

“Climate indices in historical climate reconstructions: A global state-of-the-art”

Changes to manuscript in response to review comments

1. Response to Anonymous Reviewer #1 [RC1]

We thank the anonymous reviewer for their time and thought, which will help to improve significantly the overall quality of the manuscript. We respond to each question raised in turn:

[RC1] In the introductory part, three main categories of information are mentioned that appear in historical documents and inscriptions (lines 32–35) and in the following paragraph authors state that the generation of ordinal-scale indices is a common approach for the analysis of the third category – descriptive (or narrative) evidence. However, in the following sections, they mention numerous examples of indexing approach also for the two remaining categories – e.g. sea-ice index (Ogilvie, 1996), phenology-related phenomena from China (section 3.5) or even indices derived from early instrumental measurements (Figure 5 or section 6.2, lines 519–520). I would very recommend to provide somewhere in the introductory part at least some explanation why such type of information (quite often already existing at least on the ordinal scale) is transformed to indices. It would be quite useful to add some simple categorization of indices.

[Response to reviewer] The reviewer makes a good point here. We do indeed include examples where ordinal scale data are converted to indices as part of the reconstruction process, and this is especially true for regions outside Europe or at its margins, where narrative information is less available. Almost invariably this occurs when quantitative data are integrated with information from narrative sources to generate indices. Even where instrumental measurements or quantifiable phenological data exist, it may be desirable to develop ordinal indices so that these quantitative data can be combined with descriptive, qualitative information. In this way, it is possible to develop longer, more continuous and homogenous series with a consistent resolution (monthly or seasonal) and hopefully reconstruct both low-frequency and high-frequency variability. To address this point, we will add additional text to the introduction to explain why this is the case and reiterate this point where appropriate in relevant sections of the manuscript (e.g. in the sections on African and Asian index series).

[Changes made] Additional text has been added to the final paragraph of section 1 to address this point, with examples flagged in sections 2 to 7.

[RC1] Sections 2–7 provide a detail overview of various index types that different authors compiled at individual continents and ocean according to the meteorological element reconstructed. Too much space is devoted to the scale of index series. At the same time, it is mentioned several times in the text that number of points (or granularity) is dictated above all by the quality and abundance of documentary evidence (e.g. lines 136, 614). In my opinion, more information should be provided on different characteristics of the index series in this part of the text. Those are e.g. the completeness of the index series, their temporal coverage, the way the missing information is handled, meaning of the “zero” category, overlap with the target data for quantitative reconstruction and so on. Authors mention such characteristics only sporadically.

[Response to reviewer] We take the reviewers point here. We will edit the text to reduce descriptions of the scale of index series and remove any repetitive statements about how the quality and abundance of documentary evidence influences the granularity of index series. We will also add information throughout the manuscript on the completeness of index series and their temporal coverage. The comment about how missing information is handled is a

particularly important one. There are two main approaches used to define “0 index” values. One – implicit in the Pfister method – is that no description means no number: a gap in the time series rather than a 0. Other studies make an implicit assumption that, in some circumstances, no weather description can be taken as an indication of normal conditions. We will insert additional text about this in Section 8 and include a paragraph in section 9.2 where we discuss confidence and uncertainty in index-based climate reconstructions.

[Changes made] We now note in the introduction that the quantity, resolution and/or richness of the original historical evidence influences the granularity of any reconstruction and have edited down repetitive mentions elsewhere. We have flagged up how “0 index” values are generated throughout section 8 and added text on this issue in section 9.2.

[RC1] The 3.2 section provides very detailed description of diverse Chinese documentary sources, often not used for index series construction. Moreover, this part is quite long, not directly related to the topic of indices in some cases and it has no corresponding counterpart e.g. for Europe.

[Response to reviewer] We thought carefully about exactly this point when we were compiling the original manuscript. The nature of documentary sources is well discussed in climate history literature for most parts of the world. However, to our knowledge, there has been no corresponding detail made available for the diverse range of Chinese documentary sources. Hence, even though this text adds to the length of the manuscript, we consider it important for a climate history and historical climatology audience. The same is true for Japanese and early Russian materials, hence the reason we also say more about sources for these regions. We will add an objective to the paper regarding ‘the promotion of studies from regions beyond Europe’ to encourage specialists in these areas to engage in further work on climate index production.

[Changes made] We have added text to the final paragraph of section 1 to justify the inclusion of greater detail about Chinese, Japanese and Russian sources. In the case of Chinese documents there are only a handful of overviews of source types, and for Japan and Russia, to our knowledge, none.

[RC1] Section 8 on methodological approaches used to derive indices appears the most important for those searching for “good practice for future studies” and for advice how to derive indices from their own data. In this sense, however, at least some approaches mentioned here would deserve a short comment or some sort of critics (Section 8.3, end of the first paragraph: Correlation coefficient is a relative measure and the value of 0.5 means that compared data sources share only 25% of common variability. Statistical significance of the correlation would be much more relevant).

[Response to reviewer] Thank you for this observation. We will review the text in section 8.3 to ensure that the discussion of index development is sufficiently critical.

[Changes made] Two sentences of additional explanation have been added to section 8.3.

[RC1] The same holds for some statements in Section 9. Please check lines 821–829. The whole paragraph is hard to understand and it does not make sense – at least from statistical point of view. It is not clear how “... chi-square tests, comparisons with the eigenvectors ... and the standard error of the estimate” can be used “to derive transfer functions”. For instance, the standard error of the estimate is the result of the transfer function calculation. Thus, it cannot be used to derive it. Similarly: “Such correlations can further be compared and calibrated using instrumental data”. Please re-formulate as correlations (of what?) can be hardly “calibrated”.

[Response to reviewer] Thank you for this comment. We will review the text in lines 821-829 to improve readability and ensure that it is accurate in its use of statistical terminology.

[Changes made] This paragraph has been edited as identified above.

[RC1] In section 9.2 on confidence and uncertainty there is a discrepancy between the title of this section and the text that follows. Both types of uncertainty are very important, however, they have several different reasons and different origin. Unfortunately, the text provides only some examples of the second type of uncertainty (related to the index series compilation). It would be very useful to mention at least some examples of the first one (Dobrovolny et al., 2010). Ability to quantify uncertainties in the index-based reconstructions (either formally – with some statistics or less formally – by comparison with other reconstructions) makes them fully comparable to natural proxy-based quantitative reconstructions.

[Response to reviewer] Thank you for this very helpful comment. In this section, we are focusing mainly on uncertainties related to index series compilation. We will clarify the text to make sure that this is obvious to the reader, but also mention the suggested example of wider uncertainties in index-based climate reconstruction.

[Changes made] We have made edits throughout section 9.2 to (i) note the two forms of uncertainty and (ii) add in some examples of uncertainty in reconstruction.

[RC1] It is obvious that this overview cannot refer to all relevant studies. However, I would recommend to mention in the text several other studies especially from Europe. They can be an important example of the indexing approach (Koslowski and Glaser 1999; Dobrovolny et al., 2015), example of multiproxy reconstructions using temperature (Luterbacher et al. 2004) or precipitation indices (Pauling et al. 2006) or papers important from the methodological point of view (Dobrovolny et al. 2009, Brázdil et al. 2016).

[Response to reviewer] Thank you for these helpful suggestions. We will review each of the recommended papers and add them to the manuscript where appropriate.

[Changes made] Where appropriate, we have woven the recommended papers into section 2 – but note the caveat added to the end of the introduction that we do not include studies unless they include primary documentary reconstructions.

[RC1] A suggestion for the Section 10.2, concluding recommendations: Even if the index series are constructed at several-degree scales (7 or more points), indexing always means suppressed variability of index series compared either to target data (instrumental measurements) or to natural proxies (e.g. tree rings). It is advisable to sum-up index series – either in time (from monthly to seasonal or annual) or in space (put together several index series from climatologically homogeneous region). This approach may well approximate index series to natural climate variability.

[Response to reviewer] Thank you for these helpful suggestions. We will add a bullet point to this effect to the series of recommendations in section 10.2.

[Changes made] We have added a new recommendation (8) to reflect this point.

Minor comments

[RC1] Line 43 – the term "unweighted" index may be misleading here.

[Response to reviewer] Thank you. We will clarify the text.

[Changes made] We have removed the word 'unweighted' from the sentence for clarity.

[RC1] Line 396 – "...that Henry Lamb was developing..." Here should be "Hubert Lamb", I guess.

[Response to reviewer] Well spotted!

[Changes made] Text corrected.

[RC1] Line 626 – "...to define index categories: -/+180% for index values -3/+3, -/+130% for values -2/+2, and +/-65% for values +1/-1." Percent of what? This text is confusing. Please add more explanation.

[Response to reviewer] Thank you. We will clarify the text.

[Changes made] More explanation has been added.

[RC1] Line 637 – add "decadal" otherwise not clear: "where... is the DECADAL winter temperature index..."

[Response to reviewer] Thank you. We will clarify the text.

[Changes made] Corrected.

[RC1] Lines 694 – 695 "...the presence of key descriptors is used to distinguish these categories." Not clear, please re-formulate.

[Response to reviewer] Thank you. We will expand the text to clarify this.

[Changes made] Text clarified.

[RC1] Line 696 – "Algorithms are then used to weight and combine documentary and instrumental data" Not clear, please re-formulate.

[Response to reviewer] Thank you. We will expand the text to explain this more fully.

[Changes made] We have added two short sentences to section 8.3 to clarify this.

[RC1] Table 3, 5 – There are some empty fields, please add something like "not available" or "not relevant" to avoid misinterpretation.

[Response to reviewer] Thank you. In the case of Table 3, the problem arises from having five index classes in the middle column and only four classes in columns one and three. We will review to see if we can present the table more clearly. For Table 5, we will add text to the table caption to explain the empty fields.

[Changes made] We have reviewed Table 3 and can think of no better way to present the data. We have updated Table 5 to provide clarity.

[RC1] In case of Table 5 please explain "qualitative indication" XXX means the best quality?

[Response to reviewer] Thank you. We will clarify this in the table caption.

[Changes made] We have revised the table and expanded the caption.

2. Response to Anonymous Reviewer #2 [RC2]

We thank the anonymous reviewer for their time and thought, which will help to improve significantly the overall quality of the manuscript. We respond to each question raised in turn:

[RC2] Regarding objective one [of the paper: ‘provide a global state-of-the-art review of the development and application of the index approach in historical climate reconstruction’], the review is exhaustive reflecting most of the previous work based on ordinal indices that consider the departure from normality as the main criterion to produce an anomaly scale with several levels of intensity. However, references to other approaches to build ordinal indices are missing. For instance, several papers have built ENSO chronologies from documentary evidences from different areas of S America reporting different impacts associated to ENSO (Quinn and Neal, 1992; Ortlieb 2000; Garcia-Herrera et al 2008). In my view this type of approach should also be acknowledged in the paper.

[Response to reviewer] Thank you. This is a very valid point. We will add information about index-based approaches to the development of ENSO chronologies to the most relevant part of the manuscript (Section 5, dealing with ‘Climate indices in the Americas’).

[Changes made] We have added a short paragraph on ENSO chronologies to section 5 and an account of how uncertainties are dealt with as part of index compilation in section 9.2.

[RC2] Lines 777-778 In the recent years directional wind indices over the oceans have gone beyond decadal reconstructions of wind force trends, as stated in the paper. This methodology has allowed the generation of the longest series of the wind circulation in the North Atlantic and generating new indices for circulation patterns as the NAO or the East Atlantic pattern (Mellado-Cano et al 2020). Besides, they have been useful in studying different features of the global monsoon system: the impact of volcanic eruptions on the West African Summer monsoon during the 19th century (Gallego et al 2015), the onset of the Indian Summer Monsoon (Ordoñez et al 2016) or secular trends in the Australian Summer Monsoon (Gallego et al 2017) among others. Lines 865-873. Over the Oceans the uncertainties associated to the limited sampling in a given area and period have been also quantified, see for instance Gallego et al (2015).

[Response to reviewer] Thank you for these very helpful comments. We will review each of these studies and update sections 7, 8.6 and 9.1 where appropriate.

[Changes made] We have edited section 8.6 in light of these suggestions.

[RC2] In my view the second objective is not achieved because there is not a critical analysis of the work described in sections 2-9. Consequently, the link from the recorded evidences to the identification of the best practices is missing. This should have been done in section 10, but this is again very descriptive. Tables 5-7 do not identify best practice, instead they just summarize the variables studied in every region or the number of classes used. Having missed this analytical part, many of the statements lack of support. The authors claim that they are based on two previous reviews and ‘also incorporate insights from this study’. This is not evident at all from the text, because of this lack of critical analysis in the manuscript. For instance (lines 950-952) why do the authors “recognise that the most widely used approaches such as the Pfister method would require modification to be useful for temperature and/or rainfall reconstruction in all regions”? Which of the previous papers are the support of this statement? What are the main reasons for this recognition? The authors do not provide any evidence of the limitations of these approaches and they should do it based on their extensive previous review. Is it because indices derived from a certain type of documents and for a given climate cannot be applied mimetically to different

documentary sources and climates? If so, the authors should provide supporting evidence. Otherwise this is just their opinion.

[Response to reviewer] We take the reviewer's point about lines 950-952 and will expand the text as suggested. We will also review section 9.1 in general to ensure that statements are backed up with examples from the preceding sections. We do not, however, agree with the other views - particularly the suggestion there is insufficient critical analysis of the examples discussed in sections 2 to 9. We embed critical analysis throughout the manuscript. The global coverage of historical climate studies is such that, for many parts of the world, there are not overlapping series that would allow for a direct comparison of the outcomes arising from the use of different methodologies. Where historical studies do overlap, for example in Europe and Africa, we have commented on similarities and differences. Indeed, figure 2 explicitly presents the results of two different studies of overlapping areas in a European context. The purpose of section 9, and section 9.2 in particular, is to not only synthesise the different approaches used to reconstruct climate indices in different parts of the world but also to identify weaknesses. Many of these weaknesses feed directly into the recommendations in section 10.

[Changes made] We have edited the second paragraph of section 10.2 in light of the reviewer's comments about lines 950-952 in the original manuscript.

[RC2] I find several problems with the guidelines. Firstly, they should be clearly supported from the previous review, which is not the case. The review should allow identifying best practices and the analysis of these cases should lead to the guidelines, but this sequence is not followed in the paper. Best practices are not identified, and, consequently, guidelines are not supported by them.

[Response to reviewer] As discussed in our previous response we do not agree with this viewpoint. Further, section 10 is intended as a conclusion and synthesis. Adding supporting evidence to underpin each of the 12 recommendations would add unnecessary length to an already very long manuscript.

[Changes made] No changes are necessitated in response to this comment but see below.

[RC2] Additionally, I think that a climate component is missing in some of the guidelines. According to my previous experience, the final indexation should be a compromise among the historical records characteristics, their availability and the climate of the region to be studied. The mere translation of indices built for a certain climate to other areas may lead to biases or inadequacies if applied to other regions. For example, the translation of precipitation indices developed for central Europe should be applied with care to Mediterranean climates, where most of the precipitation is concentrated between September and April and occurs mostly in the form of a few intense events. Thus, I think that guideline 2 could be rephrased from: Researchers should be familiar with the strengths and weaknesses of each of their historical sources prior to their use in climate reconstruction. To something as: Researchers should be familiar with the local climate and the strengths and weaknesses of each of their historical sources prior to their use in climate reconstruction.

[Response to reviewer] We agree fully with the reviewer here. However, the recommendations need to be considered as a whole. We already discuss the idea that indices should be developed for climatically homogeneous regions in guideline 1. We do not anywhere suggest that a one size fits all approach to index development would be appropriate. There are numerous examples in the manuscript of where approaches have been tailored to suit climatic variability in an area of interest.

[Changes made] We have edited guideline 1 to reflect the need for awareness of local climatic conditions. We specifically mention the case of Mediterranean climates in para 2 of section 8.1 as part of cautionary sentences about using monthly indices.

[RC2] Guideline 3 reads: “Researchers should select an appropriate temporal resolution for their index series according to the quantity and richness (in terms of climate information) of available historical sources. This may be monthly, seasonal, annual or longer, although for information-rich areas, a monthly resolution is the most desirable.” Again the climate factor is missing, for instance if you build monthly series for precipitation in the Mediterranean, you should be aware that during the dry months, the signal-to-noise ratio should be very low and this can bias the results. This guideline should be rephrased, as, for example: “Researchers should select an appropriate temporal resolution for their index series according to the quantity and richness (in terms of climate information) of available historical sources and the local climate. This may be monthly, seasonal, annual or longer, although for information-rich areas, a monthly resolution is desirable depending on the climate type and variable studied.”

[Response to reviewer] The key aspect to this particular guideline is that researchers should select an appropriate temporal resolution for their index series based on their data. If, due to the climatic characteristics of an area, observations are relatively sparse for particular seasons then it may not be appropriate to adopt a monthly time scale, regardless of how rich the observations are for other periods of the year. We think that this is explicit in the guideline as it stands.

[Changes made] We have edited guideline 3 to make the impact of variations in source density across the year clearer. Note also the changes made in relation to the previous point, which address this issue.

[RC2] Guideline 4 reads: “Whether to develop a three-, five- or seven- (or more) point index series will also depend upon data quantity and quality but may be influenced by the legacy of previous studies in a region if direct comparisons are required”. Two comments here. I do not understand the mention to the previous legacy, I find this confusing. Do you mean that things should be done as they were done in the past, just to compare? Even if you have identified problems in the legacy? This needs to be clarified. Applying previous indexation without a careful examination of its adequacy to a given climate and data set is not a good practice. Apart from this, a mention to the climate is also required, since the number of points in the scale may also depend on the type of climate and climate variable studied. So, for instance, this could read something as: “Whether to develop a three-, five- or seven- (or more) point index series will also depend upon data quantity and quality, the local climate and climate variable to be indexed”.

[Response to reviewer] We take the reviewer’s points about reference to previous studies and will clarify the text. We disagree, however, about the need to include reference to climate and climate variability in the guideline. The temporal resolution of any index series hinges on the richness of available data across the year. If this varies seasonally then the resolution of the series for the whole year should reflect this.

[Changes made] We have edited guideline 4 to provide clarification.

[RC2] Guideline 9 reads “To maximise their wider usefulness, index series should, ideally, overlap with runs of local or regional instrumental data to permit calibration and verification. Where instrumental data are not available, overlaps with independent high resolution palaeoclimate records may be used for calibration” I think that using palaeo-climate proxies to calibrate an index is not the best recommendation. These proxies have their own weaknesses and uncertainties and using them as the ‘truth’ to calibrate an index may introduce unexpected biases. Calibrating an index with a proxy, implies two transfer

functions from the variable to the proxy and from the proxy to the index, posing additional uncertainties. I think that a comparison with proxies is fine, but using them to calibrate is far too dangerous.

[Response to reviewer] The reviewer correctly identifies that there is a controversy here. We will address this controversy through the insertion of additional text describing, for example, Andrea Kiss' or Martin Bauch and colleagues' work comparing written records and indices with the Old-World Drought Atlas, as this illustrates the issue well. We do not, however, wish to modify the recommendation. The key words here are "high resolution". We would not recommend calibration using low resolution palaeoclimate series.

[Changes made] We have added text to section 9.1 to address this and highlighted the need to use high-resolution palaeoclimate data only.

[RC2] Summing up, I think that the paper requires an extensive revision before being acceptable for publication. The good practices need to be well identified in the text and the support of the guidelines must be clearly linked to the previous evidence. The authors have made a highly valuable effort in compiling the previous work. Improving the analysis by better illustrating the good practices and providing a clear background and support to the guidelines, would lead to a highly interesting paper, but these issues need to be solved.

[Response to reviewer] Thank you for this. We will adjust the text as outlined above.

[Changes made] This comment requires no changes other than those described above.

[RC2] Minor issue Some of the authors references are made in a strange way. For instance, line 455 'Garza Merodio who was a student of : : :.' Is this so relevant? Why are not the other academic linkages mentioned? To me this is relevant if you want to tell the history of the researchers involved in this topic, which is not the case. Line 401 'the work of Coleen Vogel'. Line 407, ' Sharon Nicholson' and several others. Why some authors are cited by their full names (not the usual practice) and other just by the surname? Not clear to me.

[Response to reviewer] First names are used sparingly throughout the manuscript to flag up key researchers who made important contributions in specific regions and/or to identify distinct schools of historical climatology that have transmitted certain methodologies.

[Changes made] This comment requires no changes other than those described above.

3. Response to Short Comment by Domínguez-Castro and Vaquero [SC1]

We thank Drs Domínguez-Castro and Vaquero for their time and thought, which will help to improve significantly the overall quality of the manuscript. We respond to each question raised in turn:

[SC1] The documentary sources provide a huge quantity of information about climate variability of the past. Nevertheless, there is high variability in the quality, quantity, formats, contents, objectives, etc. of the climatic information provided by these sources. This makes very difficult to unify indexing methodologies for all the documentary sources. For this reason, a review article of the indexing methodologies and the identification of indexing best practice is necessary, pertinent, and really welcomed. We want to provide some comments that can be useful to broaden the discussion about indexing.

We consider that an article that wants to provide a “Guidelines for generating future documentary-based indices” must analyse the advantages, disadvantages as well as the limitations of the different indexing methods used until now. We think that this discussion is missing in the manuscript and we provide here some topics that could be interesting to address:

[Response to reviewer] Thank you for these observations. As noted in our response to RC2, we do not agree that there is insufficient critical analysis in the manuscript, but we respond now to each of your suggestions in turn.

[Changes made] This comment requires no changes to the text.

[SC1] Use of different kinds of documentary sources: there are indices constructed from a unique documentary source, e.g. a private diary (Brázdil et al., 2008; Domínguez-Castro et al., 2015), series of acts of municipal and ecclesiastical institutions for a location (Barriendos, 1997; Domínguez-Castro et al., 2018), series of correspondence (Fernández-Fernández et al., 2014; Rodrigo et al., 1998)... However, other indices are constructed putting together information from many different documentary sources (e.g. Brázdil et al., 2016; Camuffo et al., 2010). Probably in the first case, it is easier to detect and correct unexpected bias or homogeneity problems. In the second case you can usually analyse longer periods or larger regions, but at the risk of including inhomogeneities.

[Response to reviewer] This is a useful point. We will review the manuscript and distinguish examples where we have indices based on single documentary sources or phenological proxies. This will need to draw on examples from a wider geographical range than the European examples suggested. We will also emphasise the comparative ease of detecting unexpected bias or homogeneity problems in single source series versus multiple source series.

[Changes made] The opening paragraph of section 9.2 now directly addresses this issue. We have also checked through the text and flagged up any examples where indices are based on single documentary sources.

[SC1] Subjectivity in the indexation: probably, the unique objective indices are those based on the presence or absence of a meteorological phenomenon, e.g. rainfall, snowfall, wind, fog... If one has this information at daily scale, the index is almost directly comparable with the instrumental series e.g. Domínguez-Castro et al., 2019. It is easy to think that the subjectivity increases with the number of categories of the indices, but more categories provide more variability. On the other hand, there are also many “uncategorized indexes”, a few examples are: appearance of ice and freeze-up dates (Takács et al., 2018), flowering

/grape harvest dates (Aono and Kazui, 2008), number of days under drought conditions (Domínguez-Castro et al., 2008)... This difference could be discussed deeply.

[Response to reviewer] We take your point about subjectivity in indexation (which has already been discussed at length in the literature). However, we are not sure of the relevance of what you refer to as ‘uncategorized indices’ for our paper. Taking the example of ice phenology from Takács et al. (2018), this study focusses on the timing of freeze up rather than developing indices of winter severity from such information. The mere existence of a dataset that might be used to produce indices does not, in our eyes, qualify a study for inclusion in our already lengthy manuscript. We will, however, check the suggested literature to make sure we haven’t missed any examples of actual index approaches (note: Aono and Kazui (2008) is mentioned in section 3.1).

[Changes made] We have reviewed the text as indicated and added a caveat to the end of the introduction justifying the inclusion (and exclusion) of studies from the manuscript.

[SC1] Resolution: the time resolution of an index can be daily, weekly, monthly, seasonal, annual or larger time resolution. Are there problems when, for instance the annual/seasonal indices are built just adding seasonal/monthly indices? Frequently, the documentary sources have bias to some season due to the climate and/or the nature of the documents. It is common that the documentary sources provide more information about extreme seasons (winter or summer), or in specific periods of the year in which the climate was determinant for some agricultural labour. For this reason, one must be very carefully when computing annual index just from the addition of seasonal indices and this could be reflected in the text.

[Response to reviewer] Thank you for this comment. This is a good point and we will integrate additional text into section 8 where relevant.

[Changes made] Additional has been added to paragraph 2 of section 8.1 to address this comment.

[SC1] The meaning of “0 index”: frequently “0 index” is considered as “normal” condition, but this can be confounded with cases when no information is available. We need to be sure to assume it, that the consulted documentary sources cover the entire period with the same quality. This is easier to evaluated when only one documentary source is used for the whole studied period.

[Response to reviewer] This is an issue already referred to in the second point made by RC1 – we repeat our response here: “There are two main approaches used to define “0 index” values. One – implicit in the Pfister method – is that no description means no number: a gap in the time series rather than a 0. Other studies make an implicit assumption that, in some circumstances, no weather description can be taken as an indication of normal conditions. We will insert additional text about this in Section 8 and include a paragraph in section 9.2 where we discuss confidence and uncertainty in index-based climate reconstructions.”

[Changes made] Please see changes made in response to a similar comment by RC1.

[SC1] The distribution of the index values: For reconstruction purposes, it is useful that the index shows similar distribution to the variable to be reconstructed. Nevertheless, this is not always possible, as in some examples provided in this paper (e.g. figure 2 shows a possible bias to negative values, and Mertz’s reconstruction shows less “-1 values” (14) than “-2 values” (20), when probably the opposite is expected). Something similar happens in Figure 8 that shows less “1 values” (abundant flows, mainly concentrated in the last decades) than large swells (“2 values”). Probably, this is because the used documentary sources record

better the extreme than the more common events. Anyway, we think it is an interesting topic to discuss in the paper.

[Response to reviewer] This is a valuable point that we will mention in section 9.2 on uncertainty (also with reference to a recent publication by White, Pei 2020).

[Changes made] We have added an additional sentence on this issue to paragraph 2 of section 9.2.

[SC1] In general, the classic methodology that uses three, five or seven indices is deeply discussed in the paper compared to other methodologies for each type of documentary source that have appeared in recent years. We consider that these particular methodologies have enormous advantages for performing reconstructions and they should be discussed.

We understand that it is impossible to cite all the published articles in the field or to analyse all the methodologies developed in a single paper for all the meteorological variables or events. However, we think that it is possible to include more recent publications. Here we provide some examples.

Regarding the European temperature indices, the more recent work cited dates from 2015. We recommend to include some recent papers as (Brázdil et al., 2019; Fernández-Fernández et al., 2017; Fila et al., 2016; Filipiak et al., 2019; Mrgic, 2018; Rodrigo, 2019). Some of the methodological approaches of these papers are interesting and have not been discussed in the manuscript.

With respect to European precipitation indices, the most recent work cited dates from 12 years ago. Some proposals to update these references are (Brázdil et al., 2019; Bullón, 2011; Fernández-Fernández et al., 2015; Filipiak et al., 2019; Metzger and Tabeaud, 2017; Rodrigo, 2019). Again, these papers have interesting methodologies that have not been discussed.

[Response to reviewer] We accept the general point made here. We will evaluate the methodologies of the more recent literature mentioned to see if they warrant inclusion in the revised version of the manuscript. Again, the key criterion for the inclusion of any study is that it applies an indexing approach in the wider tradition described for Europe-focused research. Mere overviews on available documentary information for a specific period and region (e.g. Mrgic 2018 for the Balkans) will not qualify a study for integration. We realise that we handle this differently at times for non-European regions, as we see value in pointing the anglophone academic community to these regions. However, there is already plenty of information on available sources for Europe-focused climate historical research.

[Changes made] We have checked through the list of more recent publications provided by the reviewers and, where the studies fit the brief for inclusion stipulated in the revised opening paragraph of section 2.1 (i.e. they include original published index series based on primary sources and reconstruct meteorological entities), have incorporated them into the manuscript.

[SC1] Moreover, the authors affirm in the section “Climate indices in Europe” that the information about temperature is more frequent and with a better quality than the precipitation information (e.g., line 124 “Temperature is the most common meteorological phenomenon analysed in Europe” or line 154 “Often the same scale is applied for both temperature and precipitation indices; however, precipitation indices may show more gaps than their temperature counterparts as data may be seasonal or more sporadic”). We think this assumption is true for central and northern Europe, but it is not for the Mediterranean region, for example, where the references to the lack or excess of precipitation are clearly

more frequent than those for temperature. This is because the Mediterranean climate is more temperate but has important precipitation extremes. This is a very important point in the article, because the effect of the different climates in the indexing is almost missing in the review, and this is a key issue in a review wanting to cover different regions of the globe.

A mention to pro pluvia rogation ceremonies as a documentary proxy of European droughts is missing. This proxy has been used in the last decades to understand drought variability in preinstrumental period in various European countries as Spain (Barriendos, 2010; Bravo-Paredes et al., 2020; Domínguez-Castro et al., 2008, 2010, 2012; Tejedor et al., 2018), France (Garnier, 2010), Italy (Piervitali and Colacino, 2001) or Portugal (Fragoso et al., 2018). Moreover, different methodologies have been used to extract the climate information of this proxy as the liturgical act (Martín-Vide and Vallvé, 1995), the expenses (Álvarez Vázquez, 1986), the period of time with continuous rogations (Domínguez-Castro et al., 2008) or the area occupied by the text dedicated to the celebration in the chapter acts (Gil-Guirado et al., 2019). All these methodologies are different to the three, five or seven indices discussed in this paper, and have their advantages and disadvantages that would require a detailed analysis.

Moreover, it would be interesting to cite articles of other variables, as sea level (Camuffo et al., 2017; Camuffo and Sturaro, 2003) in which interesting methods have been developed or about snowfalls reconstruction (Enzi et al., 2014).

[Response to reviewer] We agree that our statement about temperature as the most common phenomenon should be differentiated with regard to the Mediterranean situation, and we will do this accordingly. The commentators identify correctly an incoherence in the manuscript. While we initially discussed and decided not to integrate proxy indices such as pro pluvia rogation ceremonies in sections concerning Europe, we softened this approach for non-European regions. We will clarify this in the revised introduction and will provide (in section 2.6) the most important references to this long-established index type. Space precludes a full coverage of these and comparable indices relating to Europe, so we will emphasise that our key focus is on narrative sources. Regarding the other variables mentioned (e.g. snowfall or sea level/submersions), most studies focus on reconstructions of occurrence or the creation of new proxy data, rather than differentiated index-like values. Hence, we do not intend to include these contributions.

[Changes made] We have differentiated between north/central and Mediterranean Europe more clearly in our statements regarding the relative importance of temperature and precipitation reconstructions. The rogation papers mentioned by the reviewer are very interesting. However, the indices applied in many of these studies indicate the types of rogation but not precipitation or temperature. This was the reason why the majority were not included in the original manuscript. We have reviewed the suggested studies, added mentions of those that include direct climate indices, and added a caveat to the start of section 2.1 stating that we only include studies that reconstruct meteorological entities in our review of the European literature.

[SC1] Additionally, we can point out one minor suggestion:

Line 474 “Finally, Dominguez-Castro et al. (2018) built a long precipitation series for 1891–2015 CE based on descriptions of rain ceremonies in Quito, Ecuador”. This is not correct. Domínguez-Castro et al. (2018) presents a precipitation instrumental series from Quito (1891-2015) and a series of wet and dry extremes from rogation ceremonies from 1600.

[Response to reviewer] Thank you for this correction, which we will apply accordingly.

[Changes made] Corrected.