Dear editor, dear referees,

We would like to thank the reviewers and the editor once more for their time spent on reviewing our manuscript and their comprehensive comments and ratings.

In the following we reply (normal font color) to the comments (red) and note implemented changes.

Best Regards

On behalf of the authors

Oliver Bothe

Editor

dear authors

I now got two reviews back on your revised version. The reviewers acknowledge the effort that have been made and they appreciate the work you have done for the revisions which have improved the workd in various aspects. Both have some additional comments/suggestions how the paper can be improved. I think both reviewers provide clear instructions for improvements. Can i therefore ask you to follow the reviewers comments and revise your manuscript accordingly?

with my best wishes and looking much forward to the new version

Jürg

Response to reviewer #1
Referee 1

Second Review of Bothe et al.

The authors have improved the paper in this iteration. Some further work remains to be done however.

First we would like to thank the reviewer for their comments.

Then, we would like to express a general comment. We read in the reviewer's comments a fundamental scepticism concerning all data sets that we use in our analyses as well as the analyses themselves.

Obviously, each of the data sets has its associated uncertainties and deficiencies. This is valid for observations, reconstructions, and model simulations. We explore here to what extent common variations can be expected and to what extent they are present in the used data sets. We do so not only for the time-series but also for further properties of the distributions of the precipitation data sets.

Our revisions try to consider the reviewer's comments as far as they are within the scope of this manuscript. We explain our reasons when we do not integrate all reviewer's suggestions.

Please see the document including tracked changes, for our edits.

The paper is longer than it needs to be, especially the discussion.

In the revised version, the discussions were notably shortened.

Following the suggestions of the reviewers we reconsidered all new results and moved part of these to the results section (see new section 4.5).

We moved all discussions of data sets and methods to the data and methods sections (see sections 2 and 3 and particularly the tracked changes) and tried to shorten these parts. We did not necessarily succeed as we see much of this as required by the previous round of revisions.

There is also work to be done on teasing out the key contributions. At least the paper should draw out where there is agreement and not between sources, ie. Provide the key periods. This would be usefully done in a table and would help focus future research to further understand the issues raised.

We do not provide a table but we try to clarify these points. Agreement and disagreement is more clearly highlighted throughout the results section (see section 4). The key points and key periods are further highlighted in the discussions (see section 5, and particularly section 5.2) and conclusions (see section 6).

The conclusions now try to clearly point to all the contributions of our work (see tracked changes in section 6).

The integration of tree rings (based on width) is done so in the full knowledge that previous work by the authors of these data sets have shown that there is a lack of stability in terms of the relationship with station based precipitation. This work confirms a well established fact, that there is inconsistency between tree ring reconstructions and station based observations. I am therefore not sure what this aspect adds. This is well known. Surely this paragraph from the discussion negates the use of the tree ring series you use for the purpose you use them for.

"Young et al. (2015) conclude that these differences make it unlikely that the tree-ring based works and their δ 18O based work represent the same environmental parameter, and they emphasize the lack of a calibration against regional precipitation data. They further discuss the reasons given by Wilson et al. (2013) for the lacking stability of the Wilson et al. reconstruction, namely, different climate-sensitivity of the trees, unreliable instrumental data, and pollution. Young et al. (2015) conclude that their own data reliably reflects precipitation while the tree-ring widths most likely represent the combination of various environmental influences on tree growth instead of a single climate parameter."

All data sets, i.e. reconstructions, models and observations, are uncertain. However, all data sets also contain some part of the information we are interested to tease out in paleoclimatology. We openly present these uncertainties.

If we take the logical consequence of the reviewer's somewhat radical interpretation and additionally consider that our supplementary analyses show comparable weaknesses in the isotope based records, then there are no reconstructions left to compare to until we obtain an isotope enabled regional simulation which we could directly compare to the isotope measurements.

Our analyses try to identify to what extent the different data sets agree or disagree in view of their respective uncertainties, and we discuss the possible reasons for these agreements and disagreements. As such, we are convinced it is a valuable contribution.

Our aims are a) "to study the consistency in the statistical properties of precipitation distributions in these sources of information" and b) to do so by using the SPI and "[motivating] the use of the Standardized Precipitation Index in hydroclimatic comparisons between different data sets in paleoclimatology". These sources of information here includes all available types of data, observations, reconstructions, and simulations.

We are convinced that the known weaknesses of the reconstructions do not negate either of these goals.

Climate models need to be tested at long multi-centennial timescales relevant for future climate change. At these long timescales, the available evidence from proxies and long instrumental records becomes more uncertain. Nevertheless, we think that our study is a valuable contribution towards that goal.

To help tease out inconsistencies I am not sure why you don't use series such as the Paris London Index, London Sea Level Pressure (both by Richard Cornes) or perhaps the Westerly Index to help point to which data sources that potential errors derive from. There is more data available to help this work. It seems a missed opportunity not to use them.

Indeed, it is of value to study every available data set giving information about British Isle climate over the last centuries - in the case mentioned by the reviewer since 1692. Our focus here is precipitation. Detailed analyses on the involved dynamics are beyond the scope of our manuscript.

This is not so much a decision on principle but the judgement that the various sources of information available (proxy-based index reconstructions, observation based indices, proxy-and observation-based field reconstructions, data assimilation efforts, and [regional] simulations) require a separate manuscript to assess them against each other. Indeed, Figures 11 and 12 of Cornes et al. (2013) indicate on the one hand large discrepancies between different estimates of North Atlantic Oscillation indices but on the other hand reasonable to strong common variations. That is, teasing out reasons for disagreement in precipitation estimates is only possible after teasing out reasons for the disagreement in circulation estimates.

More generally, using pressure related indices does not a priori provide evidence on the reasons for uncertainties and disagreement in and among the data sources we use.

Also, regarding the data, EWP is a composite drawing on stations as far north as Edinburgh. Yet the analysis, reconstructions and modelling is based on southern England. It is not surprising to me that inconsistencies exist given the study design.

We would like to point to a number of issues here.

First, the modelling is not solely based on southern England, we select only data from these domains from our simulation output covering the entire European realm.

Following the suggestions in the first round of reviews, we included a number of additional data sets. Using these data sets, we could see different levels of agreement among the observational or observation based data sets. These comparisons further show that the reconstructions compare worse to the local to regional observational data sets than the observational data sets compare among themselves (e.g., Figures 1 and 2, and Supplementary Figures 1 to 3).

In our understanding, this indicates that the inconsistencies are not due to the uncertainties in the data sets in general and not due to a biased choice of data sets in our study.

Simulation results are based on a single RCM model chain. Can this provide a robust estimate for comparison with observations and reconstructions? You have not convinced me.

To our knowledge, there are only two regional simulations for this area for the last few centuries. Both simulations are not totally independent, since both have been driven by different versions of the same global model from the Max-Planck-Institute.

Using an ensemble would of course make the results more robust. However, such an ensemble does not exist. Regional simulations are costly, and only few modelling groups run regional paleoclimate simulations.

As in the case of a single observational record or of a reconstruction without uncertainty estimates, the comparison cannot be complete. However, we are confident that there is an added value in such comparisons. Single simulations are often the only available spatially complete physically consistent information about past climates. Evaluating how they compare to reconstructions and observations is in itself of value for our understanding of past changes. We do this for this simulation, the reconstructions, and the observations in our chosen study region.

Any single simulation has to be evaluated. We do this for this simulation in this region.

The discussion is far too long and contains some repetition, making it difficult to read. It needs to be considerably shortened and sharpened in its focus on key points. It needs to distil the key contributions of this work. The structure is also odd, there are concluding remarks followed by conclusions, and section 5.3 is Discussion of Results? The way it is organised re-traces the entire paper again. It needs to get to the job it is supposed to do, i.e. discuss key points from the findings.

We shortened the discussion and tried to focus it more. The concluding remarks for the discussions were removed. Discussions of data and methods were moved from section 5 to sections 2 and 3 respectively. The additional results were moved to section 4.5 and shortened. The discussions, thus, solely focus on the discussion of our findings, their validity, and their implications (see section 5).

The conclusions in section 6 try to highlight our main points.

It is my opinion that you give too great a credit to the early observations. These are likely to be highly uncertain. Indeed for many of the series you use the early parts are composites taken from weather diaries and very early instrumentation. In addition do you use the corrected Oxford series of Burt and Howden who corrected for gauge being on a roof and changes in gauge design? Or is it the original series? You could look more critically upon the observations.

As the reviewer may see from our section on the observations and our table, the observational series stem either from the Met Office homepage or are from the Global Historical Climatology Network (GHCN) in the version as provided by the Climate Explorer. For the GHCN, please see Peterson and Vose (1997).

We already acknowledged that the observations also have large uncertainties in the previous version. This study is a comparative analysis of the available information - observations, proxies, and modelling. We provide possible explanations for the lack of agreement among these data sets when discussing the data sets in section 2 and in section 5. We cannot revisit all single time series from publically available compilations.

We do not further address the uncertainty of the observations beyond what is already in the paper.

Given that we know the reconstructions have issues, it would be useful to highlight to the reader the particular periods where we have agreement between the different sources and those that we do not. This might help focus and provide a priority for future work. What I am saying is that the results need to be consolidated and key insights distilled. What you are saying at the moment is that there is disagreement between the observations and reconstructions, disagreement between both of these and model simulations but the latter is not as different to the observations as the reconstructions are. I need more insight on this - are there periods where all agree/disagree etc.

We try to clarify this and to highlight the disagreement and the agreement throughout sections 4 to 6. Indeed, already the last version included many pointers to the specific disagreements and agreements in the text of section 4. This allows any interested reader to find periods for future research foci. We try to extend these pointers where it appeared necessary.

We note that given the heterogeneity in the patterns of agreement and disagreement between the different sources it might appear overconfident on our side to point to specific periods for showing an exceptional high degree in consistency.

Please note that the negative relationship between temperature and precipitation does not include winter where the relationship is warm and wet. The way it is stated in the text gives the sense that it is year round. It is my understanding that temperature/precipitation relationships are strongest in winter and summer and much more variable in spring, which is why the signal may be variant in your comparisons.

The manuscript stated clearly that we are only considering the extended spring season. Crhová and Holtanová show that the negative relation also holds in spring in the observations.

We tried to reduce any chance of misunderstandings in all sections.

Again, this is not helped by the choice of your study season, which is different from the tree ring season and makes the entire analysis hard to follow at times.

We are unsure what the reviewer refers to with this comment. The season we study is the season used in the publications of the tree ring-width-based reconstructions. It is always MAMJJ except when we discuss other reconstructions that were included after the suggestions in the previous round of reviews.

We tried to be more precise in the seasonal attribution of our writing to reduce the amount of misunderstandings in all sections.

Is there a mistake in stating that Rinne et al calibrated to Oxford? My memory is that it was Kew they used. I may be wrong but please double check.

Rinne et al., 2013, page 16: “Finally, for the final reconstruction of precipitation, total May--August precipitation values were obtained by scaling the Woburn chronology against the Radcliffe precipitation series. i.e. the mean and variance of the chronology were set equal to those of the instrumental series over the calibration period 1815--1893.”

Rinne et al., 2013, page 24: “The correlation coefficient between the reconstruction and the instrumental precipitation record from the Radcliffe station over the calibration and verification period exceeds the tentative minimum value required (r = 0.71) to provide acceptable results for palaeoclimatological studies (McCarroll and Pawellek, 2001). The precipitation series, which was formed by combining reconstructed values (1613--1893) and instrumental data (1894--2003), indicates significant decadal and centennial precipitation variability culminating in dry conditions in the early-middle 17th century and the late 20th century.”

A key objective of the paper is to show the value of the SPI over moving windows allows a better comparison of datasets. You need to spell out this contribution more, be explicit. I have not been convinced of its merits above other approaches from this paper. For instance, in the abstract you state that SPI is a valuable tool for bridging part of the problem in assessing agreement and disagreement – what are the problems it addressed.

Simulations, reconstructions, and observations refer to different spatial scales and show different skillful scales. Therefore, we do not expect agreement between them in mean or variations a priori, especially for variables with a very localized and complex nature like precipitation.

The SPI effectively leads to the comparison of climatologies over moving windows. It allows to filter out higher frequency variations and allows to compare moments and percentiles of climatologies. It allows addressing general tendencies towards dryness or wetness relative to the location.

We do not state that the SPI is outperforming other approaches in any aspect but that it extends and supplements the other approaches.

I think that this is an important part of the paper as mentioned above we know there is disagreement between reconstructions and observations – what new does the SPI analysis bring, what further insight?

We try to clarify the insights in the variations of the percentiles and of dryness and wetness, which are easily obtained by the SPI. That is, among other things, the SPI provides information on the consistency of percentiles, moments and how these change over time and, thus, allows a deeper, more informative coherent comparison of hydrologic variables between different data sets.

Again in the abstract, what do you mean by the most recent historical period. This is not defined.

We thank the reviewer for pointing this out. We clarified this.

The table on date sources used needs to include the years covered by each

This is now included.

Response to reviewer #2

Referee 2

General Remarks: This is an update to a previously reviewed manuscript. The writing and clarity are significantly improved, however, there are a number of changes that must be made before publication. Please see the specific comments below.

We thank the reviewer for their valuable comments and suggestions that were mostly implemented as suggested. For a detailed response please see our replies below and compare the document with the tracked changes.

The one more general comment that must be addressed is that the discussion and conclusion sections need to be more clearly written. In particular, there are parts of these sections that read like methodology and results. These parts should be replaced with clear and concise statements on what can and cannot be taken away from the results—what do the results tell us and why is that important?

The discussion and conclusion sections have been shortened. Parts considering data and methods have been moved to sections 2 and 3 and shortened. Discussion of results has been moved to section 4.5. Paragraphs containing introductory statements or summaries of the results have also been moved or removed. The conclusions are now listed in a hopefully clearer manner (compare section 6).

Page 1, Line 1: Suggest changing “The Standardized Precipitation Index is a valuable tool for bridging part of the problems in assessing agreement and disagreement between the different sources of information. We assess the agreement in the temporal evolution of percentiles of the precipitation distributions” to “The Standardized Precipitation Index is a valuable tool for assessing agreement between the different sources of information, as it allows for a comparison of the temporal evolution of percentiles of the precipitation distributions”.

Done

Page 1, Line 8: Suggest changing “consistent relations” to “consistency”.

Done

Page 1, Line 9: Suggest changing “purely” to “impact of”.

Done

Page 1, Line 9: I do not understand what is meant by the sentence that starts on this line.

We change the sentence as follows: The disagreement between sources of information reduces our confidence in inferences about the origins of hydroclimate variability for small regions.

Page 1, Line 22: What is meant by “supports”, can you be more specific here?

We add: It strengthens our confidence in inferences from the consistent sources of information.

Page 2, Line 3: Suggest removing “and inconsistency”.

Done

Page 2, Line 21: Suggest removing “complementing the current suite of statistical diagnostics for data-model comparisons”.

Done

Page 3, Line 25: Suggest removing “regional instrumental series”.

While we still think it is important to stress this difference to the previous work, we remove this.

Page 3, Line 28: Add “at” between “looking how”.

Done

Page 4, Line 4: I have never heard the term “document asset”. My understanding is that Climate of the Past uses the term supplementary material, although that may have changed. This note is just to make sure to use the correct term throughout.

To ensure stability of our supplements, we aim at providing them via https://osf.io/duyqe/ instead of uploading them to the Climate of the Past homepage. Compare https://publications.copernicus.org/for_authors/manuscript_preparation.html:

“Authors have the opportunity to submit supplementary material with their manuscript, such as additional figures and tables, highly detailed and specific technical information, user manuals, maps, or very large images. Supplements to articles should not be used as an archive for data. Therefore, data sets, movies, animations, or computer programme code should be deposited in FAIR-aligned data repositories and to insert a persistent identifier, ideally a DOI, in the manuscript.”

Therefore we chose the word asset. We have modified this now.

Page 4, Line 7: Change “most” to “mostly”.

Done

Page 5, Line 16: Suggest removing “long”.

Done

Page 5, Line 26: Suggest adding “records” after “temperature”.

Done

Page 6, Line 9: Suggest adding “observations” after “precipitation”.

Done

Page 6, Line 14: Suggest removing “in the past”.

Done

Page 6, Line 14: Suggest changing “relations” to “the relationship”.

Done

Page 7, Line 31: Suggest changing “resolved” to “resolution”.

Changed to: high-resolution modelling

Page 8, Line 10: Suggest changing “of the methods” to “associated with this method”.

Done

Page 8, Line 24: What was the rational for choosing 40 data points. If it was arbitrary it would be useful to briefly defend this choice.

We add: The choice of 40 data points is an ad hoc decision that lies between the recommendation by McKee et al. (1993) of 30 samples and our window length of 51 years.

Page 9, Line 2: Suggest changing “which” to “what”.

Done

Page 9, Line 8: Suggest adding “can” before “point out”, changing “that” to “if”, and removing “completely”.

Changed as suggested

Page 9, Line 14: Some background on what the Weibull standard deviation can tell us would be useful here.

To provide more information about the precipitation distributions, we choose to provide the square root of the second moment (the variance) to visualize our results. We add: this provides an additional clarification of how the precipitation distribution changes over time.

Page 9, Line 26: Change “Figures” to “figures”.

Done

Page 11, Line 4: Suggest adding “sections” after “following”.

Done

Page 12, Line 14: Why the choice of non-overlapping 11-year averages?

We add at the end of the paragraph: We choose 11-year non-overlapping windows to balance the number of available data points and the filtering of interannual variability.

Page 12, Line 16: Suggest changing “gives also” to “also gives”.

Done

Page 14, Line 2-3: Can you quantify this?

We add: e.g., interannual pairwise correlation coefficients are about 0.9 between the simulated East Anglia data and the other two regions, while the simulated England-Wales precipitation correlates at approximately 0.97 with the simulated Southern-Central England data. Absolute interannual precipitation differences between the three data sets are at a maximum approximately 151 mm/season.

Page 14, Line 13: Suggest changing “relation” to “relationship”.

Done

Page 14, Line 16: Suggest removing “though possibly not surprisingly”.

Done

Page 14, Line 31: Suggest changing the part of the sentence after “period of” to “overlap”.

Changed “the observational England-Wales time series” to “their overlap”

Page 14, Line 32: I think you mean “shorter timescale” here, “smaller scale” might be confusing since it could also refer to spatial scales.

We mean smaller amplitude and corrected this.

Page 15, Line 1: Suggest changing “lack the clear overall trend” to “do not have a long term trend”.

Done

Page 15, Line 7: Suggest changing the part of the sentence after “indicate” to “the opposite”.

Done

Page 15, Line 9: I would reference the most relevant figure for the “smoothed evolution” at the end of this sentence.

Done

Page 15, Line 13: Add “us” before “to”.

Done

Page 15 Line 14: Why define an acronym for standard deviation here? If you wish to do so, it makes sense to do it at first mention in the methods section. You also do not use “SD” in the manuscript text, only in the figure labels—perhaps put the acronym definition in the figure caption.

We add it to the figure caption but also keep the note that SD in the figure refers to the Weibull Standard Deviation.

Page 15, Line 15: The Southern-Central reconstruction appears to have an increasing trend.

The reviewer is right. We amend this sentence: The reconstruction for East Anglia does not show a clear evolution in the Weibull standard deviations, whereas there is an increasing trend in the Weibull standard deviations for the Southern-Central England data.

Page 16, Line 12: It is worth being more specific here, something like “…do not agree on the time evolution of precipitation percentiles”.

Done

Page 16, Line 12: There really does not seem to be much agreement. I think that this sentence confuses that point. Suggest reformatting to “Any agreement between the simulation and reconstructions appears random, with a tendency instead towards an opposite evolution—particularly for the dry percentile”. You can then go into a bit more detail and the explanation in the following paragraph.

We change this to: Simulations and reconstructions do not agree on the time evolution of precipitation percentiles (Figures 4 to 6). Any hint of an agreement between reconstructed and simulated data is likely due to randomness (compare Figure 4). There is instead a tendency towards opposite time evolutions between the data sources. This is best seen in the dry percentiles from the mid-18th to mid-20th century (Figure 5).

Page 17, Line 3: Suggest changing to "…would show a range of trajectories and thus these results do not preclude that the model is capturing…"

We change this to: Obviously, using an ensemble of regional simulations would show a range of trajectories. Therefore, these results do not preclude per-se that the model is capturing basic physical characteristics of precipitation variability in northwestern Europe.

Page 17, Line 7: Suggest changing to "…more or less in opposition (Figure 3)."

Done

Figure 7 caption: I was not able to understand what this caption means.

We change this (and the following) to: Visualisation of how percentile-values change for over windows. We show, which percentile the 93.3th percentile MAMJJ precipitation amount for a reference window represents over time for England-Wales (green solid lines), Southern-Central England (blue dashed lines), and East Anglia (black dash-dotted lines) in a) reconstructions and observations, and b) simulations. The reference window is centred in 1815 CE.

Page 18, Line 7: Why 1815?

As stated on page 9, line 17 and now repeated in section 4.3: The year 1815 CE is included in all data sets and it allows equivalent analyses of the PMIP3 past1000 simulations (e.g., Schmidt et al., 2011).

Page 19, Line 14: Suggest removing "prior to approximately the year 1850 CE".

Changed as suggested.

Page 20, Line 11: I am not sure that I understand why this is specifically relevant to a comparison of precipitation and temperature. It seems more relevant to the comparisons of, for instance, model simulations to reconstructions.

Actually, our aim was to use this link to compare the different data sets. The statistical link is most likely a reflection of a physical mechanism, which ideally should be represented in all data sets.

We modify the subsection title to: Relation between Temperature and Precipitation in different data sources.

We modify the whole paragraph to read: We only shortly explore the interrelation between the regional temperature and precipitation variability. We show how interannual correlations between the precipitation records and temperature series evolve over time. Figure 10 plots sliding interannual correlations for 51-year windows between the observed and reconstructed precipitation data and the Central England temperature as well as the correlation between simulated England-Wales precipitation and simulated Central England temperature. We plot correlations for the untransformed precipitation records. All records are for the MAMJJ-season. Obviously, the high amount of internal variability on local and regional scales complicates the comparison among different data sources when studying small regions.

Page 20, Line 12: Suggest adding “thus” after “We” and removing “the” before “precipitation”.

See reply to previous comment.

Page 22, Line 12: Add “the” before “environment”.

Done

Page 22, Line 17: Suggest adding “inherent to SPI” after “standardization”.

Added “inherent to the SPI”

Page 22, Line 21: Suggest removing the part of the sentence after “as well”.

Done

Page 23, Line 16: Add “the” between “in” and “form”.

Done. Also changed at two other instances.

Page 23, Line 17: Suggest putting a period after the Parker reference then starting a new sentence: “In addition, there are long instrumental station observations of precipitation…”.

Done

Page 23, Line 18: I would remove these two sentences as there is not much value in discussing what you do not analyze.

OK. These were in response to earlier referee or editor comments.

Page 24, Line 5: Suggest moving “ACRE” into the parentheses and the full written out version into the main part of the sentence.

Done

Page 24, Line 10: You have already used PDSI at this point.

Thanks.

Sections 5.1 and 5.2: Much of these sections is redundant. I think the manuscript would be fine with all of this removed.

The structure of the discussion has been changed (see new section 5). These subsections have been removed here but a large part of the information was added to sections 2 and 3.

We do not completely remove these parts as they were included as response to concerns from the previous review-iteration.

Please see tracked changes in the discussions section, and earlier sections.

Page 27, Line 1: Suggest removing the period and reformatting as ", however, there is occasional temporal consistency between the simulations and observation based data."

This sentence and a good part of this paragraph was removed in response to general requests for shortening of the manuscript.

Page 27, Line 6 to 5.3.3: This is just more results and should either be moved to the results section or removed entirely.

We moved it to the results section.

Page 28, Line 3: Again this paragraph is largely results.

See previous response.

Page 28, Line 16: But it is really the reconstructions that look inconsistent with both the observations and simulations and thus I would focus predominantly on that in this paragraph (at the moment only the final sentence discusses this disagreement).

The paragraph changed to: The differences between simulation and observations may imply either shortcomings of any of the observational data sets in the early period or that the simulation presents a too stable relationship. Explanations might be physical inconsistencies within the simulations. More generally, any of the data sources may lack the physical relation between the temperature and precipitation records in the chosen season. Another possibility is that internal large-scale climate factors influencing the relation between both parameters evolve differently in simulation and reality. Assuming that the observations are the more reliable data set, we tend to the inference that the disagreement between observations and reconstructions suggest major shortcomings in the reconstructions set.

Page 28, Line 14: I would reformat this as “If we expect temporal consistency among the different sources of information, then we are assuming that all the sources of information are responding to the impact of external climate forcing”.

Done

Page 28, Line 27: I would revisit this sentence with any eye towards clarity.

We rephrase: We have to ask, what is our expectation of consistency between simulated and observed responses to exogenous influences?

Page 28, Line 31: Change “frequent” to “frequency”.

Done

Page 28, Line 34: Suggest changing to “However, the analysis period includes relatively large…”

Done

Page 29, Line 12: Does this not require that the opposite behavior is “significant” relative to randomness? This statement seems too speculative. Likewise, the observations and reconstructions often do not agree…

We remove the two sentences.

Page 30, Line 1: Change “frequent” to “frequency”.

Done

Section 5.3.5: This discussion feels a bit too general and should be refocused towards discussing the results and implications of this study

We like to keep this but shorten it. Please refer to the new section 5.3 for details, e.g., in the tracked changes below.

List of relevant changes:

Abstract:

- Minor clarifications

Introduction:

- Minor clarifications
- Minor restructuring
- Overall possibly slightly shortened

Data:

- Added more information to Table 1
- Extended discussions on the individual sources of information
- Major clarifications and restructuring

Methods:

- Added more information
- Extended discussions of the method
- Major restructuring and clarifications

Results:

- Major clarifications
- Added additional analyses that were included in the Discussion section in the previous version

Discussions:

- Removed multiple subsections (content partly moved to other sections)
- Shortened remaining subsections
- Clarifications of the discussions

Conclusions:

- Clarifications and restructuring
- Highlighted main findings

Some references became obsolete.

Inconsistencies between observed, reconstructed, and simulated precipitation indices for England since the year 1650 CE

Oliver Bothe[1], Sebastian Wagner[1], and Eduardo Zorita[1]

[1]Helmholtz Zentrum Geesthacht, Institute of Coastal Research, 21502 Geesthacht, Germany

Correspondence: Oliver Bothe (ol.bothe@gmail.com)

Abstract. The scarcity of long instrumental records, uncertainty in reconstructions, and insufficient skill in model simulations hamper assessing how regional precipitation changed over past centuries. Here, we use standardised precipitation data to compare a regional climate simulation, reconstructions, and long observational records of seasonal (March to July) mean precipitation in England and Wales over the past 350 years. The Standardized Precipitation Index is a valuable tool for ~~bridging part of the problems in assessing agreement and disagreement~~ assessing agreement between the different sources of information~~. We assess the agreement in~~ , as it allows for a comparison of the temporal evolution of percentiles of the precipitation distributions. These evolutions are not consistent among reconstructions, a regional simulation, and instrumental observations for severe and extreme dry and wet conditions. The lack of ~~consistent relations~~ consistency between the different data sets may be due to the dominance of internal climate variability over the ~~purely~~ impact of natural exogenous forcing conditions on multi-decadal time-scales. ~~This, in turn, questions our ability to make dynamical inferences about~~ The disagreement between sources of information reduces our confidence in inferences about the origins of hydroclimate variability for small regions. However, it is encouraging that there is still some agreement between a regional simulation and observational indices. Our results emphasize the complexity of hydroclimate changes during the ~~most recent historical period~~ recent centuries and stress the necessity of a thorough understanding of the processes affecting forced and unforced precipitation variability.

1 Introduction

Confidence in future climate projections of, e.g., drought and wetness conditions requires understanding of past climate and hydroclimate variability and its drivers (e.g. Schmidt et al., 2014). Focussing on the hydroclimate, estimates of past and future changes are still highly uncertain for precipitation at regional scales. Indeed, our understanding of internal, naturally forced, and anthropogenically forced variability is weaker for precipitation than for temperature due to the more complex controls on precipitation variability (e.g. Zhang et al., 2007; Hoerling et al., 2009; Iles et al., 2013; Fischer et al., 2014) and the more local-scale nature of precipitation-processes.

Consistency among estimates from early instrumental observations, paleo-reconstructions from environmental archives (i.e., paleo-observations), and climate simulations supports our understanding of past changes. It strengthens our confidence in inferrences from the consistent sources of information. Here, consistency among estimates simply means that various sources of information do not contradict each other. ~~Consistency is a weaker requirement than to expect consilience, i.e., it is weaker~~

than requiring that the evidence from the different data sets converges. Despite being a more rather liberal metric, consistency is an appropriate measure in view of the multiple sources of uncertainty in inferring past hydroclimate and precipitation variability.

Here, we explore consistency and inconsistency of observations, reconstructions, and simulations for one small region and focusing only on precipitation changes. Specifically we set out to study the consistency in the statistical properties of precipitation distributions in these sources of information.

Comparing precipitation among different data sources poses various challenges. Problems relate not only to pronounced biases in the simulated precipitation, especially derived from raw global models, and to differences in representation or, in the case of data fields, the grid resolution. In the context of long observational time series, data inhomogeneities due to changes in instrumentation, measuring techniques, and changes in locations can further influence estimates of longer-term trends (Frank et al., 2007; Wilson et al., 2005; Böhm et al., 2010) (e.g., Frank et al., 2007; Wilson et al., 2005; Böhm et al., 2010; Burt and Howden, 2011; Craddock and Craddock, 1977). Reconstructions likely represent only part of the variability spectrum of, e.g., precipitation, dependent on the strength of the climatic signal in the original data and on further shortcomings of the underlying paleo-observations (compare discussions by PAGES Hydro2k Consortium, 2017). .

The PAGES Hydro2k Consortium (2017) discuss in more detail the problems in comparing hydroclimatic variables between reconstructions and simulations.

The PAGES Hydro2k Consortium (2017) They developed recommendations for the comparison of hydroclimate representations in simulations and paleo-observations, emphasizing the uncertainties of estimates from both sources. They stress the complementary nature of simulated and environmental information considering their respective uncertainties. Estimates have to represent the same parameters on related spatial and temporal scales. Only then, a comparison can be valid. We need appropriate methods to bridge the gap between the local or regional reconstruction and the simulation output that represents aggregates over larger spatial scales. Proxy system models are one means to achieve this (Evans et al., 2013; PAGES Hydro2k Consortium, 2017). We argue that the standardisation of precipitation estimates is a simple means to compare the statistical properties of hydroclimatic parameters in simulations and paleo-observations complementing the current suite of statistical diagnostics for model-data comparisons. It is of value for periods with and without comprehensive sets of climate and weather observations.

Transforming precipitation estimates to the Standardized Precipitation Index (SPI; McKee et al., 1993) facilitates the comparison of different sources of information on precipitation in view of the mentioned challenges. It provides a common basis for comparisons between different locations, periods or seasons. The core of the SPI calculation is the fit of a distribution function to the precipitation estimates. We argue that the transformation of precipitation estimates to the SPI is a simple means to compare the statistical properties of hydroclimatic parameters in simulations and paleo-observations. It is of value for periods with and without comprehensive sets of climate and weather observations.

Previous usage of the SPI in paleoclimatology generally focussed on the index series (compare, e.g., Domínguez-Castro et al., 2008; Seftigen et al., 2013) and did not consider further information available through the transformation. We apply the SPI over moving windows of 51 years to study variations in the properties of precipitation distributions on multi-decadal time-scales. We concentrate on a regional domain where all sources of data, i.e., observations, reconstructions, and simulations are available. By applying the SPI-transformation over moving windows, we are able to evaluate and compare percentiles of the estimates as well as the moments of the distributions and the temporal changes of these distributional properties. We are essentially comparing sequences of climatologies.

Long observationally based records allow us to assess how the statistics of observed precipitation have changed over the last couple of centuries. They, in turn, provide the basis for evaluating how state-of-the art regional or global climate model simulations and reconstructions for the Common Era (CE) compare in domains co-located with the available observations. We choose southern Great Britain as our domain of interest since there are precipitation observations available in the form of the England-Wales precipitation data set (Alexander and Jones, 2000, for the period 1766 CE to present), its subdivisions, and instrumental records for Oxford (cf. Radcliffe), Pode Hole and Kew Gardens. The instrumental records start in 1767 CE, 1726 CE, and 1697 CE, respectively.

A number of precipitation reconstructions are available for the South of Great Britain. We choose the millennium-long tree-ring based data by Cooper et al. (2013) and Wilson et al. (2013) for East Anglia and Southern-Central England, respectively. We focus on an extended spring season (MAMJJ). The next section discusses our decision to concentrate on this data instead of the $\delta^{18}O$ ~~-based~~ based scaling approaches by Young et al. (2015, covering the period 1766 to present) and Rinne et al. (2013, reconstructed values from 1613 to 1893 CE).

Regional simulations for the last 500 to 2000 years are rare. ~~Among studies on these , **?** describe a regional~~ To our knowledge there are only two transient regional simulations. Gómez-Navarro et al. (2015) compare one of these simulations, a simulation with the model MM5~~for Europe. Gómez-Navarro et al. (2015) compare this simulation ,~~ , to reconstructions for various parameters over larger regional domains within Europe. For precipitation, they compare the simulation to the gridded precipitation reconstructions of Pauling et al. (2006) for Western Europe, which is based on a set of dendroclimatological and other natural proxies and documentary information. Gómez-Navarro et al. (2015) find rather good agreement in the evolution of median precipitation amounts between the reconstruction and their regional simulation for a domain including the British Isles and Ireland for the summer season. The agreement is much weaker for the spring season. They also emphasize model shortcomings and the lack of agreement in the representations of extreme climate anomalies. On the side of the reconstructions, Gómez-Navarro et al. (2015) stress the inconsistencies among the reconstructions of different parameters (i.e., temperature, precipitation, and sea level pressure).

Here, we compare observations and paleo-observations with each other. We additionally compare them to output from a regional simulation with the model CCLM for the European domain over the period 1645 to 1999 CE (compare Gómez-Navarro et al., 2014; Bierstedt et al., 2016). Our comparison differs from Gómez-Navarro et al. (2015) by using a different regional model, focussing on a smaller region, and by using regional time series reconstructions instead of deriving records from gridded products. Moreover, our general focus is on precipitation~~, including regional instrumental series~~.

Our focus is not least to motivate the use of the Standardized Precipitation Index in hydroclimatic comparisons between different data sets in paleoclimatology. We use the SPI to study the consistency of the different sources of precipitation infor-

Table 1. List of data sets by region, parameter, type of data, period covered, season used, and source for obtaining the data.

Location/Region	Parameter	Type	Period CE	Season	Source
England-Wales	Precipitation	Observations	1766–2018	MAMJJ	https://www.metoffice.gov.uk/hadobs/hadukp/
South-West England	Precipitation	Observations	1873–2018	MAMJJ	https://www.metoffice.gov.uk/hadobs/hadukp/
South-East England	Precipitation	Observations	1873–2018	MAMJJ	https://www.metoffice.gov.uk/hadobs/hadukp/
Central England	Precipitation	Observations	1873–2018	MAMJJ	https://www.metoffice.gov.uk/hadobs/hadukp/
East-Anglia	Precipitation	Reconstruction	900–2009	MAMJJ	https://www.ncdc.noaa.gov/paleo-search/study/12
Southern-Central England	Precipitation	Reconstruction	950–2009	MAMJJ	https://www.ncdc.noaa.gov/paleo-search/study/12
Southern England	Precipitation	Reconstruction	1613–1893	MJJA	Correspondence with original author
United Kingdom	$\delta^{18}O$	Observations	1766–2012	JJA	https://doi.org/10.1007/s00382-015-2559-4
Central England	Temperature	Observations	1659–2018	MAMJJ	https://www.metoffice.gov.uk/hadobs/hadcet/
Kew Gardens	Precipitation	Instrumental	1697–1999	MAMJJ	https://climexp.knmi.nl/
Pode Hole	Precipitation	Instrumental	1726–1994	MAMJJ	https://climexp.knmi.nl/
Oxford	Precipitation	Instrumental	1767–1996	MAMJJ	https://climexp.knmi.nl/
Europe	Precipitation	~~CCLM~~ Regional climate model simulation	1645–1999	MAMJJ	http://doi.org/10.6084/m9.figshare.5952025
Europe	Temperature	~~CCLM~~ Regional climate model simulation	1645–1999	MAMJJ	http://doi.org/10.6084/m9.figshare.5952025

mation for approximately the last 350 years. That is, we are looking at how well the sources of information compare among each other. This is a limited aim, which is appropriate considering the various uncertainties especially in simulations, and reconstructions, but also in observations. We explicitly do not expect the simulation output to agree with the instrumental and paleo-observation data on the mean precipitation amount since spatial representations differ. We also do not expect them necessarily to agree on decadal variations in precipitation because of the presence of internal variability (compare, e.g., Deser et al., 2012b, a; Swart et al., 2015) potentially masking commonly forced external signals. ~~Thus, even~~ Even a large ensemble of simulations may not necessarily represent these variations (see, e.g., Annan and Hargreaves, 2011). Since we transform precipitation to the Standardized Precipitation Index over moving windows, our analyses essentially become comparisons between series of climatologies, thus potentially filtering shorter term internal variability.

In the following, we first introduce and discuss our choices on data sets and methodology before comparing the data sets and discussing the results. A ~~document asset supplements this manuscript but provides only~~ supplementary document to this manuscript provides additional analyses that are non-essential for our conclusions.

2 ~~Data~~Data

Hydroclimatic changes affect humans and the environment ~~most~~ mostly on the local and regional scale. Therefore, we focus on small domains and use precipitation data. Precipitation is a more tangible variable than, e.g., drought indicators like the Palmer Drought Severity Index (PDSI). ~~We only use the single time series records instead of gridded products to avoid the possibly spurious non-climatic variance and other stastical artifacts potentially introduced by reconstruction techniques.~~

We aim at describing how much agreement we can find between different sources of information for precipitation in a small domain over a period with limited instrumental data, i.e., a period when we have to rely on reconstructions from paleo-observations. Such an assessment helps to increase our confidence in the estimates by the different sources of information. In turn, it also increases our understanding of past hydroclimatic variability.

We use observationally derived data sets, reconstructions, and simulation output in our main analyses. ~~Additionally we~~ We use further observationally derived records and instrumental station observations for assessing the quality of our main data sets. Table 1 lists the sources of information. For all analyses, we use primarily the spring-summer season from March to July (MAMJJ).

~~Data availability motivates the choice of the regional domain~~Starting from the available regional climate simulation (see below), we choose the region for our study based on the availability of precipitation observation and reconstruction data. There are long records of instrumental measurements of climate parameters for a number of locations in Europe. For southern Great Britain, there exist observational regional domain composite records for temperature and precipitation, precipitation reconstructions, and long instrumental records.

2.1 Observations

~~We choose the South of Great Britain as our domain of interest since there are precipitation observations available in~~ The British Isles are unique because there exist long observation based indices for precipitation and temperature in the form of the England-Wales precipitation data ~~set (Alexander and Jones, 2000), its subdivisions, and instrumental records for Oxford (cf. Radcliffe), Pode Hole and Kew Gardens. Furthermore, there is also a long observational temperature record available for additional comparison,~~ (Alexander and Jones, 2000) and the Central England Temperature ~~series (Parker et al., 1992). Croxton et al. (2006) find that the England-Wales precipitation and the Central England Temperature well represent the climate of the South of Great Britainin the late 20th century~~data (Parker et al., 1992). In addition, there are long instrumental station precipitation observations available, e.g. in southern Great Britain, for Kew Gardens, Oxford, and Pode Hole.

Alexander and Jones (2000; see also Wigley et al., 1984) describe the England-Wales precipitation (EWP) data. It is available from the Met Office Hadley Centre at monthly resolution extending back to the year 1766. The Met Office Hadley Centre also provides subdivisions of the data. We use those for South-West, South-East, and Central England. ~~We concentrate on the England-Wales domain because there is also temperature data available in form of the long Central England Temperature series (Parker et al., 1992).~~ Alexander and Jones (2000) describe the automated method of updating long precipitation series like the data by Wigley et al. (1984) while also ensuring the homogeneity of the data. Parker et al. (1992) similarly describe the production of temperature data to complement long-running series while maintaining the Central England Temperature data and how to maintain quality-control and homogeneity.

The Central England and England-Wales observation indices are good representations of the late 20th century climate of the South of Great Britain according to Croxton et al. (2006). Note that the composite series naturally rely on the instrumental series.

The Climate Explorer (http://climexp.knmi.nl/) provides access to a number of long series of monthly instrumental precipitation observations from the Global Historical Climatology Network (Peterson and Vose, 1997). We use those from Oxford, Kew Gardens, and Pode Hole in addition to the observationally derived Met Office Hadley Centre data sets. The Climate Explorer provides monthly data for these locations from 1697 to 1999, 1726 to 1994, and 1767 to 1999 CE, respectively. The later years in the Oxford record include missing values and we therefore only use data from 1767 to 1996 CE.

Frank et al. (2007) noted the uncertainties in early instrumental temperature observations. Additionally, the very early data in the Central England Temperature data includes non-instrumental indirect data to infer past temperature. Similarly, early precipitation observations require rigorous quality control (e.g., Burt and Howden, 2011). Woodley (1996) reviews the history of precipitation data for England and Wales as well as Scotland.

2.2 Reconstructions

There are a number of gridded reconstructions of hydroclimatic parameters covering the European domain. Continental domain gridded precipitation reconstructions are, e.g., Pauling et al. (2006), Casty et al. (2007), and Franke et al. (2017). Reconstructions of drought indices like the PDSI exist as gridded products, too, for various regions of the world including Europe (The Old World Drought Atlas, Cook et al., 2015). These products allow assessing the quality of the hydroclimate in paleoclimate simulations (Smerdon et al., 2015).

We decide to use regional precipitation reconstructions for our domain instead of gridded products to minimise the effect of the reconstruction method on the results. We focus on precipitation as it allows the direct comparison with long instrumental records and it is a parameter directly experienced by people.

To our knowledge, there are three precipitation reconstructions for small domains from the South of Great Britain, i.e., approximately within the domain of the England-Wales precipitation and the Central England temperature. These are for East Anglia (Cooper et al., 2013), for Southern-Central England (Wilson et al., 2013), and the reconstruction for Southern England by Rinne et al. (2013). The former two use tree-ring width data for their reconstructions, the latter uses tree-ring oxygen isotopes. There is additionally the work by Young et al. (2015), who scale a $\delta^{18}O$ composite record from Great Britain to the England-Wales precipitation.

We decide only to use the two tree-ring width based records. The main reason for excluding the Rinne et al. record is that it concatenates instrumental data from Radcliffe (cf. Oxford) station for 1894 to 2003 to the reconstructed values from 1613 until 1893. This reduces the time of overlap with the England-Wales precipitation data. The reconstruction by Rinne et al. (2013) is not publicly available, but the lead author provided us with the data. We provide a short assessment of the data in a supplementary manuscript asset.

~~Similarly, Young et al. (2015) scale their input $\delta^{18}O$ records by precipitation and provide the input series as supplement to their paper. Our supplementary manuscript asset provides a short assessment of a scaling using this data.~~

In the main manuscript, we only use the data by Cooper et al. (2013) and Wilson et al. (2013) for, respectively, East Anglia and Southern-Central England in March, April, May, June, July (MAMJJ). Cooper et al. (2013) and Wilson et al. (2013) identified this extended spring as the season their tree-ring width records are sensitive to for their reconstructions of precipitation. ~~In the following, we compare the England-Wales precipitation with the two reconstructions for the South of Britain.~~

~~Wilson et al. (2013) and Cooper et al. (2013) already~~ These authors calibrate their tree-ring data against gridded precipitation beyond their target regions of Southern-Central England and East Anglia, respectively. Thereby the reconstructions are possibly biased beyond their respective regions of interest. They compare their reconstructions against the long instrumental records and find a lack of stability of the relation to the instrumental data. They discuss the limitations of their ~~respective reconstructions. Both reconstructions represent between 30% and 35% of regional interannual~~ reconstructions representing less than 40% of the regional precipitation variance over the 20th century. Obviously, the reconstructions suffer from the limited lengths of the available tree ring samples. This ~~has an effect on how much low frequency variability the reconstructions can resolve.~~ may limit the resolution of precipitation variability at low frequencies in the reconstructions.

Although the reconstructions show a notable amount of low frequency variability, Cooper et al. (2013) cautions against too much confidence in the reconstructed low frequency precipitation variability. Cooper et al. (2013) explicitly call their work "preliminiary" with respect to reconstructing low frequency precipitation variability. Wilson et al. (2013) and Cooper et al. (2013) emphasize the weaknesses of their reconstructions in representing extreme years. On the other hand, both are confident in the mid- to high-frequencies of their reconstructions.

The authors note variable relationships between tree growth and environmental controls for their regions in the past. Indeed there are periods when ~~relations~~ the relationships between trees and precipitation are not significant. ~~Both studies are confident in the mid- to high-frequencies of their reconstructions but emphasize that their reconstructions have weaknesses in representing extreme years when compared to the observations. Cooper et al. (2013) explicitly call their paper "preliminiary" with respect to reconstructing low frequency precipitation variability~~Wilson et al. (2013) and Cooper et al. (2013) discuss the possibility that the tree-species used for their reconstructions were less sensitive to precipitation over certain periods, e.g., the early 19th century. That is, the proxies, theoretically representing a precipitation signal, also contain a temperature signal, for instance, if they are sensitive to soil moisture. Wilson et al. further suggest an effect of the Industrial Revolution and the associated pollution on the trees in their selection. Wilson et al. (2013) also discuss the reliability of the instrumental data but conclude this is likely not an issue.

The works by Rinne et al. (2013) and Young et al. (2015) use their $\delta^{18}O$ data to reconstruct precipitation for Southern England and Great Britain respectively. We shortly discuss results for both reconstructions below and give some more details in the supplementary document.

Rinne et al. (2013) calibrate and scale their local isotope data from 1613 to 1893 CE against the station observations from Oxford for the period 1815 to 1893 CE and concatenate the reconstruction with the observations for 1894 to 2003 CE. They target an extended summer season from May to August.

Young et al. (2015) ~~find that the two reconstructions from tree-ring widths strongly differ from their own~~ use the England-Wales summer, June to August, precipitation as scaling target for a composite of eight isotope records from Scotland, Wales, and England for the period 1766 to 2012 CE. They provide the input series as supplement to their paper.

Both publications by Rinne et al. (2013) and Young et al. (2015) note the differences of their scaled $\delta^{18}O$ data ~~. The~~ to the tree-ring width based works by Wilson et al. (2013) and Cooper et al. (2013). Young et al. (2015) emphasize that the extended spring reconstructions are basically unrelated to the $\delta^{18}O$ data. ~~Young et al. (2015), therefore, question whether both approaches reliably represent precipitation in the South of Great Britain. After discussing possible reasons for the disagreement, Young et al. (2015) conclude that the reconstructions~~ Young et al. conclude that these differences make it unlikely that the tree-ring based works and their $\delta^{18}O$ based work represent the same environmental parameter. They highlight the lack of a calibration against regional precipitation data. Young et al. (2015) conclude that their own data reliably reflects precipitation while the tree-ring widths most likely represent the combination of various environmental influences on tree growth instead of a single climate parameter.

Despite the conclusions of Young et al. (2015) we decide to focus in the main manuscript on the two tree-ring width based records by Cooper et al. (2013) and Wilson et al. (2013)~~are valid representations of oak growth in England, but they are not reliable representations of regional precipitation variations in contrast to~~. The main reason for excluding the Rinne et al. record is that it concatenates instrumental data from Radcliffe (cf. Oxford) station for 1894 to 2003 to the reconstructed values from 1613 until 1893. This reduces the time of overlap with the England-Wales precipitation data.

We do not focus on the data by Young et al. (2015) for two reasons. Firstly, the authors do not provide the full reconstruction, and, secondly, the data starts at the earliest in 1766 CE, which again minimises the period available for comparing it to the simulation data. However, this is exactly the time period of the ~~$\delta^{18}O$ data of Young et al. (2015)~~England-Wales precipitation series.

We think the focus on the tree-ring width based reconstructions is appropriate to present the possibilities of using the SPI and to highlight potential consistencies and inconsistencies between the different data sources. In the following, we compare the two reconstructions for the South of Britain with the England-Wales precipitation observations.

2.3 Simulations

We compare the observations and the reconstructions to output from a regional simulation with the model CCLM for the European domain over the period 1645 to 1999 as also used by Gómez-Navarro et al. (2014) ~~. Forcing for the regional simulation is from a global simulation with the MPI-ESM climate model in its COSMOS set-up (see below)~~and Bierstedt et al. (2016). We use output from 1652 onwards (Gómez-Navarro et al., 2014). This simulation is one of only two transient regional simulation for this region and the last fast few centuries, which exist to our knowledge.

~~The lateral forcing of~~ Forcing for the regional simulation is ~~output from the Millennium-simulation COSMOS-setup of~~ from a global simulation with the Max-Planck-Institute Earth System Model (MPI-ESM) in its Millennium-simulation COSMOS-setup. For details, see Jungclaus et al. (2010). This version of MPI-ESM couples the atmosphere model ECHAM5, the ocean model MPI-OM, a land-surface module including vegetation (JSBACH), a module for ocean biogeochemistry (HAMOCC), and an interactive carbon cycle. For the simulation, ECHAM5 was run with a T31 horizontal resolution and with 19 vertical levels. MPI-OM used a variable resolution between 22 and 250 km on a conformal grid for this simulation. The ensemble used diverse forcings. The driving simulation for the regional simulation with CCLM is one MPI-ESM simulation with all external forcings and a reconstruction of the solar activity based on Bard et al. (2000), i.e. with a comparatively large amplitude of solar variability.

The regional climate model CCLM simulation (Wagner, personal communication; see also Gómez-Navarro et al., 2014; Bierstedt et al., 2016) uses adjusted forcing fields relevant for paleoclimate simulations as also used with the global MPI-ESM simulation. These include orbital forcing and solar and volcanic activity. Since the regional model does not represent the stratosphere, the regional simulation considers the effect of volcanic aerosols as a reduction in solar constant equivalent to the net solar shortwave radiation at the top of the troposphere in MPI-ESM. CO_2 variability is prescribed and changes in greenhouse gases CO_2, CH_4, and N_2O are based on data by Flückiger et al. (2002). Land-cover changes are included as external lower boundary forcing using the same data set as the MPI-ESM simulation (Pongratz et al., 2008). The presented CCLM simulation uses a rotated grid with a horizontal resolution of 0.44 by 0.44 degree and 32 vertical levels. The sponge zone of seven grid points at each domain border is removed and fields are interpolated onto a regular horizontal grid of 0.5 by 0.5 degree.

We choose the domain including grid points closest to the longitudinal and latitudinal borders 5.5W to 1.5E and 50.5 to 54.5N to represent the England and Wales precipitation domain. This selection is somewhat arbitrary but we assume it sufficiently represents the England-Wales precipitation domain to allow meaningful comparison of changes in percentiles, although not in absolute percentile values. We choose the domain 5 to 0W and 50 to 55N as simulated counterpart of the Central England Temperature. The simulated East Anglia series represents the domain 0E to 2E and 52N to 53N, and we choose the domain 2.5W to 0E and 51N to 52.5N as equivalent for Southern-Central England. All analyses are for ~~an~~ the extended spring season~~from March to July~~ , MAMJJ, since this is the seasonal focus of the reconstructions. The appendix provides a short evaluation of the simulation against the observational CRU-data (Harris et al., 2014) over the European domain. We do not apply any bias correction to the simulation output.

So far, global simulations for the last millennium have notably coarser resolutions than the 0.44 by 0.44 degree of the regional simulation we use here (compare, e.g., Fernández-Donado et al., 2013; PAGES 2k-PMIP3 Group, 2015). However, ~~compared to other regional simulationsthis is only a~~ in contrast to present-day and future scenario regional simulations, a 0.44 by 0.44 degree resolution represents a comparatively coarse resolution dynamical downscaling. As a review by Ludwig et al. (2018, including two of the present authors) highlights, this is because the demand for long simulation periods limits applications of regional models in paleoclimatology to relatively coarse setups. Thus, one may question the benefits of the approach compared to more recent higher-resolution global simulations, e.g., with the global models CCSM4 and CESM1 (Landrum et al., 2012; Lehner et al., 2015), which have resolutions of $0.9^{\circ} \times 1.25^{\circ}$.

~~A review by Ludwig et al. (2018, including two of the present authors) emphasizes that the demand for long simulation periods limits applications of regional models in paleoclimatology to relatively coarse~~ Sørland et al. (2018) discuss the benefits of regional climate simulations in studies on regional climates. Besides other models, they also use CCLM in a 50km ~~setups.~~

Ludwig et al. (2018) conclude that regional simulations provide setup comparable to the simulation used here. They note that improved representation of regional climate in a regional simulation is not solely due to the increased resolution but may be due to different strategies in model-building and tuning. Pinto et al. (2018) explain differences in results from regional, including CCLM, and global simulations for southern Africa by an interplay between the representation of sub-grid-scale processes in the different models and factors related to the increased resolution.

Blenkinsop and Fowler (2007) find that regional climate models may be deficient in their ability to model persistent low precipitation episodes for the British Isles, which has repercussions for their representation of drought events. The review by Ludwig et al. (2018) reports more realistic distributions for precipitation in the paleo-context. regional paleoclimate simulations.

Flato et al. (2013, chapter 9 of the IPCC AR5) are more ambiguous in their review but they emphasize the value review the progress of regional downscaling as a tool in addition to higher resolved global simulationsand high-resolution modelling. They emphasize that the skill of such exercises depends on the model used, the season, the domain of interest, and the considered meteorological variable. They highlight studies showing that there is not a linear increase in simulation skill towards higher resolutions. Higher resolutions typically provide more reliable estimates of extremes, including heavy rainfall.

The quality of the simulated precipitation still strongly depends on the parameterisations implemented in the regional climate model. Precipitation, especially convective precipitation events, are still sub-grid processes, even within regional climate models. Concentrating on accumulated amounts on seasonal time-scales and their long-term changes, however, allows at least a more robust comparison of simulated precipitation to observed and reconstructed data.

3 Methods

One objective of this manuscript is to highlight how the concept of the Standardised Precipitation Index (SPI, McKee et al., 1993) adds additional perspectives on comparing various sources of information for periods with and without instrumental observations. Therefore, we shortly introduce the SPI-transformation procedure and how we use this information to subsequently compare precipitation estimates from observations, reconstructions, and a regional climate simulation.

3.1 The Standardized Precipitation Index – SPI

Standardising precipitation data facilitates comparing precipitation distributions between different locations, time-scales, periods, and data sources. For this purpose, McKee et al. (1993) introduced the Standardized Precipitation Index (SPI).

The Interregional Workshop on Indices and Early Warning Systems for Drought proposed the SPI as common index to facilitate comparability between meteorological drought estimates for different regions (Hayes et al., 2011, see also Keyantash et al. (2002)). The SPI should complement previously used indices.

Raible et al. (2017) find the SPI to be a reliable drought index for Western Europe including the British Isles. The standardisation inherent to the SPI allows further applications, e.g., flood monitoring (Seiler et al., 2002), and the easy comparison of normal, dry, and wet conditions between different sources of data. Indeed the UK drought portal (https://eip.ceh.ac.uk/droughts) relies on the SPI, and there are recommendations to use the SPI in operational monitoring of meteorological drought (e.g. Hayes et al., 2011).Sienz biases of the methods. Sienz et al. (2012) discuss associated biases of the approach.

The SPI uses only precipitation, which makes it an ideal and relatively straightforward tool for comparing hydroclimatic data between different data sources. Precipitation is a standard output of simulations, long instrumental records exist for various locations, and a number of reconstructions exist as well.

However, as the SPI uses only precipitation, it is of less value when the interest is in, e.g., the water supply, runoff, or streamflow (but see Seiler et al., 2002). The focus on precipitation also limits the applicability for studying temperature sensitive parts of the hydrological cycle and impacts on biological and anthropogenic systems (e.g., PAGES Hydro2k Consortium, 2017; Keyantash et al., 2002; Hayes et al., 2011; Van Loon, 2015).

Previous usage of the SPI in paleoclimatology generally focussed on the index seriesand did not consider further information available through the transformation from precipitation to SPI. For example, Domínguez-Castro et al. (2008) and Machado et al. (2011) compare SPI-series to differently derived hydroclimatic indices over approximately the last 500 years. Other studies reconstructed the SPI instead of absolute precipitation amounts (e.g. Seftigen et al., 2013; Yadav et al., 2015; Tejedor et al., 2016; Klippel et al., 2018). Lehner et al. (2012) use the SPI to compute pseudo-proxies from re-analysis data and long simulations with global climate models to test a reconstruction-method.

3.1.1 Transformation

The standardized precipitation index SPI requires fitting a distribution function to the precipitation data. There are various candidate distributions as, e.g., Sienz et al. (2012, and their references) discuss (see also Stagge et al., 2015)(e.g., Sienz et al., 2012; Stagge et al., 2015, and their references).

In our analyses, we fit a Weibull distribution. Sienz et al. (2012) highlighted that the Weibull distribution performed better in transforming the England-Wales precipitation data on a monthly time-scale compared to a number of other distributions. However, other distributions outperformed the Weibull distribution for other data sets and other SPI time-scales. Results differ only little if we fit Gamma or Generalised Gamma distributions (not shown). Our procedure of the SPI-calculation follows the detailed description by Sienz et al. (2012).

McKee et al. (1993) recommend at least 30 data points for successful distribution fits, but Guttman (1994) notes the lack of stability for small sample sizesand shows that higher order L-moments only converge for samples larger than about 60 data points.

. We fit distributions over moving sliding 51-year windowsand . Thus, we use more data points than recommended by McKee et al. (1993) but still less than the 60 points for which Guttman (1994) finds convergence of higher order L-moments. Appendix Figure B1 shows 95% intervals of a bootstrap procedure samples sampling 1000 times 40 data points from each window to provide an estimate of sampling variability (presented in Appendix Figure B1). Our procedure of the SPI-calculation follows the detailed description by Sienz et al. (2012)and fitting distributions to these samples. The choice of 40 data points is an ad hoc decision. We could also have chosen sample sizes of 49 data points.

3.1.2 Evaluation

Standardising precipitation data ~~can avoid or at least~~ at least can attenuate some of the problems mentioned in the introduction. Transforming precipitation to standardised values provides further means to study the agreement or the lack thereof between different data sources.

By transforming to ~~Standardized Precipitation Indices~~ the SPI over moving windows, we essentially compare climatologies and potentially filter shorter term internal variability. If this climatology for the observations is the target climatology, an ensemble of climate simulations should sample this distribution interannually following the paradigm of a statistically indistiguishable ensemble (Annan and Hargreaves, 2011). Our analyses compare how well the climatologies agree in different sources of data.

One particular interest is to consider to ~~which~~ what extent the different data sources describe comparable evolutions in various percentiles, e.g., representing extremes. The SPI-transformation simplifies this. If the transformation over moving windows filters a certain amount of internal variability, if boundary and forcing conditions are sufficiently equivalent in the simulation compared to the observed climate, and if simulated precipitation and the observed climate react equivalently to these conditions, precipitation distributions and their properties may change consistently between different sources of information. The results of Gómez-Navarro et al. (2015) give some indications that this expectation may be warranted. In the worst case, our analyses point out that one of the ~~sources of information completely contradicts the other data sets~~data sources contradicts the others.

For any given window, the fitted distribution parameters allow calculating various properties. For example, we can consider the changing amount of precipitation, which one would describe as average, extremely high, or extremely low for subsequent periods. In the SPI-literature, the 6.7th and 93.3th percentiles represent traditionally the regions of severe (and extreme) dryness/wetness of the probability density function. Accordingly, we subsequently compare 6.7th and 93.3th percentiles for the fitted distributions over time. Further, we can compare the moments of the distributions. We choose to show the square-root of the Weibull distribution variance, i.e., the Weibull standard deviation over sliding windows. This provides an additional clarification of how the precipitation distribution changes over time. The Appendix C shows parameters for the distribution fits.

The fitted parameters allow further analyses, e.g., we can compare how likely a reference amount of precipitation is for different periods. We do this for 50th, 6.7th, and 93.3th percentiles in a reference year. We choose 1815 CE as reference year, since it is included in all data sets and it allows potentially equivalent analyses of the PMIP3 past1000 simulations (e.g., Schmidt et al., 2011).

~~Agreement on changes in percentiles and standard deviation increases our confidence in our understanding of forced and unforced changes in precipitation variability and projected future precipitation variations. Disagreement on estimated changes may highlight differing internal climate variability between observed, reconstructed, and simulated data or it may signal that the simulation does not correctly capture forced variations.~~

3.2 Smoothing

Performing the transformation to standardised precipitation over 51-year windows results in smoothed estimates. For convenience, we additionally plot smoothed time series in a number of Figuresfigures. Filtered series are solely used for visualisation.

We use a Hamming window. In most cases, this has a length of 51 points but we also occasionally use different window lengths. The 51-point Hamming filter represents a different frequency cut-off than a simple 51-year moving median or moving mean as can be obtained from fitting the distributions over 51-year moving windows.

4 Results

4.1 Relations among data sets

4.1.1 Observational data and reconstructions

Figure 1 provides a first impression of the observational and reconstruction data we use in the following sections. All series are for the extended spring season from March to July on which we focus. Panels show 31-point Hamming-filtered time series. These allow a better qualitative assessment of the commonalities between the data sets and the differences compared to, e.g., 11-point or 51-point Hamming-filtered time series.

Observational precipitation series from the Met Office Hadley Centre for South-West, South-East, Central England, and England-Wales show high agreement in their variations on these time-scales for the period of overlap (see Figure 1a, particularly the period 1890 to approximately 1980). The instrumental time series for Kew Gardens and Pode Hole show more disagreement in certain periods for the considered smoothing, i.e., they even evolve oppositely at certain times, e.g., the mid-19th and mid-20th centuries (see Figure 1b). The instrumental data for Oxford appears to agree better with the data for Kew Gardens, which is to be expected from the geographic locations of the stations. Visually, both reconstructions agree less well with the observational series and with each other than the observational data does (see Figure 1c). This holds for their variations and their overall level of variability. Figure 1d adds the Central England temperature data for MAMJJ for completeness sake.

Correlation matrices (Figure 2, and supplementary manuscript assetdocument) and scatterplots (see manuscript assetsupplementary document) emphasize the differing agreement between the various data sources even more clearly on interannual time-scales. Figure 2 presents the correlation matrix for complete observations, i.e. for the period 1873 to 1994 when all records have data. Correlation coefficients change slightly if we consider pairwise complete records. Relations among precipitation data sets are always positive. They are very strong between the England-Wales data and its subdivions, between the Kew Gardens series and the South-East England data, between the Pode Hole series and the Central England data, and between the Oxford record and the South-East England data as well as the England-Wales precipitation. The relation between the two reconstructions is also rather strong over the sub-period. Correlations are, however, weaker between the reconstructions and the observed series.

There is a generally negative relation between the Central England temperature and the precipitation data sets for the chosen extended spring season from March to July. It is weakest for the Southern-Central England reconstruction but also rather

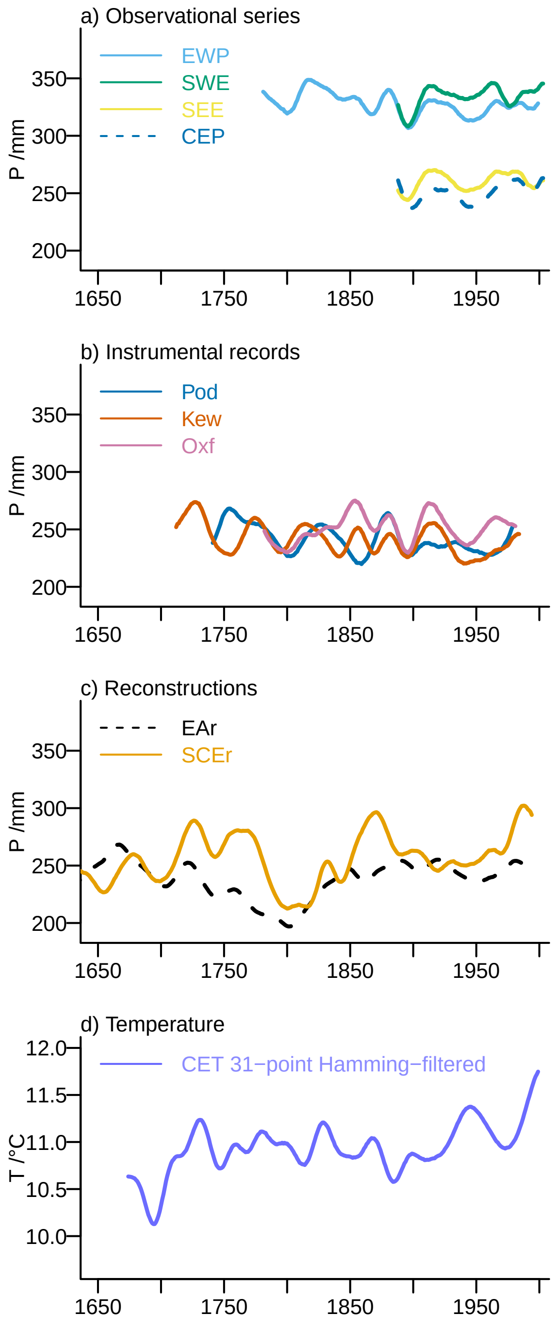

Figure 1. Visualisation of the observation based records for the extended spring season March to July (MAMJJ). We show 31-point Hamming-filtered time series for a) the Met Office Hadley Centre observational precipitation series for England-Wales (EWP), South-West (SWE), South-East (SEE), and Central England (CEP), b) the instrumental precipitation series for Pode Hole (Pod), Kew Gardens (Kew), and Oxford (Oxf), c) the precipitation reconstructions for East Anglia (EAr) and Southern-Central England (SCEr), and d) the Central England Temperature (CET) data.

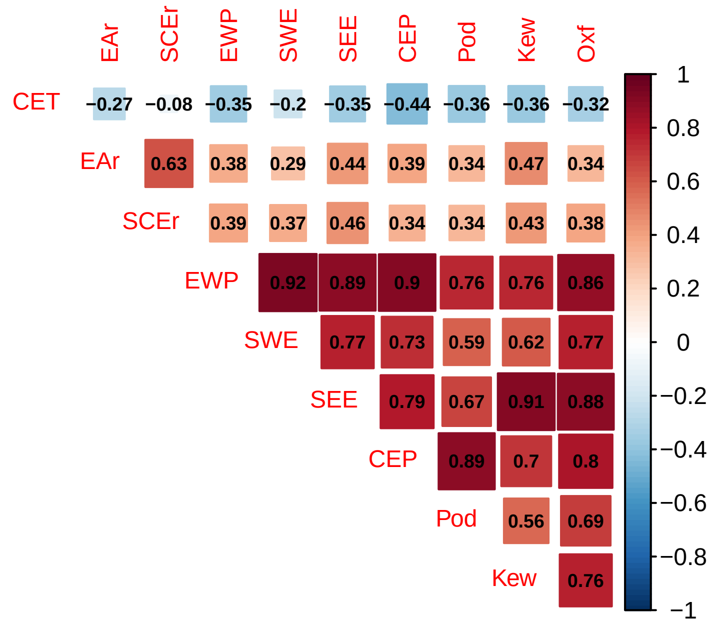

Figure 2. Correlation matrix for complete correlations between the observation or paleo-observation based data sets Central England Temperature (CET), East Anglia precipitation reconstruction (EAr), Southern-Central England precipitation reconstruction (SCEr), England-Wales precipitation (EWP), South-West England precipitation (SWE), South-East England precipitation (SEE), Central England precipiation (CEP), Pode Hole precipitation (Pod), Kew Gardens precipitation (Kew), and Oxford precipitation (Oxf). Complete correlations mean, we only use the years 1873 to 1994 for which all records have data. The season for all records is MAMJJ.

weak for the East-Anglia reconstruction and the South-West England record from the Met Office Hadley Centre. Scatterplots emphasize that even the temperature-precipitation relations with larger correlations scatter widely (not shown). Temperature-relations are stronger for the observationally based data from the Met Office Hadley Centre and the instrumental series for the summer season June to August (not shown).

Correlations for non-overlapping 11-year averages are positive and strongest between the England-Wales precipitation and the two instrumental series (not shown, see supplementary ~~manuscript asset~~document, calculated for the period 1767 to 1986). This analysis ~~gives also~~ also gives reasonable correlations ($r \approx 0.51$) between the pair of reconstructions and between the instrumental series. Otherwise, correlations for this resolution are weak. Correlations for the extended spring season with the Central England temperature data are largest for the non-overlapping 11-year averages of the Kew Gardens instrumental series. We choose 11-year non-overlapping windows to balance the number of available data points and the filtering of interannual variability.

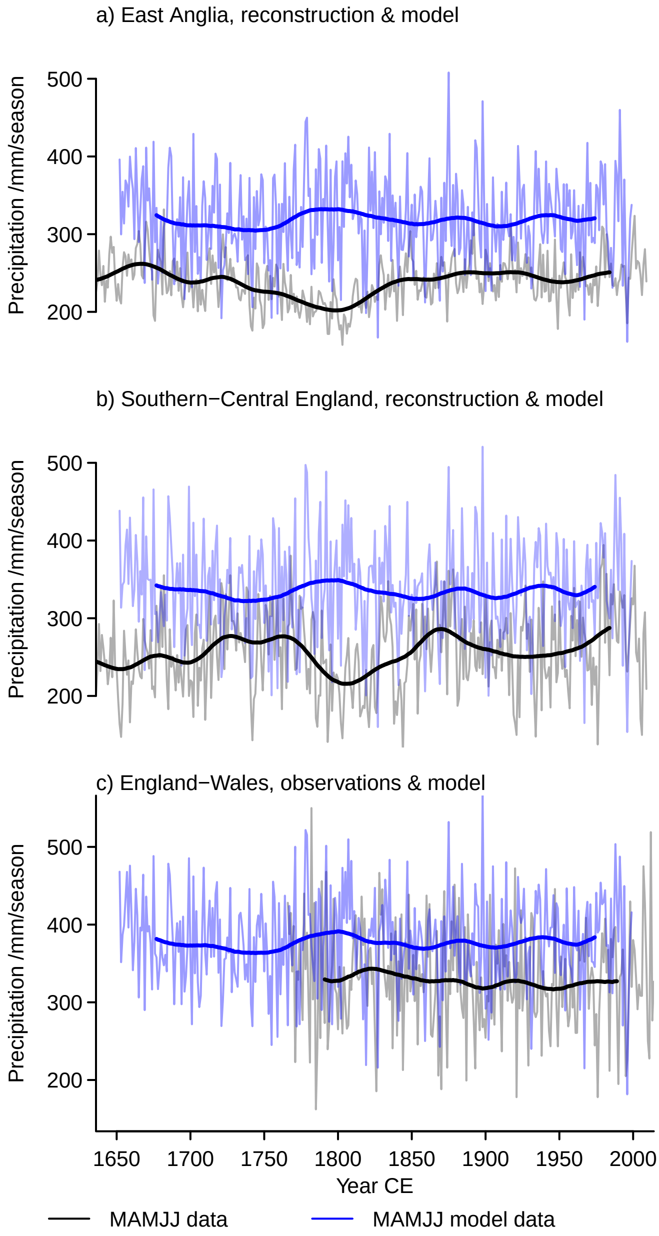

Figure 3. Extended spring (MAMJJ) precipitation in (paleo-)observation based data and simulation output, a) East Anglia precipitation in reconstruction (black) and regional model (blue), b) Southern-Central England precipitation in reconstructions (black) and regional simulation (blue), and c) England-Wales precipitation in observational data (black) and regional simulation (blue). We show interannual data (light colors) and 51-point Hamming-filtered data (solid colored).

4.1.2 (Paleo-)observational data and regional simulation output

Figure 3 presents the two reconstructions and the England-Wales precipitation in comparison to the respective data from the regional simulation. All data are again for the extended spring season from March to July (MAMJJ), and the panels zoom in on the period of the regional simulation. We show the interannual time series and the 51-point Hamming-filtered representation.

Considering the evolution of the records, the 51-point Hamming-filtered time series show pronounced differences besides some common features for the reconstructions for Southern-Central England (Wilson et al., 2013) and East Anglia (Cooper et al., 2013) (black lines in Figure 3a and b) similar to the representations in Figure 1. Both reconstructions feature a relative precipitation minimum centered on approximately the year 1800. The Southern-Central England reconstruction additionally displays a relative minimum in the early 20th century.

The observed England-Wales precipitation is available at monthly resolution from the year 1766 onward. The Hamming-filtered time series shows markedly less multi-decadal to centennial variability compared to the reconstructions, but the observations have much more interannual variability than the reconstruction for East Anglia and slightly more variability than the reconstruction for Southern-Central England (Figure 3c, black line). The filtered England-Wales time series also displays a slightly negative trend.

Differences between the simulated regional records are generally smaller (blue lines in Figure 3). Existing differences highlight the spatial heterogeneity of precipitation, e.g., interannual pairwise correlation coefficients are about 0.9 between the simulated East Anglia data and the other two records, while the simulated England-Wales precipitation correlates at approximately 0.97 with the simulated Southern-Central England data. Absolute interannual precipitation differences between the three data sets are at a maximum approximately 151 mm/season (not shown). A general feature for all regions is that excursions of the filtered simulation output often, but not always, are opposite to those of the reconstructions or observation time series.

There is an obvious bias in the absolute amounts between the simulation output and the other data sets. The simulation output series give larger precipitation amounts. We do not try to attribute this difference. We note that it is not as prominent for the more local comparison with the data from Rinne et al. (2013) for May to August and the bias is generally slightly negative for the summer season June to August for England-Wales precipitation (not shown, see supplementary ~~manuscript asset~~document). We assume that the differing spatial representations sufficiently explain the mismatch. However, the change of sign in the bias for the summer season suggests that the simulation overestimates spring precipitation, underestimates summer precipitation, and the positive spring bias is larger than the negative summer bias. See also Appendix A for a comparison of the simulation to observational data over the full European model domain. Figure 3 shows a common feature for all three comparisons. Simulated records appear to show opposite evolutions compared to the (paleo-)observations overall but particularly in the late 18th to early 19th century and in the early to mid-20th century.

This initial comparison already shows varying levels of agreement for the chosen data sets derived from observations and the reconstructions. It highlights that the ~~relation~~ relationship between the reconstructions and the observational data sets are weaker than between the instrumental data and the observational indices on interannual time-scales. Note that the regional

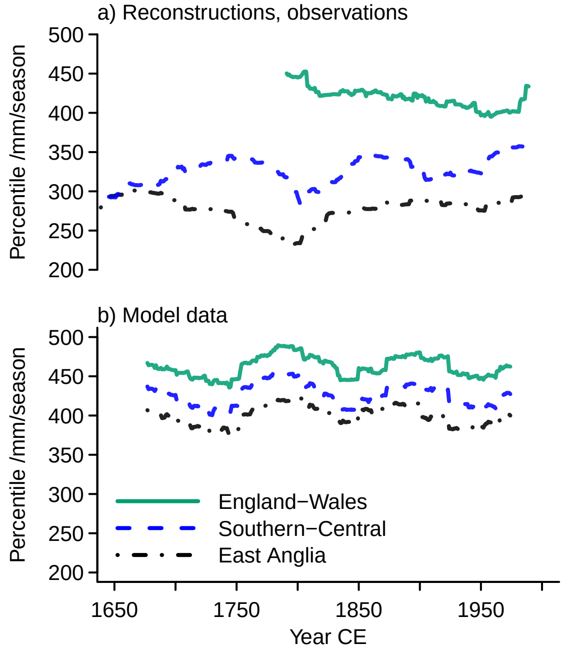

Figure 4. Visualisation of the MAMJJ precipitation amount identified as severely wet (93.3th percentile) over 51-year windows for England-Wales (green solid lines), Southern-Central England (blue dashed lines), and East Anglia (black dash-dotted lines) in a) reconstructions and observations, and b) simulations.

observational indices include information from the instrumental data. On longer time-scales the reconstructions align less well among each other than the observationally derived time series. However~~though possibly not surprisingly~~, the local, purely instrumental series also show more disagreement among each other than the derived larger domain products. Filtered regional time series evolve often visually oppositely in the simulation compared to the reconstructions and the observations.

So far, we used the precipitation and temperature data. In the following, we mainly use the information obtained via the transformation to standardised precipitation indices.

4.2 Comparing standardised precipitation data

Figure 4 to 6 add, respectively, the comparisons of the wet, i.e. 93.3th, percentile, the dry, i.e. 6.7th, percentile, and the square root of the Weibull distribution variance to the comparison of the interannual and filtered time series in the previous section.

4.2.1 Observations vs. Reconstructions

Since they represent different regions, we do not expect agreement in the absolute precipitation amounts representing wet conditions between the England-Wales precipitation data and the reconstructions in Figure 4a. We note that the difference between the wet percentile for the England-Wales precipitation and the reconstructions is larger than for the average amounts,

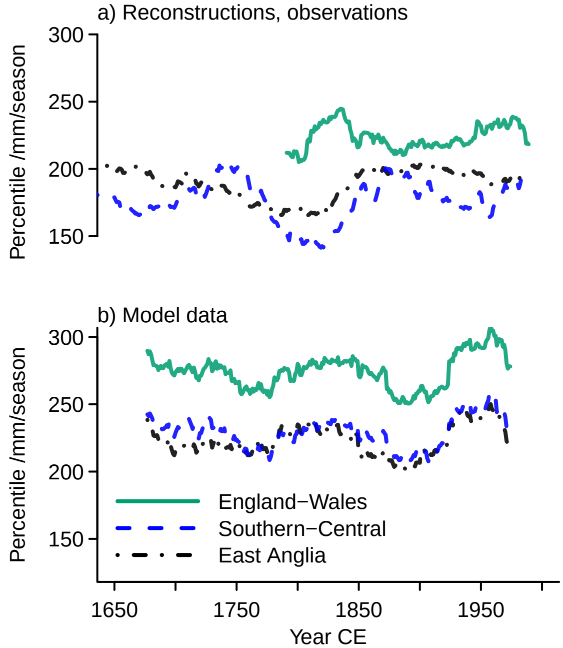

Figure 5. Visualisation of the MAMJJ precipitation amount identified as severely dry (6.7th percentile) over 51-year windows for England-Wales (green solid lines), Southern-Central England (blue dashed lines), and East Anglia (black dash-dotted lines) in a) reconstructions and observations, and b) simulations.

indicating a wider distribution for the data based on instrumental observations. Precipitation histograms confirm this (not shown). On the other hand, differences are smaller for the dry percentile (Figure 5). Nevertheless, this is a sign that the reconstructions underestimate the width of the precipitation distributions of 51-year window climatologies.

Reconstructed and observation-based time series show a slightly opposite trend for the wet percentile over much of the period of ~~the observational England-Wales time series~~ their overlap (Figure 4) ~~. Smaller scale~~ from the second half of the 18th century to the mid-20th century. Smaller amplitude variations in the beginning of the wet percentile series are also opposite. The dry percentile series ~~lack the clear overall~~ do not have a long term trend but multidecadal variations evolve oppositely between reconstructed and observed dry percentiles from the end of the 18th century to the early 20th century (Figure 5).

The opposite trends in the wet percentiles mean that the observed 93.3th, i.e. wet, percentile represents lower precipitation amounts in the middle of the 20th century compared to the late 18th century, while the reconstructed wet percentile represents larger precipitation amounts in the middle of the 20th century compared to the late 18th century (Figure 4). Similarly the opposite multidecadal variability in the 6.7th, i.e. dry, percentiles of reconstructions and observations means that when the reconstructions represent a drying of the dry percentiles, the observations indicate ~~that very dry conditions are already identifiable for larger precipitation amounts in a period~~ the opposite and vice versa (Figure 5). Generally, the series for the severe to extreme dryness and wetness percentiles reflect the smoothed evolution of the respective data set (compare Figure 3).

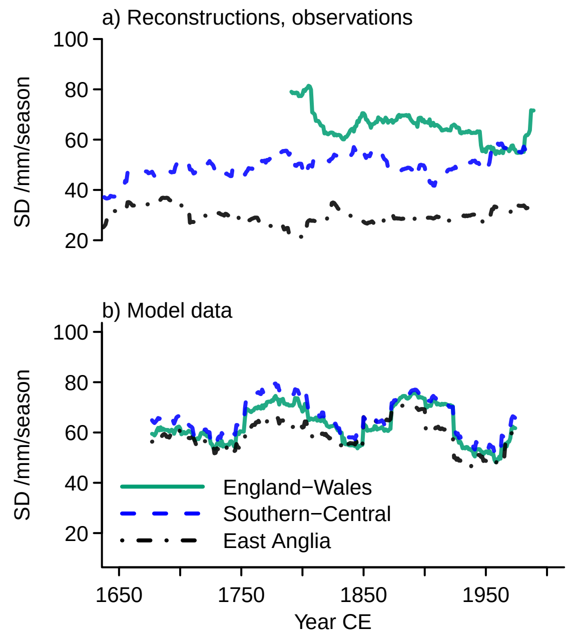

Figure 6. Visualisation of Weibull standard deviations (SD) over 51-year windows for MAMJJ precipitation for England-Wales (green solid lines), Southern-Central England (blue dashed lines), and East Anglia (black dash-dotted lines) in a) reconstructions and observations, and b) simulations.

We note that the data of Rinne et al. (2013) for Southern England in summer display an apparent opposite evolution of wet percentiles for the period of overlap between reconstruction and observations from the late 18th to the late 19th century. On the other hand dry percentiles agree well over this period (not shown, see supplementary manuscript assetdocument).

Parameters for the fitted distributions also allow us to evaluate the moments of the distributions. Estimates for the Weibull standard deviations (SD , in Figure 6) differ between observations and reconstructions as expected from the previously noted differences in percentiles. The reconstructions do reconstruction for East Anglia does not show a clear evolution in the Weibull standard deviations, whereas there is an increasing trend in the Weibull standard deviations for the Southern-Central England data. The observations show a slight reduction in the standard deviation until the middle of the 20th century, with a strong increase afterwards.

4.2.2 Simulation output

The simulated time series in Figure 3 show large similarities between regions. This is also the case for the wet and dry percentiles as well as for the standard deviations. Indeed, the respective statistics evolve simultaneously among the different regions, and the standard deviations overlap (Figures 4 to 6).

Thus, differences between regional domains are smaller for their simulated representations compared to the observed or reconstructed records. They are slightly more notable for the moving window statistics compared to the Hamming-filtered series. Dry percentiles are very similar for East Anglia and for Southern-Central England in the simulation but wet conditions require larger precipitation amounts for Southern-Central compared to East Anglia. Appendix B highlights that this may be due to sampling variability. Smoothed simulated data and wetness percentiles evolve similarly, but opposite evolutions of the dryness and wetness percentiles result in widening and shrinking of the distributions after approximately the year 1800.

4.2.3 Simulation output vs. observationally derived data and reconstructions

Simulations and reconstructions agree only minimally do not agree on the time evolution of precipitation percentiles (Figures 4 to 6). The simulation appears to agree sligthly with the reconstruction for Southern-Central England in the late 19th century in the wet percentile (Any hint of an agreement between reconstructed and simulated data is likely due to randomness (compare Figure 4). However, then the dryness percentile evolve in an opposite way There is instead a tendency towards opposite time evolutions between the data sources. This is best seen in the dry percentiles from the mid-18th to mid-20th century (Figure 5).

This apparent opposite evolution is the most common feature when comparing the percentiles derived from the simulation and from the reconstructions. When the percentile series for the reconstructions show minima, the simulation commonly shows maxima and vice versa. Obviously, using an ensemble of regional simulations probably would show different trajectories. This does would show a range of trajectories. Therefore, these results do not preclude per-se that the model is capturing basic physical characteristics of precipitation variability in northwestern Europe.

The smoothed representations of the simulation output and the smoothed observed England-Wales precipitation show only small multidecadal variations, which appear to be more or less opposite in simulated and observed estimates (compare above in opposition (Figure 3). The wet percentiles do not show any agreement although they both have a relative maximum in the late 18th century (Figure 4). On the other hand, the dry percentiles show approximate agreement in their evolutions with over the full time period of their overlap. Particularly noteworthy are approximately concurrent maxima in the early 19th century and in the middle of the 20th century (Figure 5). Similarly the Weibull standard deviations show some commonalities between the simulated representation of the England-Wales precipitation and the observations (Figure 6) over the full period of their overlap.

We note that there is neither any clear commonality nor any overly opposite evolution in the dry percentiles when comparing the regional simulation to the reconstruction for Southern England summer precipitation by Rinne et al. (2013, not shown, see supplementary manuscript asset)Rinne et al. (2013, not shown, see supplementary document). The wet percentiles, however, evolve oppositely in the 18th century but then show a common positive trend in the 19th century (not shown, see manuscript assetsupplementary document).

Figure B1 provides uncertainty estimates for part of our analyses. The figure shows 95% intervals of a bootstrap procedure sampling 1000 times 40 data points from the time windows each window and fitting distributions to these samples. The choice of 40 data points is an ad hoc decision that lies between the recommendation by McKee et al. (1993) of 30 samples and our window length of 51 years. Uncertainty on the fitted distributions varies in size over time and between data sets. Indeed, there are periods when sampling variability is so large that apparent differences in distributional properties between periods are not significant for most sources of information.

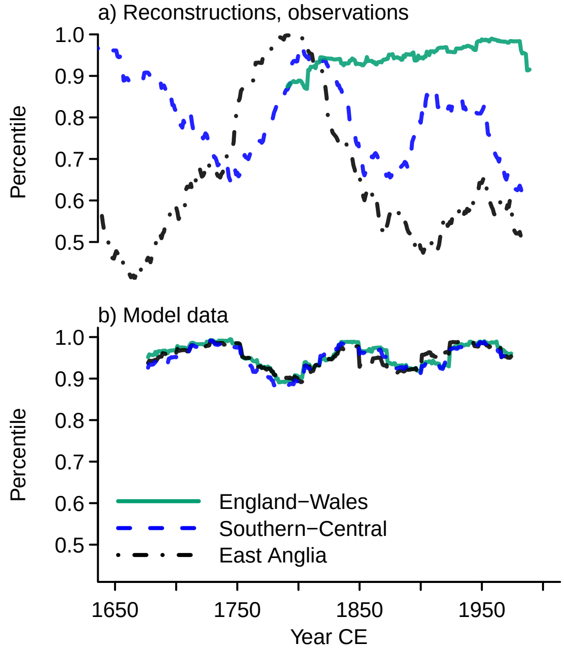

Figure 7. Visualisation of how percentile-values change for over windows. We show, which ~~percentiles~~ percentile the the 93.3th percentile MAMJJ precipitation amount for ~~the~~ a reference window ~~centered in year 1815 CE~~ represents over time for England-Wales (green solid lines), Southern-Central England (blue dashed lines), and East Anglia (black dash-dotted lines) in a) reconstructions and observations, and b) simulations. The reference window is centred in 1815 CE.

4.3 Changes in probability of certain precipitation amounts

In the methods section, we describe the procedure of calculating standardized precipitation indices over moving time windows. We obtain a distribution fit for each time window. The parameters of the fit for a window allow us to identify the probability of a precipitation amount for the respective window. Figures 7 to 9 present changes in the probability of certain amounts of precipitation, i.e. lines are the changing percentiles represented by a given amount of precipitation over time. The Figures show these changes for the precipitation amounts representing 93.3th, 50th, and 6.7th percentiles, respectively, in a reference window. For this comparison, the reference is the distribution of precipitation in the window centered around the year 1815 CE. The year 1815 CE is included in all data sets and it allows equivalent analyses of the PMIP3 past1000 simulations (e.g., Schmidt et al., 2011). We estimate and plot the percentiles that correspond to these reference precipitation amounts in other time windows.

The England-Wales precipitation shows a slight increase over time in the reference 93.3th percentile in the year 1815 CE (Figure 7a). Recently, there is a steep decrease in the series. Similarly, the 50th percentile for 1815 CE represents slightly larger percentiles over time (Figure 8a). On the other hand, there are weak multi-decadal variations in the series for the 6.7th percentile in the observations, and the 6.7th percentile from 1815 CE may become slightly less likely over time (Figure 9a).

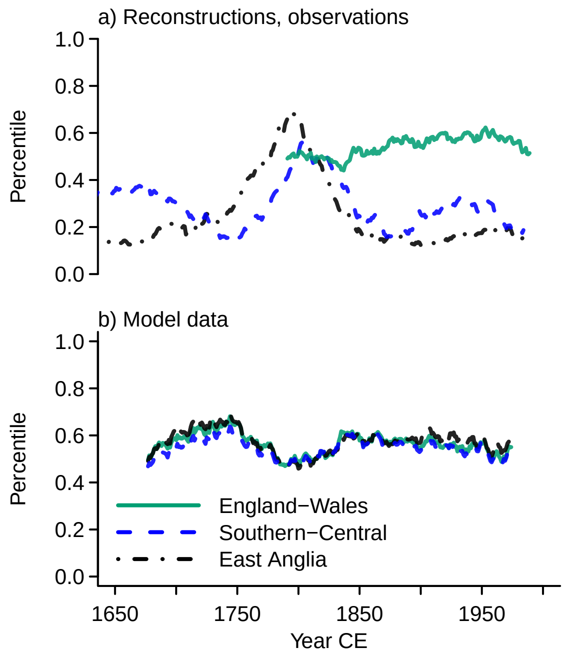

Figure 8. Visualisation of how percentile-values change for over windows. We show, which percentiles percentile the 50th percentile MAMJJ precipitation amount for the a reference window centered in year 1815 CE represents over time for England-Wales (green solid lines), Southern-Central England (blue dashed lines), and East Anglia (black dash-dotted lines) in a) reconstructions and observations, and b) simulations. The reference window is centred in 1815 CE.

Before turning to the reconstructions, we shortly note that the simulations show similar trajectories for all three percentile values and all three regions. There are not any obvious trends, but the series show multidecadal variations. The window centered in the year 1815 CE falls within a minimum or at the end of a minimum. The respective precipitation amount generally represents larger percentiles before the time window centered in 1815 CE. After this time window, the 6.7th and 93.3th percentiles both approach a maximum in the series (Figures 7b and 9b). However, the 93.3th percentiles reach it about the year 1850 CE and the 6.7th percentile only in approximately the year 1900 CE, when the 93.3th percentile is again in a relative minimum. Thus, the wet and dry percentiles evolve oppositely from the early 19th century onwards, i.e. the distribution widens and shrinks since approximately the year 1850 CE. The median reference for 1815 CE also represents larger percentiles later but there is a slight decreasing trend from approximately the mid-19th century to the end of the simulation (Figure 8b).

The reconstructions for East Anglia and Southern-Central England have some peculiar features (Figures 7a to 9a). For one, it is not ideal to choose a reference year from the period around 1800 CE. The 6.7th percentile in 1815 CE is much less likely earlier and later in both regions (Figure 9a). Similarly, average precipitation around 1815 CE represents approximately the 20th percentiles in earlier and later periods for East Anglia (Figure 8a) but also represents much smaller percentiles in later periods for Southern-Central England. Severe and extreme wet conditions from this period may even represent long-term average conditions for East Anglia (Figure 7a). We note that comparisons to the data by Rinne et al. (2013) do not feature

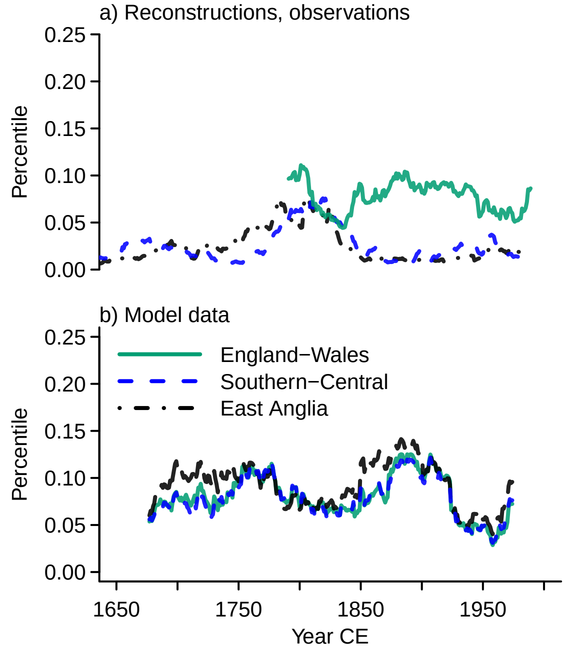

Figure 9. Visualisation of how percentile-values change for over windows. We show, which ~~percentiles~~ percentile the 6.7th percentile MAMJJ precipitation amount for ~~the~~ a reference window ~~centered in year 1815 CE~~ represents over time for England-Wales (green solid lines), Southern-Central England (blue dashed lines), and East Anglia (black dash-dotted lines) in a) reconstructions and observations, and b) simulations. The reference window is centred in 1815 CE.

such peculiarities (not shown) but using a simple scaling approach for the $\delta^{18}O$ data of Young et al. (2015) gives similar results ~~prior to approximately the year 1850 CE~~ (not shown, but compare information given in the supplementary ~~manuscript asset~~document).

In general, there are not any clear common evolutions between the different data sets before the 20th century. Only the dry percentiles in the simulation and the observations may evolve similarly in the period of their overlap (Figure 9). Interestingly, there is an apparent contrast between simulation and reconstructions with potentially opposite evolutions in the period of their overlap prior to the 20th century in all shown series. In the 20th century, on the other hand, some commonalities may be inferred at least for the series representing the reference 93.3th percentile (Figure 7).

Most prominent in these analyses is that the distributions for reconstructed precipitation show large shifts to larger precipitation amounts compared to the simulations and observations. In contrast, the simulation and observations vary only within a rather narrow range. This may relate to the weaknesses of the reconstructions in representing not only low-frequencies but also extremes (compare Cooper et al., 2013; Wilson et al., 2013). The regional simulation and the reconstructions show again an apparent opposite evolution for East Anglia and Southern-Central England. All sources of information tend to show shifts in the probability of precipitation amounts.

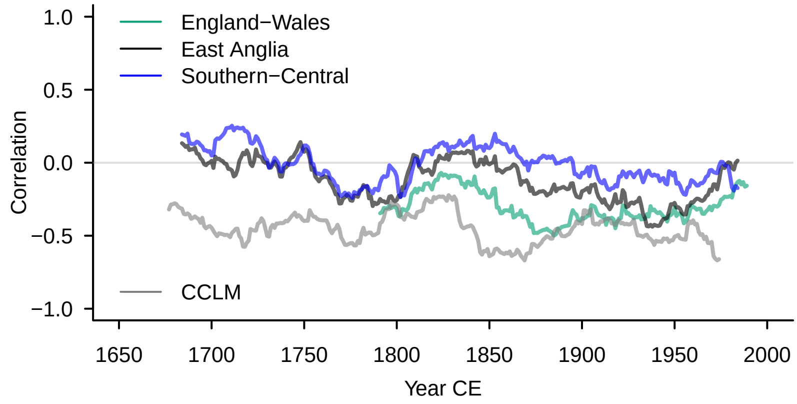

Figure 10. Interannual correlations over 51-year windows between extended spring (MAMJJ) Central England Temperature and various precipitation records: extended spring (MAMJJ) precipitation series for observational England-Wales-precipitation (green), reconstructed East Anglia precipitation (black), reconstructed Southern-Central England precipitation (blue). The grey line is for the simulated representations of the England-Wales MAMJJ precipitation and the Central England Temperature in MAMJJ.

4.4 Relation between Temperature and Precipitation in different data sources

~~The high amount of internal variability on local and regional scales complicates the comparison among different data sources when studying small regions.~~ We only shortly explore the interrelation between the regional temperature and ~~the precipitation variability .~~ precipitation variability focussing on the extended spring season from March to July.

We show how interannual correlations between the precipitation records and temperature series evolve over time for the chosen season. Figure 10 plots sliding interannual correlations for 51-year windows between the observed and reconstructed precipitation data and the Central England temperature as well as the correlation between simulated England-Wales precipitation and simulated Central England temperature. We plot correlations for the untransformed precipitation records. All records are for the MAMJJ-season. Obviously, the high amount of internal variability on local and regional scales complicates the comparison among different data sources when studying small regions.

We expect variability of moving correlation coefficients simply due to sampling variability (Gershunov et al., 2001). For example, a bootstrap procedure following Gershunov et al. (2001) suggests a 90% credible interval for 51-year moving window correlations of between approximately -0.59 and approximately -0.21 for a correlation of approximately -0.43 between simulated Central England Temperature and England-Wales precipitation over the full period. That is, variations in Figure 10 are probably within the sampling variability estimates for 51-year moving window correlations. That further implies that for overall uncorrelated data we can expect some windows to show statistically significant correlations. We do not show significance levels in Figure 10. We note that for 51-year windows and the time series characteristics of the data (e.g., approximately uncorrelated noise for seasonal precipitation), one may regard absolute values of correlation coefficients larger than 0.23 as statistically significant at the 5% level.

On interannual timescales and over 51-year moving windows, all data sets evolve similarly in Figure 10 for the extended spring season. However, observed and reconstructed data show weaker correlations in the late 20th century, while the correlation strength increases in the regional simulation. Both reconstructions do not show any statistically significant relation between temperature and precipitation over the full period. The reconstruction for East Anglia is intermittently negatively correlated with the temperature data. The observations show a notable negative relation from the second half of the 19th to the mid-20th century. Only correlations between the regional simulation temperature and precipitation are negative and relatively strong ($r \approx 0.5$) throughout the full period.

The observed negative relation is well known. For example, Crhová and Holtanová (2018) show a slightly negative correlation between temperature and precipitation in observations over the southern British Isles in spring and summer. They also show that regional climate simulations usually capture this feature successfully.

5 Discussions

Our understanding of hydroclimatic changes for future and past climates increased notably in recent years (compare, e.g., PAGES Hydro2k Consortium, 2017), especially for drought (see **?**). Nevertheless comparing our various sources of information for past hydroclimatic changes remains challenging (compare PAGES Hydro2k Consortium, 2017).

Hydroclimate comparisons between different data sources may focus on floods, on drought using indices like the PDSI, or on precipitation, including the SPI used in this study. Humans notice effects of climatic changes mostly on local to regional scales. We expect that changes in precipitation are of larger immediate relevance for local communities than changes in drought indices. Thus, we study precipitation changes in regional small domains.

In this section, we more extensively discuss the SPI, our data choices, and our results. We also discuss additional data sets presented in the supplementary manuscript asset.

4.1 Method

Much research on hydroclimatic variability focusses on drought because of its effects on society and environment. Based on criteria suggested by Keyantash et al. (2002), the Interregional Workshop on Indices and Early Warning Systems for Drought proposed the Standardized Precipitation Index (SPI, McKee et al., 1993) as common index to monitor meteorological drought (Hayes et al., 2011). The SPI should complement previously used indices and facilitate comparability between different regions. Raible et al. (2017) find the SPI to be a reliable drought index for Western Europe including the British Isles. The standardisation allows further applications, e.g., flood monitoring (Seiler et al., 2002), and the easy comparison of normal, dry, and wet conditions between different sources of data.

The SPI uses only precipitation, which makes it an ideal and relatively straightforward tool for comparing hydroclimatic data between different data sources. Precipitation is a standard output of simulations, long instrumental records exist for various locations, and a number of reconstructions exist as well although paleo-observations may represent soil moisture rather than precipitation.

4.1 Considering further reconstructions and global simulations

However, as the SPI uses only precipitation, it is of less value when the interest is in, e.g., Here, we shortly describe additional results. If we perform similar analyses as described above but on a selection of the water supply, runoff, or streamflow (but see Seiler et al., 2002). The focus on precipitation also limits the applicability for studying temperature sensitive parts of the hydrological cycle and impacts on biological and anthropogenic systems (e.g., PAGES Hydro2k Consortium, 2017; Keyantash et al., 2002; Hayes et al., 2011; Van Loon, 2015).

Most importantly, the interpretation of our results relies on the robustness of the SPI-transformations. Sienz et al. (2012) highlighted that the Weibull distribution performed better in transforming the England-Wales precipitation data on a monthly time-scale compared to a number of other distributions. However, other distributions outperformed the Weibull distribution for other data sets and other SPI time-scales.

We fit distributions over sliding 51-year windows. While we thus use more data points than recommended by McKee et al. (1993), we still use less than the 60 points for which Guttman (1994) finds convergence of higher order L-moments. Appendix Figure B1 shows 95% intervals of a bootstrap procedure sampling 1000 times 40 data points from each window and fitting distributions to these samples. Uncertainty on the fitted distributions varies in size over time and between data sets. Indeed, there are periods when sampling variability is so large that apparent differences in distributional properties between periods are not significant for most sources of information. PMIP3-ensemble of global simulations (Schmidt et al., 2011), we do not find commonalities between the simulations or between the simulations and the other sources of information (not shown, see supplementary document). If we use different reconstructions, agreement between simulation and reconstructed precipitation does not necessarily increase, but differences between reconstructions and observations may be reduced (not shown, see supplementary document).

In a sense, the SPI calculations provides us with information on the climatological precipitation distributions over moving windows. The comparison becomes therefore an assessment of the changes in the climatology between different 51-year periods. This climatology does not only provide information on the mean state but also further derived statistics, the extreme percentiles for the individual windows, We use two different reconstructions based on $\delta^{18}O$. For one, we obtain the precipitation reconstruction by Rinne et al. (2013) for Southern England for the May to August extended summer season. Secondly, we use the isotope records for England and the variability in these periods.

If this climatology for the observations is the target climatology, an ensemble of climate simulations should sample this distribution interannually following the paradigm of a statistically indistiguishable ensemble (Annan and Hargreaves, 2011). Our analyses compare how well the climatologies agree in different sources of data.

4.2 Data

4.1.1 Observations

Starting from the available regional climate simulation, we choose the region for our study based on the availability of precipitation observation and reconstruction data. There are long records of instrumental measurements of climate parameters for a number of locations in Europe. The British Isles are unique because there exist long observation based indices for precipitation and temperature in form of the England-Wales precipitation data (Alexander and Jones, 2000) and the Central England Temperature data (Parker et al., 1992) besides the long instrumental precipitation observations, e.g. in southern Great Britain, for Kew Gardens, Oxford, and Pode Hole. Additionally there are instrumental records from Inverness, Edinburgh, and Manchester starting in the 1780s, which we do not use because of their northern locations. For Ireland, **?** provide a monthly rainfall series starting in the year 1711, which we do not discuss here either because of its distance to our study region.

The Central England and Wales by Young et al. (2015) and scale the composite against the observational England-Wales observation indices are good representations of the late 20th century climate of the South of Great Britain according to Croxton et al. (2006). Our Figure 2 also shows the strong correlation between the various precipitation records based on observations. Note that the composite series naturally rely on the instrumental series. Weakest relations occur for the instrumental series of Pode Hole with the sub-regional series for South-West England and for the relation between the two instrumental records from Pode Hole and Kew Gardens. Frank et al. (2007) noted the uncertainties in early instrumental temperature observations. Additionally, the very early data in the Central England Temperature data includes non-instrumental indirect data to infer past temperature. Similarly, early precipitation observations require rigorous quality control. In this context, Woodley (1996) reviews the history of England and Wales as well as Scotland precipitation data.

Figure 2 further shows the negative relation between temperature and precipitation for our domain of interest. Tout (1987) does not find any changes in the negative relation between England-Wales precipitation andCentral England Temperature We follow the procedure described by Young et al. (2015) but for two seasonal estimates, the extended spring from March to July used in our main analyses and, following Young et al. for the summer season from June to Augustbetween 1766 and 1980.

4.1.1 Reconstructions

Paleo-reconstructions of the recent past have made notable progress both in the spatial coverage and in the quality of the reconstructions by incorporating so far unexplored data sources and new methods. **?**, for example, highlight the importance of ship-based observations recorded in log books for reconstructing large-scale fields. Initiatives like or ACRE (Atmospheric Circulation Reconstructions over the Earth,) are invaluable for such efforts and also aid reanalysis projects like the twentieth century reanalysis (**?**), the reanalysis of global fields for the period1600 to 2005 by Franke et al. (2017), or the last millennium climate reanalysis (**?**).

For the hydroclimate, there are a number of gridded reconstructions covering the European domain. Continental domain gridded precipitation reconstructions are, e.g., Pauling et al. (2006), Casty et al. (2007), and Franke et al. (2017). Reconstructions of drought indices like the Palmer Drought Severity Index (PDSI) exist as gridded products, too, for various regions of the world including Europe (The Old World Drought Atlas, Cook et al., 2015). These products allow assessing the quality of the hydroclimate in paleoclimate simulations (Smerdon et al., 2015).

The availability of observational data and regional reconstructions motivates our domain choice covering southern Great Britain. We decide to use regional precipitation reconstructions instead of gridded products to minimise the effect of the reconstruction method on the results. We focus on precipitation as it allows the direct comparison with long instrumental records and it is a parameter directly experienced by people.

We mainly focus on two reconstructions based on tree-ring widths measurements (Cooper et al., 2013; Wilson et al., 2013). These authors calibrate their tree-ring data against gridded precipitation beyond their target regions of Southern-Central England and East Anglia, respectively. Thereby the reconstructions are possibly biased beyond their respective regions of interest. They compare their reconstructions against the long instrumental records and find a lack of stability of The supplementary document provides some details for our summer season scaling of the isotope data of Young et al. (2015). The most striking feature is again a notable difference in the percentiles prior to time windows approximately centered in the year 1850 compared to the later period. This feature resembles the behavior of the relation to the instrumental data. They discuss the limitations of their reconstructions representing less than 40% of the regional precipitation variance over the 20th century.

Although the reconstructions show a notable amount of low frequency variability, Cooper et al. (2013) cautions against too much confidence in the low frequency precipitation variability in their reconstruction. Wilson et al. (2013) and Cooper et al. (2013) emphasize the weaknesses of their reconstructions in representing extreme years. On the other hand, both are confident in the mid- to high-frequencies of their reconstructions.

Both, Wilson et al. (2013) and Cooper et al. (2013) discuss the possibility that the tree-species used for their reconstructions were less sensitive to precipitation over certain periods, e.g., the early 19th century. That is, the proxies, theoretically representing a precipitation signal, also contain a temperature signal, for instance, if they are sensitive to soil moisture. Wilson et al. further suggest an effect of the Industrial Revolution and the associated pollution on the trees in their selection. Wilson et al. (2013) also discuss the reliability of the instrumental data but conclude this is likely not an issue.

Besides these two tree-ring width based reconstructions, the works by Rinne et al. (2013) and Young et al. (2015) use $\delta^{18}O$ data to reconstruct precipitation for Southern England and Great Britain respectively. We shortly discuss results for both reconstructions below. Rinne et al. (2013) calibrate and scale their local isotope data from 1613 to 1893 CE against the station observations from Oxford for the period 1815 to 1893 CE and concatenate the reconstruction with the observations for 1894 to 2003 CE. They target an extended summer seasonfrom May to August. Young et al. (2015) use the England-Wales summer, June to August, precipitation as scaling target for a composite of eight isotope records from Scotland, Wales, and England for the period 1766 to 2012 CE. Both publications by Rinne et al. (2013) and Young et al. (2015) note the differences of their reconstructions to the tree-ring width based works by Wilson et al. (2013) and Cooper et al. (2013).

Young et al. (2015) conclude that these differences make it unlikely that the tree-ring based works and their $\delta^{18}O$ based work represent the same environmental parameter, and they emphasize the lack of a calibration against regional precipitation data. They further discuss the reasons given by Wilson et al. (2013) for the lacking stability of the Wilson et al. reconstruction, namely, different climate-sensitivity of the trees, unreliable instrumental data , and pollution. Young et al. (2015) conclude that their own data reliably reflects precipitation while the tree-ring widths most likely represent the combination of various environmental influences on tree growth instead of a single climate parameter. width based reconstructions. While this may be due to the chosen calibration method and period, it appears more likely that there is a problem in the relation between isotopes and precipitation for this early period.

4.1.1 Simulations

Comparing our extended spring season scaling to the equivalent observations, there is limited agreement for the dry percentile after approximately the year 1850 (not shown) but otherwise we cannot find any consistency of this data compared to the observational counterparts. We also see no agreement between the data by Young et al. and the regional simulation output.

In contrast to present-day and future scenario regional simulations, the 0.44 by 0.44 degree resolution of our CCLM simulation represents a comparatively coarse resolution dynamical downscaling. Sørland et al. (2018) discuss how the use of a model-chain including global and regional climate simulations assists studies on regional climates. Besides other models, they also use CCLM in a 50km setup comparable to the simulation used here. Their work emphasizes that improved representation of regional climate in a regional simulation is not solely due to the increased resolution but may be due to different strategies in model-building and tuning. Pinto et al. (2018) study global and regional simulations, including CCLM-simulations, for southern Africa. They explain differences in results from regional and global simulations by an interplay between the representation of sub-grid-scale processes in the different models and factors related to The period covered by the data of Rinne et al. (2013) only shortly overlaps with the period of the observational data. For this overlap dry percentiles tend to agree with the observations but wet percentiles evolve oppositely (compare supplementary document). The change in average precipitation for a reference year also agrees between both data sets for the time of overlap (not shown). Compared to the increased resolution. Blenkinsop and Fowler (2007) note that regional climate models may be deficient in their ability to model persistent low precipitation episodes for the British Isles, which has repercussions for their representation of drought events.

The review by Ludwig et al. (2018) reports more realistic distributions for precipitation in regional paleoclimate simulations. Flato et al. (2013, chapter 9 of the IPCC AR5) review the progress of regional downscaling and high-resolution modelling. They emphasize that the skill of such exercises depends on the model used, the season, the domain of interest, and the considered meteorological variable. They highlight studies showing that there is not a linear increase in simulation skill towards higher resolutions. Higher resolutions typically provide more reliable estimates of extremes, including heavy rainfall. Flato et al. (2013) view regional modelling as a valuable extension of global climate modelling.

The quality of the simulated precipitation still strongly depends on the parameterisations implemented in the regional climate model. Precipitation, especially convective precipitation events, are still sub-grid processes, even within regional climate models. Concentrating on accumulated amounts on seasonal time-scales and their long-term changes, however, allows at least a more robust comparison of simulated precipitation to observed and reconstructed data. regional simulation output, evolutions tend to be opposite.

Shortcomings of the various data sets limit our expectations to what extent they can reflect comparable variations among each otherIf we consider the relation between temperature and precipitation in the additional data sets and their respective seasons, the disagreement between data sources changes compared to our main analyses (not shown). The observations show consistently negative correlations for the summer season, and the scaled isotope data by Young et al. (2015) agrees quite well with the summer observations except for a large part of the 20th century when it shows a markedly weaker negative correlation (not shown). The simulation shows again generally stronger correlations compared to the other data in summer and shows some agreement with the observations in the industrial period since approximately the year 1850 (not shown). If we correlate the scaled isotope data to the temperature for an extended spring season from March to July, the correlations are quite similar to those for the larger domain simulation output but differ notably from the observations (not shown). The extended summer (MJJA) reconstruction by Rinne et al. (2013) agrees well with the respective observations in a consistently negative correlation (not shown). The relation is weaker for the reconstruction prior to the period of the Oxford precipitation observations (not shown).

4.2 ~~Discussion of Results~~

5 Discussions

5.0.1 ~~Validity of approach~~

5.1 Validity of approach

Information from reconstructions of climate parameters and from simulation output together increase our understanding of past climates. The PAGES Hydro2k Consortium (2017) made recommendations for valid and appropriate comparisons of hydroclimate data from both sources of information. Here, we consider approximately the last 350 years by comparing both estimates to long instrumental data. We have to consider whether our analyes are appropriate in the sense of the recommendations concerning uncertainties, the properties compared, and the expectations underlying the comparison (PAGES Hydro2k Consortium, 2017).

The observational England-Wales precipitation data is a weighted composite of regional series which again are based on instrumental information. The input changed over time. Similarly, the reconstructions combine spatially distributed proxy, e.g., tree-ring width series into regional scale composite series (Cooper et al., 2013; Wilson et al., 2013) to maximise the common signal between different locations. On the other hand, the simulations are aggregations of various grid-point time series from the simulation output. We assume that the compositing and the aggregation have similar effects in removing local variability. In this respect, records from different sources are similar to each other and thus our comparison appears valid.

Explicit uncertainty estimates are only available for the reconstruction for East Anglia and only for a low-pass filtered version of the data (Cooper et al., 2013). Our results as well as the discussions of Cooper et al. (2013), Wilson et al. (2013), Rinne et al. (2013), and Young et al. (2015) emphasize that uncertainties for the reconstructions are potentially large and that even the relation to precipitation is not necessarily valid for ~~parts of them~~some periods. Similarly, uncertainties affect the simulations not only with respect to our domain choice but also with respect to the algorithms and parametrisations implemented for simulating precipitation in the regional climate model.

Considering the limitations of ~~the~~ any simulation and the a priori known shortcomings of the reconstructions, questions may arise on the validity and robustness of our analyses. ~~We~~ Even if one assumes that prior discussions on the reconstructions invalidate their use, they would at least be a useful data source for our first goal of highlighting the benefits of adding the SPI to our set of tools for studying past precipitation variability.

However, we do not agree with such an assumption. The reconstructions are still, at least 'preliminary' (as stated by Cooper et al., 2013), estimates of past precipitation for the southern British Isles. As such, it is of value to include them in a comparison of distributional precipitation characteristics between different data sources for this domain. It is further of interest to highlight for any available reconstruction in which properties the reconstructed precipitation distributions agree or disagree with the other sources of information. That is, understanding our sources of information about past climates requires to identify their strengths as well as their shortcomings.

More generally, we argue that the transformation to standardized indices provides a sound basis for equivalence between the different precipitation estimates for subsequent comparisons of the distributional properties.

Then, we assume that the comparison becomes informative for changes over time between these distributions. While we cannot expect accurate or even approximate temporal agreement between time series from simulation output and observation based data on either interannual or multi-decadal time-scales because of internal variability, the transformation makes our comparison one of climatologies. Furthermore, one may assume that the evolution of percentiles and variability may be more consistent between the different data sets than the average conditions.

5.1.1 Additional analyses

We find that the considered observations, reconstructions, and a regional simulation only show limited agreement in their representation of precipitation for a small regional domain covering southern Great Britain for approximately the last 350 years. Striking are the differences between the tree-ring width based reconstructions (Cooper et al., 2013; Wilson et al., 2013) and the observations, which again highlight the shortcomings of the two considered reconstructions (compare Young et al., 2015). It is noteworthy that there are multiple periods where simulation output and reconstructions evolve oppositely. Possibly surprising is occasional temporal consistency in some of the measures between regional simulation and England-Wales precipitation data.

We performed similar analyses on a selection of the PMIP3-ensemble of global simulations (Schmidt et al., 2011). There, we see no commonalities between the different simulations or the simulations and the other sources of information in the analyses of precipitation distribution properties (not shown, see supplementary manuscript asset).

If we use different reconstructions, agreement between simulation and reconstructed precipitation does not necessarily increase, but differences between reconstructions and observations may be reduced (not shown, see supplementary manuscript asset). We use two different reconstructions based on $\delta^{18}O$. For one, we obtain the precipitation reconstruction by Rinne et al. (2013) for Southern England for the May to August extended summer season. Secondly, we use the isotope records for England and Wales by Young et al. (2015) and scale the composite against the observational England-Wales precipitation data. We follow the procedure described by Young et al. (2015) but for two seasonal estimates, the extended spring from March to July used in our main analyses and, following Young et al. for the summer season from June to August. See the supplementary manuscript asset for some details of the comparison to the summer scaling.

For the scaled data by Young et al. (2015), the most striking feature is again a notable difference in the percentiles prior to time windows approximately centered in the year 1850 compared to the later period. This feature resembles the behavior

~~of the tree-ring width based reconstructions. While this may be due to the chosen calibration method and period, it appears more likely that there is a problem in the relation between isotopes and precipitation for this early period. Comparing the data to the extended spring observations, there is limited agreement for the dry percentile after this early period (not shown). For other periods, the moving window distributions show prominent inconsistencies compared to their observational counterparts. Comparing the data by Young et al. to the regional simulation also does not show any agreement.~~

~~The period covered by the data of Rinne et al. (2013) only shortly overlaps with the period of the observational data. For this overlap dry percentiles tend to agree with the observations but wet percentiles evolve oppositely (compare supplementary manuscript asset). The change in average precipitation for a reference year also agrees between both data sets for the time of overlap (not shown). Compared to the regional simulation output, evolutions tend to be opposite.~~

5.1.1 ~~Implications of main results~~

5.2 Implications of the main results

Our analyses highlight the shortcomings of different reconstructions relative to observations. We also see that differences to observations may be comparable for reconstructions and simulations. Our ~~methods also show~~ approach further shows that apparently the reconstructions and the simulations ~~frequently evolve differently~~occasionally evolve in opposite directions. This may signal that we indeed do not perform a valid comparison, that simulations may misrepresent forced responses, or, considering the reconstructions' relation to temperature, that the reconstructions do not fully relate to precipitation.

We expect disagreement between simulations and observations ~~on some levels,~~ not least because of differing influences of internal variability (see discussions below). More critical is the lack of consistency between reconstructions and observations. Most notably the reconstructions show unrealistically large changes in the cumulative probabilities represented by certain precipitation amounts for the extended spring season MAMJJ (compare Figures 7 to 9). The reconstructions do not reliably represent the extended spring precipitation distributions in specific periods. ~~Plotting the anomalies for the observations and reconstructions (not shown) displays much stronger variability over the common period in the reconstructions compared to the observations and at times opposite trends.~~

One result is the inconsistency of the relations between temperature and precipitation in the data sets for the considered domains for the extended spring season from March to July. Tout (1987) and Crhová and Holtanová (2018) both note the negative relation between temperature and precipitation observations for Britain. ~~We find this~~ Tout (1987) does not find any changes in the negative relation between England-Wales precipitation and Central England Temperature for the summer season from June to August between 1766 CE and 1980 CE. We find the negative relation for the extended spring only consistently in the simulation, and ~~over the more recent period~~ from approximately 1850 CE to 1950 CE also in the observations. The tree-ring width based reconstructions do not show any clear relation for the extended spring season. ~~If we consider other seasons, the~~ The disagreement between data sets changes for other seasons (not shown). ~~The observations show consistently negative correlations for the summer season, and the scaled isotope data by Young et al. (2015) agrees quite well with the summer observations except for a large part of the 20th century when it shows a markedly weaker negative correlation (not shown). The~~

~~simulation shows again generally stronger correlations compared to the other data in summer and shows some agreement with the observations in the industrial period since approximately the year 1850 (not shown). If we correlate the scaled isotope data to the temperature for an extended spring season from March to July, the correlations are quite similar to those for the larger domain simulation output but differ notably from the observations (not shown). The extended summer (MJJA) reconstruction by Rinne et al. (2013) agrees well with the respective observations in a consistently negative correlation (not shown). The relation is weaker for the reconstruction prior to the period of the Oxford precipitation observations (not shown).~~

~~Explanations for the different temperature-precipitation relations might be either~~

The differences between simulation and observations may imply either shortcomings of any of the observational data sets in the early period or that the simulation presents a too stable relationship. Explanations might be physical inconsistencies within the simulations~~or a lack of~~. More generally, any of the data sources may lack the physical relation between the temperature and precipitation records ~~. A third~~in the chosen season. Another possibility is that internal large-scale climate factors influencing the relation between both parameters evolve differently in simulation and reality. ~~This implies a dominant influence of internal variability on the considered regional and temporal scales which we discuss in the next sub-section. Even though reconstructions and observations represent different regional domains~~Assuming that the observations are the more reliable data set, we tend to the inference that the disagreement between ~~the~~ observations and reconstructions ~~suggests~~ suggest major shortcomings in the reconstructions~~, if we assume the observations to be the more reliable data set~~.

5.2.1 ~~Internal vs. forced variability~~

5.3 Internal vs. forced variability

If we ~~look for~~ expect temporal consistency among ~~different source of informations, we assume that all~~ the different sources of information, then we are assuming that all the sources of information ~~reflect similar impact of the~~ are responding to the impact of external climate forcing~~. Moreover, we then also believe~~ , and that the regional simulation ~~to be skillful in representing~~ skillfully represents the climate response to these conditions. ~~We also have to be aware that~~ Nevertheless, internal climate variability may dominate even for large exogenous forcing (compare, e.g. Deser et al., 2012a). We ~~can frame this as the question to what extend one can expect simulations and observation based data sets to reflect consistently these exogenous influences.~~

~~We assume that the transformation to distributional properties smooths out some of the temporal and structural differences from the different evolutions of internal variability expected between simulations and observational data. However, influences from low frequent climate modes may still have different phases in different data sets. In this sense, it is encouraging that the regional simulation shows some commonalities with the observed statistics.~~ have to ask, what is our expectation of consistency between simulated and observed responses to exogenous influences?

The instrumental period overlaps with the industrial period of large anthropogenic climate forcing. Earlier ~~in our period of interest~~ exogenous forcing is potentially weak ~~. However the period includes periods of~~ despite relatively large variations in solar activity ~~like the late Maunder Minimum (~1645 to ~1715 CE), the Dalton Minimum in the early 19th century, and~~

a period with relatively strong solar activity inbetween as indicated by sunspot numbers (Clette et al., 2014). Furthermore, (Clette et al., 2014), and the occurrence of a number of strong tropical volcanic eruptions occurred during this period , i.e. in ~1809 CE (unknown location), 1815 CE (Tambora), and 1835 CE (Cosigüina) during the period of interest (e.g., Schmidt et al., 2011).

Fischer et al. (2014) show that forced Forced precipitation signals can agree in simulations, e.g., the CMIP5 21st century global projections . The lack of consistent relations between different data sets under purely externally naturally forced and internal variability on multi-decadal time-scales questions our ability to make dynamical inferences about hydroclimate variability of small regions.

(Fischer et al., 2014). A lack of an identifiable relation relationship to the forcing between different data sources in our study does not necessarily imply that the underlying climate data are wrong but may simply suggest that internalnatural variability dominates, e.g., oceanic, atmospheric, or coupled climate variability mask, modulate, or counteract masks, modulates, or counteracts an external forcing influence. That is, the lack of consistent evolutions points to shortcomings of the data sources or an overwhelming influence of internal variability. The sporadic opposite behavior make the first more likely without negating the latter. That is, we interpret the opposite behavior as reactions to the forcing but different reactions of simulated and observed climate.

In this context, we We have to emphasize that the regional simulation and its driving MPI-ESM-COSMOS simulation both use variations of the total solar irradiance forcing that could be unrealistically wide. Furthermore, , and neither simulation includes a resolved stratosphere to account for potential UV-related top-down mechanisms (Anet et al., 2013, 2014).

Furthermore, since In addition, our regional focus is close to the western boundary of the domain of the regional simulation, and, thus, we expect a rather strong influence of the dynamical evolution of the driving coarse-resolution simulation with MPI-ESM-COSMOS. Indeed, Blenkinsop and Fowler (2007) report a strong influence of the driving general circulation model on the representation of drought in regional climate simulations in southern Great Britain.

Thus, while the regional simulation appears to present similar variations compared to the observations during some periods, it is unclear whether it does so for the right reasons.

5.3.1 Relation to dynamics

Our Relatedly, since the regional focus is a small domain. Thus, we should not expect simulations to agree with observations on the evolution of regional climate parameters and even an ensemble may show diverse behavior since , the influence of natural internal variability is likely large, e.g. in the case of the British Isles, the North Atlantic Oscillation (Gómez-Navarro et al., 2012; Gómez-Navarro and Zorita, 2013). **?** note the general importance of the storm track over the North Atlantic as a control on precipitation variability, and **?** show this for the England-Wales summer season precipitation. Hall and Hanna (2018) study to what extent North Atlantic circulation indices explain precipitation in the United Kingdom. They note a negative correlation of summer precipitation with indices representing jet-latitudes, which include the NAO. **?** detail the large-scale influences, e.g., the wave-train pattern on the jet stream, on the flooding events in the UK in 2007. Earlier, **?** noted the strong relation between the England-Wales precipitation and Lamb's cyclonic British Isles weather type (compare, e.g., ?) in spring, while recently Matthews et al. (2016) emphasize the importance of the high-frequent weather variability, i.e. cyclones, for seasonal precipitation amounts over the British Isles and particularly the summer season.

We note that Cooper et al. (2013) do not find any significant influence of the North Atlantic Oscillation (NAO) on precipitation or tree growth in East Anglia over the 20th century. They use the NAO as a measure of large scale influences on western European precipitation coherence.

Thus, internal climate variability is an integral expression of the circulation over the North Atlantic region (Gómez-Navarro et al., 2012; Gómez-Navarro and Zorita, 2013; Hall and Hanna, 2018; Matthews et al., 2016). Thus, we should not expect simulations to agree with observations on the evolution of regional climate parameters and even an ensemble may show diverse behavior. Differences in internal variability between models, observations, and paleo-observations may also include instabilities of dominant large scale patterns. That is, we cannot reject the idea that include their representation of past changes in the relationship between the regional climate and the large-scale circulation changed in the past. Lehner et al. (2012) describe the importance of such changes for inferring past states of the North Atlantic Oscillation from sparse proxy data. The importance of changes in the large-scale circulation becomes even more clear when considering the stability in centers of action in the North Atlantic sector over longer time-scales (Pinto and Raible, 2012; Raible et al., 2014)(Pinto and Raible, 2012; Lehner et al., 2012; Raible et al., 2014).

That isThus, while the forcing history suggests notable variationsand , and the large-scale temperature records indicate an imprint of the forcing history on hemispheric and global temperatures, internal variability may dominate on smaller regional scales (e.g., Deser et al., 2012b). This is despite the fact that, e.g., the large scale storm track is indeed sensitive to solar (e.g., Ineson et al., 2015) and volcanic forcing (e.g., Fischer et al., 2007; Trouet et al., 2018). Considering the possibly large role of internal variability on regional scales and the limitations of simulations in representing regional scale precipitation, the occasionally consistent variations in precipitation distribution properties increase our confidence in forced changes.

5.3.1 Concluding remarks

In summarising, for the chosen regional domains, we do not find consistency among the various data sets. However, each of these data sets is associated with its own uncertainties, which put various caveats on the interpretation of the lacking consistency and its sources. Encouragingly simulations and observations appear to agree on certain features occasionally but maybe for different However, while the regional simulation appears to present similar variations compared to the observations during some periods, we cannot say whether it does so for the right reasons.

6 Conclusions

This study pursued two goals. For one, we wanted to show that comparing precipitation in reconstructions, climate model simulations, and observations based on the Standardized Precipitation Index (SPI) over moving windows allows for helps in the rigorous comparison of these data sets and extends the common set of tools for such analysesdifferent sources of precipitation information over paleoclimatic time-scales. The information on precipitation distributions obtained by the SPI-approach eases comparing how different sources of information represent climatologies of precipitation. Second, by using this approach, we studied the consistency of the various sources of information for precipitation variations in a small regional domain ~~. We chose a domain~~ in southern Great Britain~~and compared long-term trends, decadal variability, and the probability distributions for the period since approximately 1650 CE~~.

~~Fitting distributions over moving windows provides the opportunity to compare how the different sources of information represent various percentiles and moments of the distributions over time in the presence of varying external forcings. It further allows to compare which percentile a reference precipitation amount represents over time; more loosely spoken, one compares how the probability of a reference precipitation amount changes over time.~~

~~For~~ Regarding the results for our specific study domain, first we did not find any clear ~~common~~ consistency for precipitation signals among a regional climate model simulation, an observational data set, and two local domain reconstructions. ~~The regional simulation shows only limited agreement with its observational target but less so with the reconstructions. The~~ We conclude that the considered reconstructions indeed appear to be unreliable representations of the observational series. ~~Relations between temperature and precipitation share some common co-variance on interannual time scales between the sources of information~~

Second, the regional simulation shows limited agreement with its observational target, the observational England-Wales precipitation data. Particularly the variability in both data sources shows comparable changes for the full period of the observations. This is possibly mainly due to comparable changes in dryness, which also show some level of agreement over the full period. This partial agreement between variability and dryness of the regional simulation and observations is encouraging. However, considering all associated uncertainties, we conclude that it may be due to different processes in the respective data sources.

~~A further~~ Third, the simulation data does not agree with the reconstructions. Nevertheless, an interesting result is the at times opposite evolution of the reconstructions and the regional simulations considering regional dryness and wetness~~. However~~, e.g., between 1750 and 1850. Again, considering all sources of uncertainty, we cannot attribute ~~it~~ this to the external forcing or to errors in either data source. ~~The partial agreement between variability and dryness of the regional simulation and observations is encouraging but may be due to different processes in the respective data sources~~

Fourth, our data sources do not agree on the strength of the relation between temperature and precipitation. However, the relations between both parameters share some common co-variance on interannual time scales between the sources of information for the season from March to July, e.g. in the 19th century.

Generally, a dominant role of internal variability could explain the lack of consistency in standardised precipitation measures in the different data sets on the temporal and spatial scales we consider here; the relative role of the external climate forcing generally becomes weaker at smaller spatial and shorter temporal scales (Deser et al., 2012b). The lack of general consistency and slightly differing interannual relations between temperature and precipitation still require a closer look at the uncertainties of observations, the methods and input data of reconstructions, and dynamical and thermodynamical representations of regional climate in regional ~~and global~~ simulations.

A supplementary ~~asset~~ document for this manuscript will be deposited at https://osf.io/duyqe/.

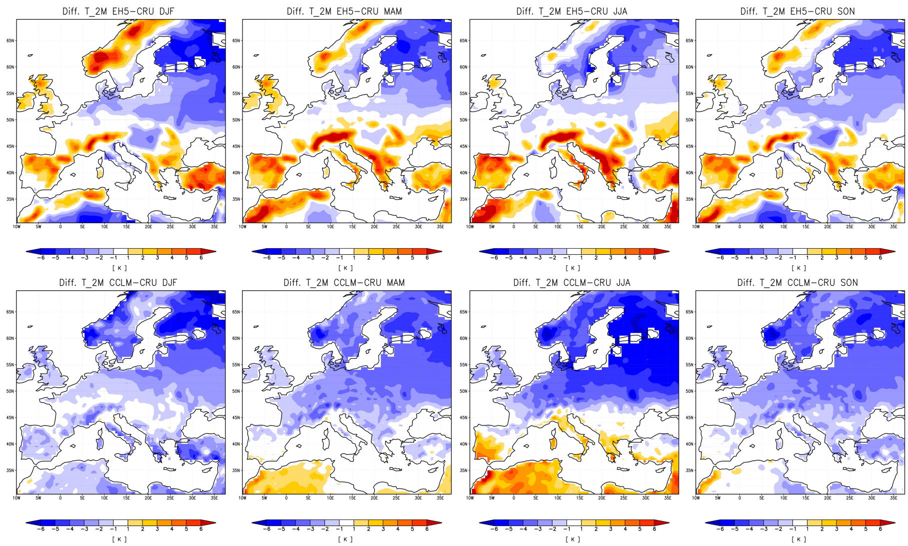

Figure A1. Top: Difference between the driving MPI-ESM simulation and the CRU data for seasonal near surface air temperature. Bottom: Difference for CCLM

Data availability. The Central England Temperature data is available from the Met Office, https://www.metoffice.gov.uk/hadobs/hadcet/.

The England-Wales Precipitation data is available from the Met Office, https://www.metoffice.gov.uk/hadobs/hadukp/ as are the subdivisions for South-East, South-West, and Central England.

Station data for Oxford, Kew Gardens and Pode Hole is available at, e.g., the Climate Explorer (http://climexp.knmi.nl/) of the Koninklijk Nederlands Meteorologisch Instituut (KNMI).

The reconstruction data for Southern-Central England and East Anglia are available from the NOAA National Centers for Environmental Information at, respectively, https://www.ncdc.noaa.gov/paleo-search/study/12907 and https://www.ncdc.noaa.gov/paleo-search/study/12896.

Temperature and precipitation fields from the regional simulation with CCLM are available at http://doi.org/10.6084/m9.figshare.5952025 (PRIME2, 2018).

If deemed relevant for future work, we are going to provide the standardised data as well via a public repository.

Considering the data used in the supplementary document, we are unable to provide the data by Rinne et al. (2013) as we only obtained it from the original author. The $\delta^{18}O$ data from Young et al. (2015) is available from https://link.springer.com/article/10.1007%2Fs00382-015-2559-4, if the editor deems it necessary we are going to provide the scaled series produced from this input.

Appendix A: Evaluation of the simulation setup against the CRU-data

We shortly describe the performance of the COSMOS-MPI-ESM-CCLM-setup compared to the observational CRU-data (Harris et al., 2014; University of East Anglia Climatic Research Unit et al., 2017). We used version CRU TS 3.10, which has subsequently been superseded. The current version CRU TS 4.01 is available at http://doi.org/10/gcmcz3 with further information also given at https://crudata.uea.ac.uk/cru/data/hrg/ (last visited 20 September 2018).

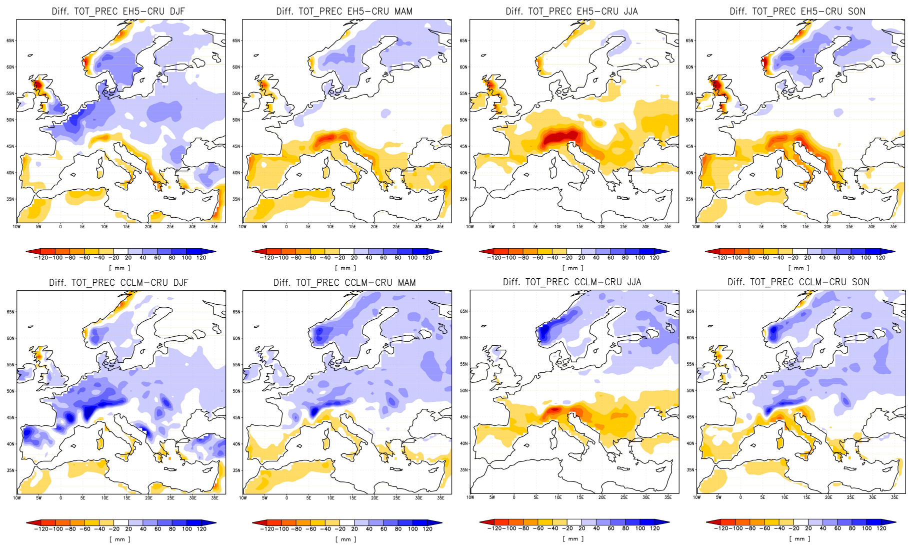

Figure A2. As Figure A1 but for the precipitation

The mean climate of the driving COSMOS-MPI-ESM simulation is too warm for much of the British Isles, the Scandinavian Alps, northern North Africa, Iberia, the Alps, southern France, Turkey, and Greece for all seasons over the period 1951-2000 (Figure A1, top). It is generally too cold over the Baltic region, the eastern part of the model domain, the southern border of the domain over Africa, and central Europe. High elevation and southern area warm biases frequently exceed 6K. Cold biases exceed 2 to 4K occasionally over northeastern Europe and at the southern border of the domain. We attribute these biases to some extent to the cruder representation of the European orography and, possibly related to that, to biases in the modelled atmospheric circulation. However, the specific choice of forcings may also influence the climatology.

In the regional CCLM simulation (Figure A1, bottom), warm biases for 1951-2000 are confined to the Atlas Mountains in all seasons and to the South of the domain in spring and summer. Cold biases are common otherwise and are largest over the Northeast frequently exceeding 3-4K.

For precipitation, summer is frequently too dry in central Europe in COSMOS-MPI-ESM and especially at the west coast of Scotland and in the Alps (Figure A2, top row). The southern domain is generally too dry in spring when Scandinavia is slightly too wet. Coastal and mountainous regions as well as Iberia, Italy, and southern France are more likely to be too dry in autumn and winter. Scandinavia is also too wet in autumn. The COSMOS-MPI-ESM winter climatology is too wet over much of central, eastern, and northern Europe.

In CCLM, too dry conditions are generally confined to southern Europe and North Africa and areas affected by the storm track, i.e. the coasts of Scotland and Norway (Figure A2, bottom row). They extend to southern central Europe only in summer. The climate is too wet in Scandinavia and northeastern Europe in most seasons. Large parts of Europe are too wet in all seasons except summer. Noteworthy is the excess precipitation at the northern flank of the Alps from autumn to spring. Part of these discrepancies are possibly attributable to a too zonal airflow outside the summer season.

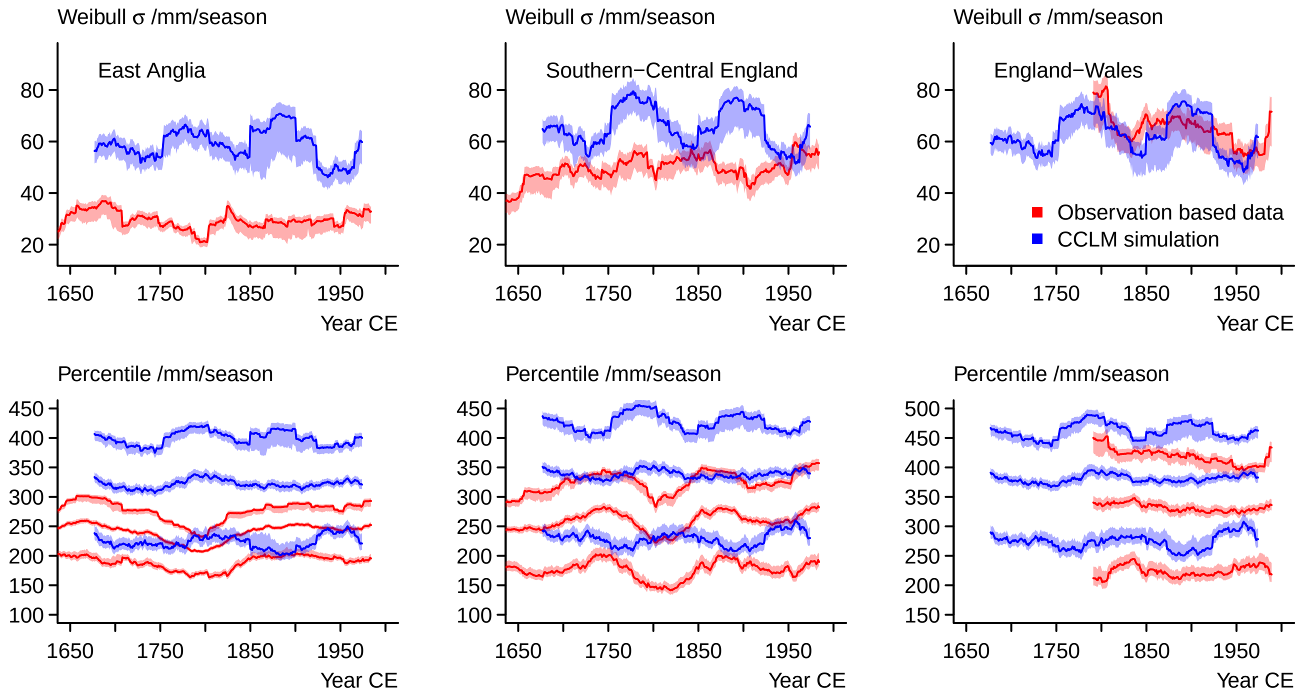

Figure B1. Visualisation of uncertainty of the distributional properties. We use a bootstrap procedure on running estimates. We resample 40 year samples a thousand times from moving 51-year windows. Units are precipitation amounts. Shading are 95% intervals, lines are medians. Top row: Weibull standard deviation. Bottom row: 93.3th, 50th, and 6.7th percentiles. Red: Reconstruction and observations. Blue: CCLM. The left column is for East Anglia, the middle column for Southern-Central England, and the right column for the England-Wales precipitation.

In summarizing, the model presents a too strong latitudinal temperature gradient over the European domain. The annual cycle of temperature is apparently too strong in the South with warm biases in summer but cold biases in winter and it is slightly too weak in the North with cold biases being stronger in summer than in winter. Similarly to temperature, the gradient in precipitation also appears to be too strong and the annual cycle amplitude differs between simulation and gridded observational estimates especially for Central Europe. Specifically, autumn to spring are wetter in the simulation while summer conditions differ only slightly or are too dry, which implies a weaker annual cycle compared to observations.

Appendix B: Uncertainty of running measures

Figure B1 shows bootstrap estimates over thousand 40-year samples for each 51-year window. The estimates are for the running measures for reconstructions and observations for the three regions of interest (red) and the regional simulation (blue). The top row are Weibull standard deviations and the bottom row is for the percentiles.

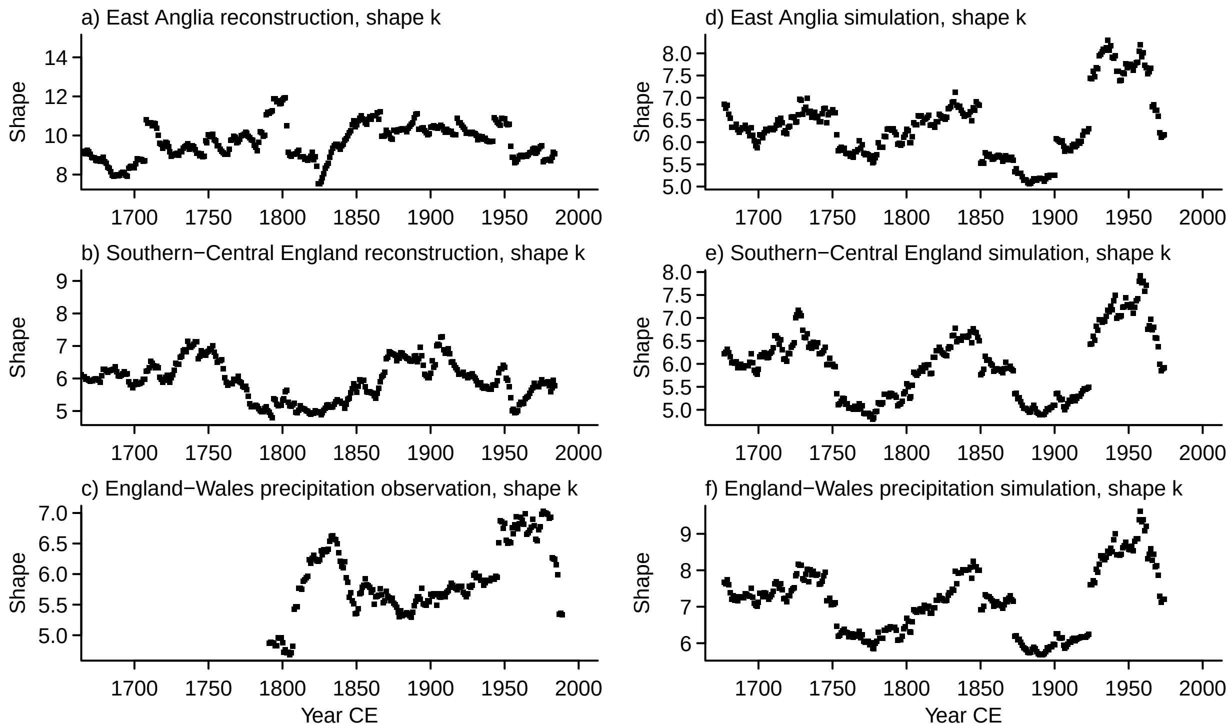

Figure C1. Evolution of the shape parameter k for the Weibull distribution fits for the a) East Anglia reconstruction, b) Southern-Central England reconstruction, c) England-Wales precipitation observational data, d) East Anglia regional simulation, e) Southern-Central England regional simulation, f) England-Wales precipitation regional simulation.

The Figure highlights that sampling variability is generally larger for the simulated data. Indeed sampling variability may render differences between periods non-significant. However, also the bootstrap distributions appear strongly skewed.

Appendix C: Distributional parameters

The Weibull distribution is a two parameter distribution with a scale and a shape parameter. See, e.g., Sienz et al. (2012), for more details and how the distribution compares to other distributions in computing the Standardised Precipitation Index.

Figures C1 and C2 present the shape, k, and scale, λ, parameters of our Weibull distribution fits for the reconstructions for East Anglia and Southern-Central England, the observational England-Wales precipitation, and the respective time series in the simulation.

Results for the simulation show very similar evolutions among regions highlighting the homogeneity of the simulation data. There are also similarities between the two reconstructions. One could argue the shape parameters evolve similarly in observation and simulation.

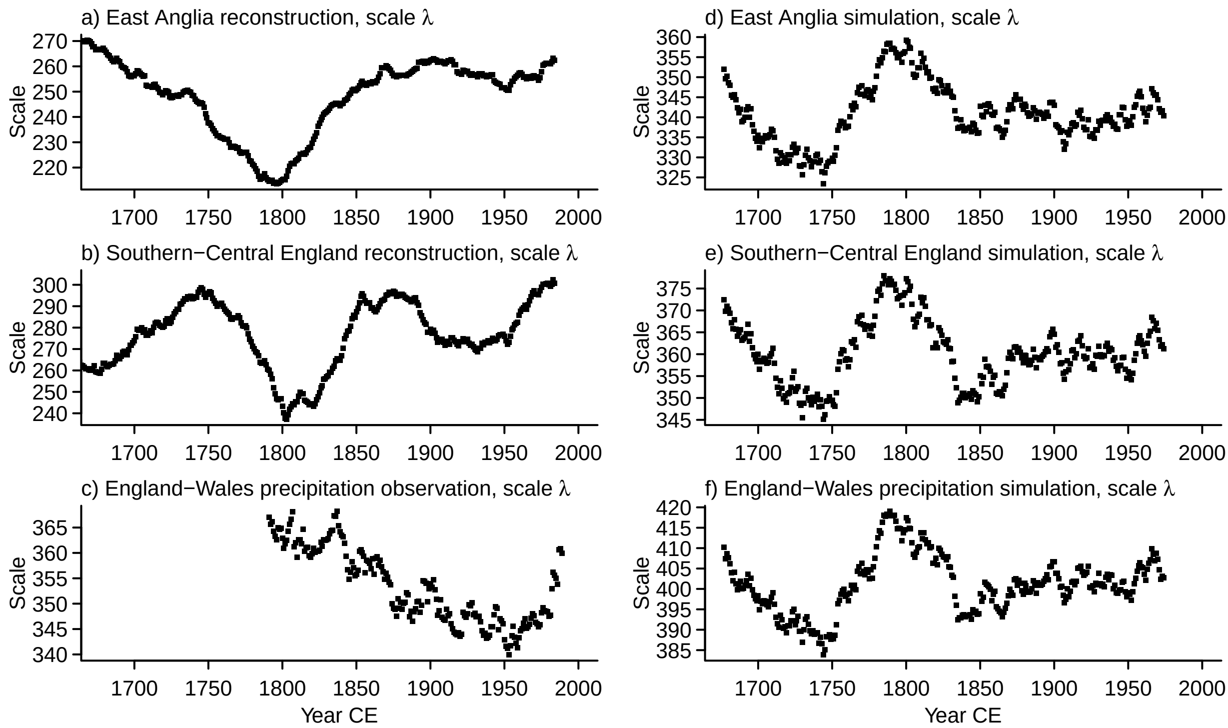

Figure C2. Evolution of the scale parameter λ for the Weibull distribution fits for the a) East Anglia reconstruction, b) Southern-Central England reconstruction, c) England-Wales precipitation observational data, d) East Anglia regional simulation, e) Southern-Central England regional simulation, f) England-Wales precipitation regional simulation.

The shape parameter determines the 'shape' of the distribution. In our cases, changes in this parameter are rather small (compare Figure C1). Nevertheless they can result in notably different widths of distributions for a specific data set over time. It is interesting that there is only small overlap between the range of shape parameters for the East Anglia reconstruction and all other series.

Larger scale parameters for a constant shape parameter result in a flatter distribution that extends further to larger values. Smaller values result in a narrower distribution with larger probability density at its peak.

The evolution of the shape parameter reflects, in our cases, the evolution of the skewness of the distributions (not shown). All distributions show negative skewness, and the amplitude increases with increases in the shape parameter.

Figure C3 shows the excess kurtosis over the period of interest. The most common feature for the different records is a negative excess kurtosis. Interestingly, the East Anglia reconstructions shows large positive values. The simulation data has a period with positive, or for the simulated England-Wales precipitation larger positive, values in the middle of the 20th century, and the observed England-Wales precipitation shows only negative excess kurtosis. The scaling of the kurtosis-axes for the reconstructions highlights that they show much larger values earlier in the last millennium (not shown, compare the supplementary ~~manuscript asset~~document).

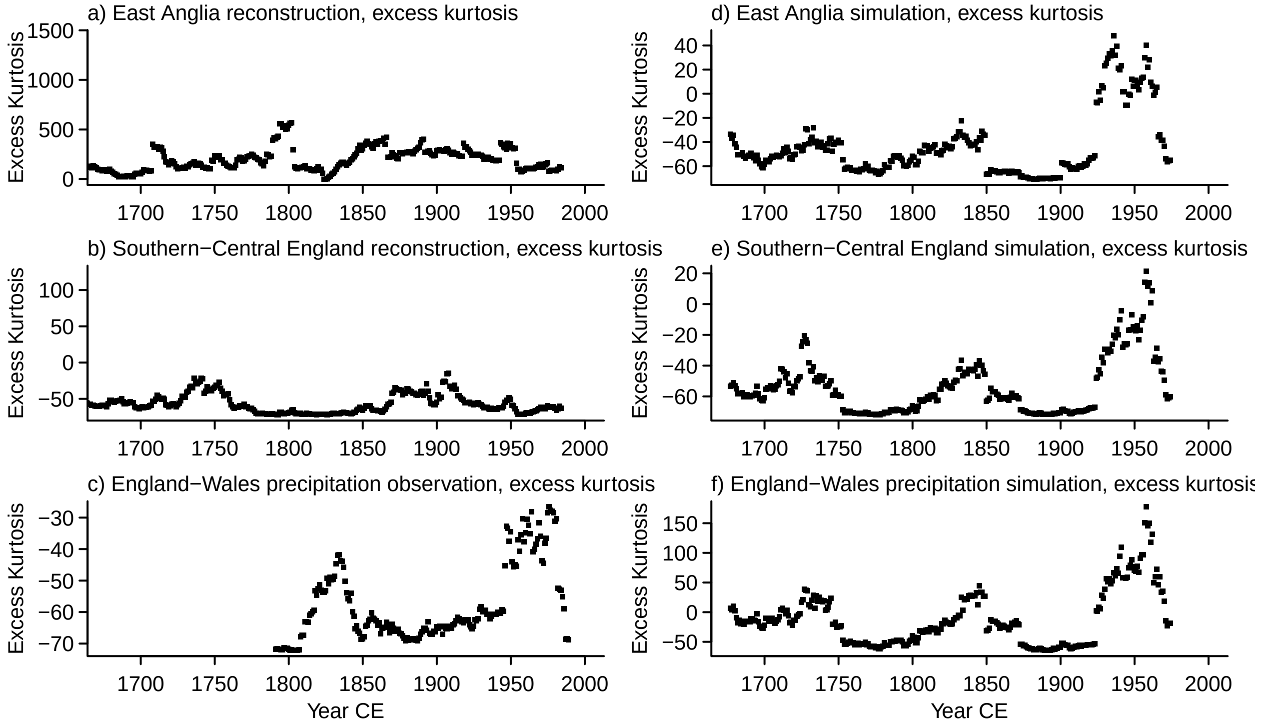

Figure C3. Evolution of the excess kurtosis of the fitted Weibull distributions for the a) East Anglia reconstruction, b) Southern-Central England reconstruction, c) England-Wales precipitation observational data, d) East Anglia regional simulation, e) Southern-Central England regional simulation, f) England-Wales precipitation regional simulation.

Appendix D: External code

This manuscript uses a number of external software-packages. File-manipulations used the Climate Data Operators (cdo, https://code.mpimet.mpg.de/projects/cdo/). Furthermore, the following R (R Core Team, 2018) packages helped in the work: gtools (Warnes et al., 2018), corrplot (Wei and Simko, 2017), ncdf (Pierce, 2015), VGAM (Yee, 2015), MASS (Venables and Ripley, 2002), nortest (Gross and Ligges, 2015), dplR (Bunn et al., 2018), zoo (Zeileis and Grothendieck, 2005), latex2exp (Meschiari, 2015), knitr (Xie, 2015), and rmarkdown (Allaire et al., 2018). Furthermore, RStudio (RStudio Team, 2016) was essential. The manuscript was prepared using the rticles-package (no reference available).

The SPI-code bases on work by Frank Sienz (e.g., Sienz et al., 2012). Christian Zang provided a Gershunov-bootstrap procedure (compare, e.g., Gershunov et al., 2001; Zang and Biondi, 2015) that we modified.

Competing interests. The authors are not aware of any circumstances that might be seen as competing interests.

Acknowledgements. Funding in the projects PRIME2 and PALMOD (www.palmod.de) made the completion of this study possible. This study is a contribution to PALMOD, and to the PAGES 2k Network, especially its PALEOLINK project. We acknowledge the service of the Met Office Hadley Centre for Climate Change for providing the Central England Temperature and England-Wales Precipitation under the Open Government Licence (http://www.nationalarchives.gov.uk/doc/open-government-licence/version/2/), and of the NOAA Centers for Environmental Information for providing the reconstruction data by Cooper et al. (2013) and Wilson et al. (2013). We acknowledge the SPI-code provided by Frank Sienz (e.g., Sienz et al., 2012). Christian Zang provided input for a computationally efficient Gershunov-test.

References

Alexander, L. V. and Jones, P. D.: Updated precipitation series for the U.K. and discussion of recent extremes, Atmosph. Sci. Lett., 1, 142–150, https://doi.org/10.1006/asle.2000.0016, , 2000.

Allaire, J., Xie, Y., McPherson, J., Luraschi, J., Ushey, K., Atkins, A., Wickham, H., Cheng, J., and Chang, W.: rmarkdown: Dynamic Documents for R, https://CRAN.R-project.org/package=rmarkdown, r package version 1.10, 2018.

Anet, J. G., Muthers, S., Rozanov, E., Raible, C. C., Peter, T., Stenke, A., Shapiro, A. I., Beer, J., Steinhilber, F., Brönnimann, S., Arfeuille, F., Brugnara, Y., and Schmutz, W.: Forcing of stratospheric chemistry and dynamics during the Dalton Minimum, Atmospheric Chemistry and Physics, 13, 10 951–10 967, https://doi.org/10.5194/acp-13-10951-2013, , 2013.

Anet, J. G., Muthers, S., Rozanov, E. V., Raible, C. C., Stenke, A., Shapiro, A. I., Brönnimann, S., Arfeuille, F., Brugnara, Y., Beer, J., Steinhilber, F., Schmutz, W., and Peter, T.: Impact of solar versus volcanic activity variations on tropospheric temperatures and precipitation during the Dalton Minimum, Climate of the Past, 10, 921–938, https://doi.org/10.5194/cp-10-921-2014, , 2014.

Annan, J. D. and Hargreaves, J. C.: Understanding the CMIP3 Multimodel Ensemble, J. Climate, 24, 4529–4538, https://doi.org/10.1175/2011jcli3873.1, , 2011.

Bard, E., Raisbeck, G., Yiou, F., and Jouzel, J.: Solar irradiance during the last 1200 years based on cosmogenic nuclides, Tellus B, 52, 985–992, https://doi.org/10.1034/j.1600-0889.2000.d01-7.x, , 2000.

~~Bengtsson, L., Hodges, K. I., Roeckner, E., Bengtsson, L., Hodges, K. I., and Roeckner, E.: Storm Tracks and Climate Change, Journal of Climate, 19, 3518–3543, , , 2006.~~

Bierstedt, S. E., Hünicke, B., Zorita, E., Wagner, S., and Gómez-Navarro, J. J.: Variability of daily winter wind speed distribution over Northern Europe during the past millennium in regional and global climate simulations, Climate of the Past, 12, 317–338, https://doi.org/10.5194/cp-12-317-2016, , 2016.

~~Blackburn, M., Methven, J., and Roberts, N.: Large-scale context for the UK floods in summer 2007, Weather, 63, 280–288, , , 2008.~~

Blenkinsop, S. and Fowler, H.: Changes in drought frequency, severity and duration for the British Isles projected by the PRUDENCE regional climate models, Journal of Hydrology, 342, 50–71, https://doi.org/10.1016/J.JHYDROL.2007.05.003, , 2007.

Böhm, R., Jones, P. D., Hiebl, J., Frank, D., Brunetti, M., and Maugeri, M.: The early instrumental warm-bias: a solution for long central European temperature series 1760–2007, Climatic Change, 101, 41–67, https://doi.org/10.1007/s10584-009-9649-4, , 2010.

Bunn, A., Korpela, M., Biondi, F., Campelo, F., Mérian, P., Qeadan, F., Zang, C., Pucha-Cofrep, D., and Wernicke, J.: dplR: Dendrochronology Program Library in R, https://CRAN.R-project.org/package=dplR, r package version 1.6.8, 2018.

Burt, T. P. and Howden, N. J.: A homogenous daily rainfall record for the Radcliffe Observatory, Oxford, from the 1820s, Water Resources Research, 47, https://doi.org/10.1029/2010WR010336, 2011.

Casty, C., Raible, C., Stocker, T., Wanner, H., and Luterbacher, J.: A European pattern climatology 1766–2000, Climate Dynamics, 29, 791–805, https://doi.org/10.1007/s00382-007-0257-6, , 2007.

Clette, F., Svalgaard, L., Vaquero, J., and Cliver, E.: Revisiting the Sunspot Number, Space Science Reviews, pp. 1–69, https://doi.org/10.1007/s11214-014-0074-2, , 2014.

~~Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Matsui, N., Allan, R. J., Yin, X., Gleason, B. E., Vose, R. S., Rutledge, G., Bessemoulin, P., Brönnimann, S., Brunet, M., Crouthamel, R. I., Grant, A. N., Groisman, P. Y., Jones, P. D., Kruk, M. C., Kruger, A. C., Marshall, G. J., Maugeri, M., Mok, H. Y., Nordli, Ø., Ross, T. F., Trigo, R. M., Wang, X. L., Woodruff, S. D., and Worley, S. J.: The Twentieth Century Reanalysis Project, Quarterly Journal of the Royal Meteorological Society, 137, 1–28, , , 2011.~~

~~Cook, B. I., Mankin, J. S., and Anchukaitis, K. J.: Climate Change and Drought: From Past to Future, Current Climate Change Reports, 4, 164–179, , , 2018.~~

Cook, E. R., Seager, R., Kushnir, Y., Briffa, K. R., Bu ntgen, U., Frank, D., Krusic, P. J., Tegel, W., van der Schrier, G., Andreu-Hayles, L., Baillie, M., Baittinger, C., Bleicher, N., Bonde, N., Brown, D., Carrer, M., Cooper, R., Cufar, K., Dittmar, C., Esper, J., Griggs, C., Gunnarson, B., Gu nther, B., Gutierrez, E., Haneca, K., Helama, S., Herzig, F., Heussner, K.-U., Hofmann, J., Janda, P., Kontic, R., Ko se, N., Kyncl, T., Levanic, T., Linderholm, H., Manning, S., Melvin, T. M., Miles, D., Neuwirth, B., Nicolussi, K., Nola, P., Panayotov, M., Popa, I., Rothe, A., Seftigen, K., Seim, A., Svarva, H., Svoboda, M., Thun, T., Timonen, M., Touchan, R., Trotsiuk, V., Trouet, V., Walder, F., Wazny, T., Wilson, R., and Zang, C.: Old World megadroughts and pluvials during the Common Era, Science Advances, 1, e1500 561–e1500 561, https://doi.org/10.1126/sciadv.1500561, ~~,~~ 2015.

Cooper, R., Melvin, T., Tyers, I., Wilson, R., and Briffa, K.: A tree-ring reconstruction of East Anglian (UK) hydroclimate variability over the last millennium, Climate Dynamics, 40, 1019–1039, https://doi.org/10.1007/s00382-012-1328-x, ~~,~~ 2013.

Craddock, J. M. and Craddock, E.: Rainfall at Oxford from 1767 to 1814, estimated from records of Thomas Hornsby and others, The Meteorological Magazine, 106, 361–372, 1977.

Crhová, L. and Holtanová, E.: Simulated relationship between air temperature and precipitation over Europe: sensitivity to the choice of RCM and GCM, International Journal of Climatology, 38, 1595–1604, https://doi.org/10.1002/joc.5256, ~~,~~ 2018.

Croxton, P. J., Huber, K., Collinson, N., and Sparks, T. H.: How well do the central England temperature and the England and Wales precipitation series represent the climate of the UK?, International Journal of Climatology, 26, 2287–2292, https://doi.org/10.1002/joc.1378, ~~,~~ 2006.

Deser, C., Knutti, R., Solomon, S., and Phillips, A. S.: Communication of the role of natural variability in future North American climate, Nature Climate Change, 2, 775–779, https://doi.org/10.1038/nclimate1562, ~~,~~ 2012a.

Deser, C., Phillips, A., Bourdette, V., and Teng, H.: Uncertainty in climate change projections: the role of internal variability, Climate Dynamics, 38, 527–546, https://doi.org/10.1007/s00382-010-0977-x, ~~,~~ 2012b.

Domínguez-Castro, F., Santisteban, J. I., Barriendos, M., and Mediavilla, R.: Reconstruction of drought episodes for central Spain from rogation ceremonies recorded at the Toledo Cathedral from 1506 to 1900: A methodological approach, Global and Planetary Change, 63, 230–242, https://doi.org/10.1016/j.gloplacha.2008.06.002, ~~,~~ 2008.

~~Dong, B., Sutton, R. T., Woollings, T., and Hodges, K.: Variability of the North Atlantic summer storm track: mechanisms and impacts on European climate, Environmental Research Letters, 8, 034 037, , 2013.~~

Evans, M. N., Tolwinski-Ward, S. E., Thompson, D. M., and Anchukaitis, K. J.: Applications of proxy system modeling in high resolution paleoclimatology, Quaternary Science Reviews, 76, 16–28, https://doi.org/10.1016/j.quascirev.2013.05.024, ~~,~~ 2013.

Fernández-Donado, L., González-Rouco, J. F., Raible, C. C., Ammann, C. M., Barriopedro, D., García-Bustamante, E., Jungclaus, J. H., Lorenz, S. J., Luterbacher, J., Phipps, S. J., Servonnat, J., Swingedouw, D., Tett, S. F. B., Wagner, S., Yiou, P., and Zorita, E.: Large-scale temperature response to external forcing in simulations and reconstructions of the last millennium, Climate of the Past, 9, 393–421, https://doi.org/10.5194/cp-9-393-2013, ~~,~~ 2013.

Fischer, E. M., Luterbacher, J., Zorita, E., Tett, S. F. B., Casty, C., and Wanner, H.: European climate response to tropical volcanic eruptions over the last half millennium, Geophys. Res. Lett., 34, L05 707+, https://doi.org/10.1029/2006gl027992, ~~,~~ 2007.

Fischer, E. M., Sedláček, J., Hawkins, E., and Knutti, R.: Models agree on forced response pattern of precipitation and temperature extremes, Geophys. Res. Lett., pp. 2014GL062 018+, https://doi.org/10.1002/2014gl062018, ~~,~~ 2014.

Flato, G., Marotzke, J., Abiodun, B., Braconnot, P., Chou, S., Collins, W., Cox, P., Driouech, F., Emori, S., Eyring, V., Forest, C., Gleckler, P., Guilyardi, E., Jakob, C., Kattsov, V., Reason, C., and Rummukainen, M.: Evaluation of Climate Models, in: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, edited by Stocker, T., Qin, D., Plattner, G.-K., Tignor, M., Allen, S., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P., chap. 9, p. 741–866, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, https://doi.org/10.1017/CBO9781107415324.020, www.climatechange2013.org, 2013.

Flückiger, J., Monnin, E., Stauffer, B., Schwander, J., Stocker, T. F., Chappellaz, J., Raynaud, D., and Barnola, J.-M.: High-resolution Holocene N2O ice core record and its relationship with CH4 and CO2, Global Biogeochem. Cycles, 16, 10–10–8, https://doi.org/10.1029/2001gb001417, ~~,~~ 2002.

Frank, D., Büntgen, U., Böhm, R., Maugeri, M., and Esper, J.: Warmer early instrumental measurements versus colder reconstructed temperatures: shooting at a moving target, Quaternary Science Reviews, 26, 3298–3310, https://doi.org/10.1016/j.quascirev.2007.08.002, ~~,~~ 2007.

Franke, J., Brönnimann, S., Bhend, J., and Brugnara, Y.: A monthly global paleo-reanalysis of the atmosphere from 1600 to 2005 for studying past climatic variations, Scientific Data, 4, 170 076, https://doi.org/10.1038/sdata.2017.76, ~~,~~ 2017.

Gershunov, A., Schneider, N., and Barnett, T.: Low-Frequency Modulation of the ENSO–Indian Monsoon Rainfall Relationship: Signal or Noise?, J. Climate, 14, 2486–2492, https://doi.org/10.1175/1520-0442(2001)014%3C2486:lfmote%3E2.0.co;2, http://dx.doi.org/10.1175/1520-0442(2001)014%3C2486:lfmote%3E2.0.co;2, 2001.

Gómez-Navarro, J., Werner, J., Wagner, S., Luterbacher, J., and Zorita, E.: Establishing the skill of climate field reconstruction techniques for precipitation with pseudoproxy experiments, Climate Dynamics, pp. 1–19, https://doi.org/10.1007/s00382-014-2388-x, ~~,~~ 2014.

Gómez-Navarro, J. J. and Zorita, E.: Atmospheric annular modes in simulations over the past millennium: No long-term response to external forcing, Geophysical Research Letters, 40, 3232–3236, https://doi.org/10.1002/grl.50628, ~~,~~ 2013.

Gómez-Navarro, J. J., Montávez, J. P., Jiménez-Guerrero, P., Jerez, S., Lorente-Plazas, R., González-Rouco, J. F., and Zorita, E.: Internal and external variability in regional simulations of the Iberian Peninsula climate over the last millennium, Climate of the Past, 8, 25–36, https://doi.org/10.5194/cp-8-25-2012, http://dx.doi.org/10.5194/cp-8-25-2012, 2012.

~~Gómez-Navarro, J. J., Montávez, J. P., Wagner, S., and Zorita, E.: Aregional climate palaeosimulation for Europe in the period 1500–1990 – Part 1: Model validation, Climate of the Past, 9, 1667–1682, , , 2013.~~

Gómez-Navarro, J. J., Bothe, O., Wagner, S., Zorita, E., Werner, J. P., Luterbacher, J., Raible, C. C., and Montávez, J. P.: A regional climate palaeosimulation for Europe in the period 1500–1990 – Part 2: Shortcomings and strengths of models and reconstructions, Climate of the Past, 11, 1077–1095, https://doi.org/10.5194/cp-11-1077-2015, ~~,~~ 2015.

Gross, J. and Ligges, U.: nortest: Tests for Normality, https://CRAN.R-project.org/package=nortest, r package version 1.0-4, 2015.

Guttman, N. B.: On the Sensitivity of Sample L Moments to Sample Size, J. Climate, 7, 1026–1029, https://doi.org/10.1175/1520-0442(1994)007%3C1026:otsosl%3E2.0.co;2, ~~,~~ 1994.

~~Hakim, G. J., Emile-Geay, J., Steig, E. J., Noone, D., Anderson, D. M., Tardif, R., Steiger, N., and Perkins, W. A.: The last millennium climate reanalysis project: Framework and first results, Journal of Geophysical Research: Atmospheres, 121, 6745–6764, , , 2016.~~

Hall, R. J. and Hanna, E.: North Atlantic circulation indices: links with summer and winter UK temperature and precipitation and implications for seasonal forecasting, International Journal of Climatology, 38, e660–e677, https://doi.org/10.1002/joc.5398, ~~,~~ 2018.

Harris, I., Jones, P. D., Osborn, T. J., and Lister, D. H.: Updated high-resolution grids of monthly climatic observations – the CRU TS3.10 Dataset, Int. J. Climatol., 34, 623–642, https://doi.org/10.1002/joc.3711, ~~,~~ 2014.

Hayes, M., Svoboda, M., Wall, N., Widhalm, M., Hayes, M., Svoboda, M., Wall, N., and Widhalm, M.: The Lincoln Declaration on Drought Indices: Universal Meteorological Drought Index Recommended, Bulletin of the American Meteorological Society, 92, 485–488, https://doi.org/10.1175/2010BAMS3103.1, ~~,~~2011.

Hoerling, M., Eischeid, J., and Perlwitz, J.: Regional Precipitation Trends: Distinguishing Natural Variability from Anthropogenic Forcing, J. Climate, 23, 2131–2145, https://doi.org/10.1175/2009jcli3420.1, ~~,~~2009.

Iles, C. E., Hegerl, G. C., Schurer, A. P., and Zhang, X.: The effect of volcanic eruptions on global precipitation, J. Geophys. Res. Atmos., 118, 8770–8786, https://doi.org/10.1002/jgrd.50678, ~~,~~2013.

Ineson, S., Maycock, A. C., Gray, L. J., Scaife, A. A., Dunstone, N. J., Harder, J. W., Knight, J. R., Lockwood, M., Manners, J. C., and Wood, R. A.: Regional climate impacts of a possible future grand solar minimum, Nature Communications, 6, 7535+, https://doi.org/10.1038/ncomms8535, ~~,~~2015.

Jungclaus, J. H., Lorenz, S. J., Timmreck, C., Reick, C. H., Brovkin, V., Six, K., Segschneider, J., Giorgetta, M. A., Crowley, T. J., Pongratz, J., Krivova, N. A., Vieira, L. E., Solanki, S. K., Klocke, D., Botzet, M., Esch, M., Gayler, V., Haak, H., Raddatz, T. J., Roeckner, E., Schnur, R., Widmann, H., Claussen, M., Stevens, B., and Marotzke, J.: Climate and carbon-cycle variability over the last millennium, Climate of the Past, 6, 723–737, https://doi.org/10.5194/cp-6-723-2010, ~~,~~2010.

Keyantash, J., Dracup, J. A., Keyantash, J., and Dracup, J. A.: The Quantification of Drought: An Evaluation of Drought Indices, Bulletin of the American Meteorological Society, 83, 1167–1180, https://doi.org/10.1175/1520-0477-83.8.1167, ~~,~~2002.

~~Kington, J. A.: Daily weather mapping from 1781:, Climatic Change, 3, 7–36, , , 1990.~~

Klippel, L., Krusic, P. J., Brandes, R., Hartl, C., Belmecheri, S., Dienst, M., and Esper, J.: A 1286-year hydro-climate reconstruction for the Balkan Peninsula, Boreas, 47, 1218–1229, https://doi.org/10.1111/bor.12320, ~~,~~2018.

~~Küttel, M., Xoplaki, E., Gallego, D., Luterbacher, J., García-Herrera, R., Allan, R., Barriendos, M., Jones, P. D., Wheeler, D., and Wanner, H.: The importance of ship log data: reconstructing North Atlantic, European and Mediterranean sea level pressure fields back to 1750, Climate Dynamics, 34, 1115–1128, , , 2010.~~

~~Lamb, H. H.: Types and spells of weather around the year in the British Isles : Annual trends, seasonal structure of the year, singularities, Quarterly Journal of the Royal Meteorological Society, 76, 393–429, , , 1950.~~

Landrum, L., Otto-Bliesner, B. L., Wahl, E. R., Conley, A., Lawrence, P. J., Rosenbloom, N., and Teng, H.: Last Millennium Climate and Its Variability in CCSM4, J. Climate, 26, 1085–1111, https://doi.org/10.1175/jcli-d-11-00326.1, ~~,~~2012.

Lehner, F., Raible, C. C., and Stocker, T. F.: Testing the robustness of a precipitation proxy-based North Atlantic Oscillation reconstruction, Quaternary Science Reviews, 45, 85–94, https://doi.org/10.1016/j.quascirev.2012.04.025, http://dx.doi.org/10.1016/j.quascirev.2012.04.025, 2012.

Lehner, F., Joos, F., Raible, C. C., Mignot, J., Born, A., Keller, K. M., and Stocker, T. F.: Climate and carbon cycle dynamics in a CESM simulation from 850 to 2100 CE, Earth System Dynamics, 6, 411–434, https://doi.org/10.5194/esd-6-411-2015, ~~,~~2015.

Ludwig, P., Gómez-Navarro, J. J., Pinto, J. G., Raible, C. C., Wagner, S., and Zorita, E.: Perspectives of regional paleoclimate modeling, Annals of the New York Academy of Sciences, https://doi.org/10.1111/nyas.13865, ~~,~~2018.

Machado, M., Benito, G., Barriendos, M., and Rodrigo, F.: 500 Years of rainfall variability and extreme hydrological events in southeastern Spain drylands, Journal of Arid Environments, 75, 1244–1253, https://doi.org/10.1016/J.JARIDENV.2011.02.002, ~~,~~2011.

Matthews, T., Murphy, C., Wilby, R. L., and Harrigan, S.: A cyclone climatology of the British-Irish Isles 1871-2012, International Journal of Climatology, 36, 1299–1312, https://doi.org/10.1002/joc.4425, ~~,~~2016.

McKee, T. B., Doeskin, N. J., and Kleist, J.: The relationship of drought frequency and duration to time scales, in: 8th Conf. on Applied Climatology, pp. 179–184, Am. Meteorol. Soc., 1993.

Meschiari, S.: latex2exp: Use LaTeX Expressions in Plots, https://CRAN.R-project.org/package=latex2exp, r package version 0.4.0, 2015.

~~Murphy, C., Broderick, C., Burt, T. P., Curley, M., Duffy, C., Hall, J., Harrigan, S., Matthews, T. K. R., Macdonald, N., McCarthy, G., McCarthy, M. P., Mullan, D., Noone, S., Osborn, T. J., Ryan, C., Sweeney, J., Thorne, P. W., Walsh, S., and Wilby, R. L.~~

PAGES 2k-PMIP3 Group: ~~A 305-year continuous monthly rainfall series for the island of Ireland (1711–2016)~~Continental-scale temperature variability in PMIP3 simulations and PAGES 2k regional temperature reconstructions over the past millennium, Climate of the Past ~~, 14, 413–440, , , 2018.~~ Discussions, 11, 2483–2555, https://doi.org/10.5194/cpd-11-2483-2015, 2015.

PAGES Hydro2k Consortium: Comparing proxy and model estimates of hydroclimate variability and change over the Common Era, Climate of the Past, 13, 1851–1900, https://doi.org/10.5194/cp-13-1851-2017, https://www.clim-past.net/13/1851/2017/, 2017.

~~PAGES2k-PMIP3 Group: Continental-scale temperature variability in PMIP3 simulations and PAGES 2k regional temperature reconstructions over the past millennium, Climate of the Past Discussions, 11, 2483–2555, , , 2015.~~

Parker, D. E., Legg, T. P., and Folland, C. K.: A new daily central England temperature series, 1772–1991, Int. J. Climatol., 12, 317–342, https://doi.org/10.1002/joc.3370120402, ~~,~~ 1992.

Pauling, A., Luterbacher, J., Casty, C., and Wanner, H.: Five hundred years of gridded high-resolution precipitation reconstructions over Europe and the connection to large-scale circulation, Climate Dynamics, 26, 387–405, https://doi.org/10.1007/s00382-005-0090-8, ~~,~~ 2006.

Peterson, T. C. and Vose, R. S.: An Overview of the Global Historical Climatology Network Temperature Database, Bull. Amer. Meteor. Soc., 78, 2837–2849, https://doi.org/10.1175/1520-0477(1997)078%3C2837:aootgh%3E2.0.co;2, ~~,~~ 1997.

Pierce, D.: ncdf: Interface to Unidata netCDF Data Files, https://CRAN.R-project.org/package=ncdf, r package version 1.6.9, 2015.

Pinto, I., Jack, C., and Hewitson, B.: Process-based model evaluation and projections over southern Africa from Coordinated Regional Climate Downscaling Experiment and Coupled Model Intercomparison Project Phase 5 models, International Journal of Climatology, https://doi.org/10.1002/joc.5666, ~~,~~ 2018.

Pinto, J. G. and Raible, C. C.: Past and recent changes in the North Atlantic oscillation, WIREs Clim Change, 3, 79–90, https://doi.org/10.1002/wcc.150, ~~,~~ 2012.

Pongratz, J., Reick, C., Raddatz, T., and Claussen, M.: A reconstruction of global agricultural areas and land cover for the last millennium, Global Biogeochem. Cycles, 22, GB3018+, https://doi.org/10.1029/2007gb003153, ~~,~~ 2008.

PRIME2: Precipitation and Temperature from a Regional Climate Model simulation with CCLM for Europe over the period 1645-1999CE, https://doi.org/10.6084/m9.figshare.5952025, 2018.

R Core Team: R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, https://www.R-project.org/, 2018.

Raible, C. C., Lehner, F., González-Rouco, J. F., and Fernández-Donado, L.: Changing correlation structures of the Northern Hemisphere atmospheric circulation from 1000 to 2100 AD, Climate of the Past, 10, 537–550, https://doi.org/10.5194/cp-10-537-2014, ~~,~~ 2014.

Raible, C. C., Bärenbold, O., and Gómez-navarro, J. J.: Drought indices revisited – improving and testing of drought indices in a simulation of the last two millennia for Europe, Tellus A: Dynamic Meteorology and Oceanography, 69, 1287 492, https://doi.org/10.1080/16000870.2017.1296226, ~~,~~ 2017.

Rinne, K., Loader, N., Switsur, V., and Waterhouse, J.: 400-year May–August precipitation reconstruction for Southern England using oxygen isotopes in tree rings, Quaternary Science Reviews, 60, 13–25, https://doi.org/10.1016/j.quascirev.2012.10.048, 2013.

RStudio Team: RStudio: Integrated Development Environment for R, RStudio, Inc., Boston, MA, http://www.rstudio.com/, 2016.

Schmidt, G. A., Jungclaus, J. H., Ammann, C. M., Bard, E., Braconnot, P., Crowley, T. J., Delaygue, G., Joos, F., Krivova, N. A., Muscheler, R., Otto-Bliesner, B. L., Pongratz, J., Shindell, D. T., Solanki, S. K., Steinhilber, F., and Vieira, L. E. A.: Climate forcing reconstructions for use in PMIP simulations of the last millennium (v1.0), Geoscientific Model Development, 4, 33–45, https://doi.org/10.5194/gmd-4-33-2011, , 2011.

Schmidt, G. A., Annan, J. D., Bartlein, P. J., Cook, B. I., Guilyardi, E., Hargreaves, J. C., Harrison, S. P., Kageyama, M., LeGrande, A. N., Konecky, B., Lovejoy, S., Mann, M. E., Masson-Delmotte, V., Risi, C., Thompson, D., Timmermann, A., Tremblay, L. B., and Yiou, P.: Using palaeo-climate comparisons to constrain future projections in CMIP5, Climate of the Past, 10, 221–250, https://doi.org/10.5194/cp-10-221-2014, , 2014.

Seftigen, K., Linderholm, H. W., Drobyshev, I., and Niklasson, M.: Reconstructed drought variability in southeastern Sweden since the 1650s, International Journal of Climatology, 33, 2449–2458, https://doi.org/10.1002/joc.3592, , 2013.

Seiler, R. A., Hayes, M., and Bressan, L.: Using the standardized precipitation index for flood risk monitoring, International Journal of Climatology, 22, 1365–1376, https://doi.org/10.1002/joc.799, , 2002.

Sienz, F., Bothe, O., and Fraedrich, K.: Monitoring and quantifying future climate projections of dryness and wetness extremes: SPI bias, Hydrology and Earth System Sciences, 16, 2143–2157, https://doi.org/10.5194/hess-16-2143-2012, , 2012.

Smerdon, J. E., Cook, B. I., Cook, E. R., Seager, R., Smerdon, J. E., Cook, B. I., Cook, E. R., and Seager, R.: Bridging Past and Future Climate across Paleoclimatic Reconstructions, Observations, and Models: A Hydroclimate Case Study*, Journal of Climate, 28, 3212–3231, https://doi.org/10.1175/JCLI-D-14-00417.1, , 2015.

Sørland, S. L., Schär, C., Lüthi, D., and Kjellström, E.: Bias patterns and climate change signals in GCM-RCM model chains, Environmental Research Letters, 13, 074 017, https://doi.org/10.1088/1748-9326/aacc77, , 2018.

Stagge, J. H., Tallaksen, L. M., Gudmundsson, L., Van Loon, A. F., and Stahl, K.: Candidate Distributions for Climatological Drought Indices (SPI and SPEI), International Journal of Climatology, 35, 4027–4040, https://doi.org/10.1002/joc.4267, , 2015.

Swart, N. C., Fyfe, J. C., Hawkins, E., Kay, J. E., and Jahn, A.: Influence of internal variability on Arctic sea-ice trends, Nature Clim. Change, 5, 86–89, https://doi.org/10.1038/nclimate2483 10.1038/nclimate2483, 2015.

Tejedor, E., de Luis, M., Cuadrat, J. M., Esper, J., and Saz, M. Á.: Tree-ring-based drought reconstruction in the Iberian Range (east of Spain) since 1694, International Journal of Biometeorology, 60, 361–372, https://doi.org/10.1007/s00484-015-1033-7, , 2016.

Tout, D. G.: Precipitation-temperature relationships in England and Wales summers, Journal of Climatology, 7, 181–184, https://doi.org/10.1002/joc.3370070208, , 1987.

Trouet, V., Babst, F., and Meko, M.: Recent enhanced high-summer North Atlantic Jet variability emerges from three-century context, Nature Communications, 9, 180, https://doi.org/10.1038/s41467-017-02699-3, , 2018.

University of East Anglia Climatic Research Unit, Harris, I., and Jones, P.: CRU TS4.01: Climatic Research Unit (CRU) Time-Series (TS) version 4.01 of high-resolution gridded data of month-by-month variation in climate (Jan. 1901- Dec. 2016), https://doi.org/10.5285/58a8802721c94c66ae45c3baa4d814d0, 2017.

Van Loon, A. F.: Hydrological drought explained, Wiley Interdisciplinary Reviews: Water, 2, 359–392, https://doi.org/10.1002/wat2.1085, , 2015.

Venables, W. N. and Ripley, B. D.: Modern Applied Statistics with S, Springer, New York, fourth edn., http://www.stats.ox.ac.uk/pub/MASS4, iSBN 0-387-95457-0, 2002.

Warnes, G. R., Bolker, B., and Lumley, T.: gtools: Various R Programming Tools, https://CRAN.R-project.org/package=gtools, r package version 3.8.1, 2018.

Wei, T. and Simko, V.: R package "corrplot": Visualization of a Correlation Matrix, https://github.com/taiyun/corrplot, 2017.

Wigley, T. M. L., Lough, J. M., and Jones, P. D.: Spatial patterns of precipitation in England and Wales and a revised, homogeneous England and Wales precipitation series, J. Climatol., 4, 1–25, https://doi.org/10.1002/joc.3370040102, ~~,~~1984.

Wilson, R., Miles, D., Loader, N., Melvin, T., Cunningham, L., Cooper, R., and Briffa, K.: A millennial long March–July precipitation reconstruction for southern-central England, Climate Dynamics, 40, 997–1017, https://doi.org/10.1007/s00382-012-1318-z, ~~,~~2013.

Wilson, R. J. S., Luckman, B. H., and Esper, J.: A 500 year dendroclimatic reconstruction of spring–summer precipitation from the lower Bavarian Forest region, Germany, Int. J. Climatol., 25, 611–630, https://doi.org/10.1002/joc.1150, http://dx.doi.org/10.1002/joc.1150, 2005.

Woodley, M. R.: A review of two national rainfall series, International Journal of Climatology, 16, 677–687, https://doi.org/10.1002/(SICI)1097-0088(199606)16:6<677::AID-JOC43>3.0.CO;2-X, 1996.

Xie, Y.: Dynamic Documents with R and knitr, Chapman and Hall/CRC, Boca Raton, Florida, 2nd edn., https://yihui.name/knitr/, iSBN 978-1498716963, 2015.

Yadav, R. R., Misra, K. G., Yadava, A. K., Kotlia, B. S., and Misra, S.: Tree-ring footprints of drought variability in last ~300 years over Kumaun Himalaya, India and its relationship with crop productivity, Quaternary Science Reviews, 117, 113–123, https://doi.org/10.1016/J.QUASCIREV.2015.04.003, ~~,~~2015.

Yee, T. W.: Vector Generalized Linear and Additive Models: With an Implementation in R, Springer, New York, USA, 2015.

Young, G. H. F., Loader, N. J., McCarroll, D., Bale, R. J., Demmler, J. C., Miles, D., Nayling, N. T., Rinne, K. T., Robertson, I., Watts, C., and Whitney, M.: Oxygen stable isotope ratios from British oak tree-rings provide a strong and consistent record of past changes in summer rainfall, Climate Dynamics, 45, 3609–3622, https://doi.org/10.1007/s00382-015-2559-4, ~~,~~2015.

Zang, C. and Biondi, F.: treeclim: an R package for the numerical calibration of proxy-climate relationships, Ecography, 38, 431–436, https://doi.org/10.1111/ecog.01335, 2015.

Zeileis, A. and Grothendieck, G.: zoo: S3 Infrastructure for Regular and Irregular Time Series, Journal of Statistical Software, 14, 1–27, https://doi.org/10.18637/jss.v014.i06, 2005.

Zhang, X., Zwiers, F. W., Hegerl, G. C., Lambert, F. H., Gillett, N. P., Solomon, S., Stott, P. A., and Nozawa, T.: Detection of human influence on twentieth-century precipitation trends, Nature, 448, 461–465, https://doi.org/10.1038/nature06025~~,~~, 2007.