions of Kaufman et al. (2020) also suggest Late Holocene cooling. The $\delta^{18}O_{BSi}$ compilation presented here for lacustrine environments is based on diatoms whose bloom is mostly attributed to the late spring-early summer season.
Decreasing summer temperatures throughout the Holocene would manifest in decreasing T_{lake}, which would in turn lead to increasing $\delta^{18}O_{BSi}$ (Fig. 1). The observed decrease of $\delta^{18}O_{BSi}$, however, points towards lower $\delta^{18}O_{lake}$, which would be in line with decreasing $\delta^{18}O_{prec}$ due to decreasing summer T_{air} (Fig. 1). This would imply either a prevalence of summer precipitation or lake basins with sub-annual T_{res}. However, the trend is observed for lake basins with a wide range of T_{res}, suggesting millennial-scale changes in summer T_{air} to be the main driver of the observed $\delta^{18}O_{BSi}$ signal.(L632-649)

Figure 8: I am asking me whether the selection of the time series shown is significant. In contrast to the data of the present study, the time series of Marcott et al. (2013) is primarily based on marine data. I recommend that the latest reconstruction by Kaufman et al. (2020) be used here. Herzschuh et al. have just presented a new reconstruction in Climate of the Past.

Answer: We understand the reviewer concerns and suggestion, but, as commented above, this paper was submitted in December 2022, and the Herzschuh et al. 2023 paper was not published at the time of submission. To account for the reviewers' remarks, we updated the time series with newer Herzschuh et al. papers. However, the Pangaea data publication mentioned as Herzschuh et al., 2021 in the first submission relates to the same LegacyClimate 1.0 dataset based on pollen-based climate reconstructions from 2594 Northern Hemisphere sites covering the late Quaternary (Herzschuh et al. 2022 ESSD), which has been further scientifically explored in Herzschuh et al. 2023CP. We refer to all three publications now.

The Marcott et al. (2013) reconstruction is a fundamental and widely used publication, which shows a double peak in Early Holocene, also visible in our dataset, whereas most other reconstructions are rather coarse (including the suggested publication by Kaufman et al. 2020). Moreover, part of the added values comes from comparing our compilation to reconstructions from other proxies and environments (Vinther= glacial, Herzschuh= terrestrial, Lisiecki and Raymo=marine, Marcott=multi-proxy), so we think it is worthwhile including the time series of Marcott et al. (2013). We therefore would like to leave Fig 8B as Marcott 2013. However, we also implemented Kaufman et al. (2020)-based temperature anomaly for 30-60° and 60-90°N as new Figure 8E.

It would be exciting to further consider why the Eurasia and North America curves in Figure 8 F diverge in the early Holocene. How far are the detected time series temperature-sensitive or humidity/evaporation/precipitation-sensitive?

Answer: Touché. The diatom isotopes reflect the isotope composition of the host lake water, with different reaction of the individual lakes, their catchment, settings etc. to the environmental drivers. We are thus, with isotopes alone, not able to disentangle between temperature-sensitive and precipitation-sensitive records i.e. a higher amount of winter precipitation in a given period of time or a generally colder phase would react similar. Both would lead to lower $\delta^{18}O_{lake}$ and hence lower $\delta^{18}O_{diatom}$. In our compilation, however, it is obvious that on longer time scales, the records follow the insolation and NH temperatures quite well. This might definitely be different for individual records (i.e. at Bolshoye Shuchchye lake $\delta^{18}O_{diatom}$ seems to follow a long-term cooling trend over the Holocene overprinted by short episodes of snow meltwater influx, so a clearly precipitation-driven signal). As outlined, the issue of temperature vs. precipitation sensitivity is difficult to disentangle. This has also often been attempted in the original publications as well. We are aware that Herzschuh et al. (2022) GRL tried to separate precipitation and temperature sensitivities from pollen-based reconstructions. With our $\delta^{18}O_{diatom}$ alone, we are not able to give this information. We stress that we assessed the records on a different time scale than the original publications and therefore might see different factors prevailing in the binned records difficult to be linked to the original interpretations.

MIS 2 (Fig. 9): As requested above, I think it is right that the course of the individual curves in Fig. 9 A is interpreted in terms of dynamical processes. This should be attempted even more strongly; in particular, it would make sense if the origin of the individual figures were indicated in this figure.

Answer: As outlined above, and in agreement with Reviewer #2, we omitted Fig. 9B and hence, the differences between MIS1 and 2 on the spatial scale. For the time series in Fig. 9A (now Fig. 9), we added the record numbers (#3,8,15,24).

References:

Büntgen, U. et al., 2016. Cooling and societal change during the Late Antique Little Ice Age from 536 to around 660 AD. Nature Geoscience 9(3). DOI:10.1038/ngeo2652.

Kaufman, D., McKay, N., Routson, C. et al. (2020b) Holocene global mean surface temperature, a multi-method reconstruction approach. Scientific Data 7: 201.

Kaufman, D.S., Broadman, E., 2023. Revisiting the Holocene global temperature conundrum. Nature 614, 425-435. doi: 10.1038/s41586-022-05536-w.

Liu, Z. et al., 2014. The Holocene temperature conundrum. Proceedings of the National Academy of Sciences of the U.S. 111, E3501–E3505.

Wanner, H., 2021. Late-Holocene – cooler or warmer? The Holocene, 31(9), 1501-1506.

Wanner, H. et al., 2022. The variable European Little Ice Age. Quaternary Science Reviews 287, 107531.

Answer: All references listed here are now included in the new version of the manuscript