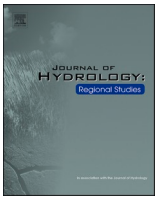
Journal of Hydrology: Regional Studies 68 (2026) 104135



Contents lists available at ScienceDirect

Journal of Hydrology: Regional Studies

journal homepage: www.elsevier.com/locate/ejrh



Assessment of whether tree-ring hydroclimate reconstructions are useful for characterising extreme drought in East Anglia (England)

Jane Thurgood *, Timothy Osborn

University of East Anglia, Norwich Research Park, Norwich, England NR4 7TJ, UK

ARTICLE INFO

Keywords:
Extreme drought
Tree-ring reconstructions
East England
Hydroclimate variability
Water resource management

ABSTRACT

Region: East of England
Focus: Paleoclimate proxies are used here to characterise extreme droughts including those corresponding to a 1:500-year return period. We evaluate four tree-ring hydroclimate (precipitation or drought index) reconstructions representing similar geographical regions across East Anglia and Southern England.
New hydrological insights: The overall agreement between the reconstructions is found to be weak or moderate, partly due to their differing seasons, regions and methodological differences. Therefore, we used all four reconstructions to quantify the sensitivity of drought to choice of reconstruction. We show that reconstruction errors influence the quantification of extreme dry seasons (here spring/summer seasons) and must be accounted for in the analysis especially if the reconstruction was calibrated using regression. We demonstrate that modelling the reconstruction error as a random term, as well as z-score standardisation, enables the reconstruction to reproduce instrumental climate behaviour such as the rank-order drought curve. Using these approaches, we quantify the occurrence probability of high threshold extreme events and evaluate the differences in severity of droughts across the reconstruction datasets (which vary in length from 800 to 1112 years). We conclude that tree-ring reconstructions can provide valuable long term hydroclimate insight, but they should be used with caution when investigating extreme droughts due to inter-reconstruction differences and the need to represent reconstruction errors.

1. Introduction

The UK has a well-established framework for water resources and drought planning. The purpose of these plans has typically been to ensure security of supplies according to a 'drought of record' (a benchmark historical event, such as 1921–22). However, regulatory changes mean that water companies are now required to assess their resilience to severe (1-in-200 year) or extreme (1-in-500 year) drought events (Environment Agency, 2025). Characterising such rare events is very challenging, and the inherent scientific questions are only beginning to be tackled, but the costs (in terms of water supply infrastructure investments) that rest on such assessments are potentially vast. This paper considers this challenge with a focus on East Anglia (the east of England) but demonstrates approaches that are applicable to other regions.

The instrumental record is not long enough to characterise such low probability events. Long records or large samples that realistically represent long-term (multidecadal to multicentennial) variability are necessary because greater long-term climate variability

* Corresponding author.
E-mail address: J.Thurgood@uea.ac.uk (J. Thurgood).

https://doi.org/10.1016/j.ejrh.2026.104135
Received 16 June 2026; Received in revised form 14 September 2026; Accepted 20 October 2026
Available online 30 October 2026
2214-5818/© 2026 Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

would enhance the intensity of extreme droughts relative to those that occur in a climate without long-term variability (Li et al., 2016; Martin et al., 2020). Multiple approaches can address this challenge by simulating long series or large samples of hydrometeorological variability (with physical climate models or statistical stochastic models), but model biases and gaps in understanding exist (Valler et al., 2024).

An alternative approach, applied in the current study, is to make use of millennial length reconstructions of past hydroclimatic variability to provide a large sample of yearly values from which the intensity of low-probability events can be estimated (Smerdon, 2012). For this purpose, tree-ring data have useful characteristics (Esper et al., 2018). Their annual resolution allows calibration against instrumental data in the time domain and the potential to record individual yearly droughts. Each record is assembled from multiple, precisely-dated samples and the level of agreement between the constituent series enables a quantitative evaluation of the reliability of the tree-growth timeseries (Wigley et al., 1984). Conversely, being complex biological organisms means that their growth is affected by multiple confounding factors (climatic and non-climatic) including a dependence on the biological age of each tree that must be accounted for with care to avoid removing real climatic trends (Briffa & Melvin, 2011).

In regions where tree growth is strongly limited by available moisture, the statistical association between measures of annual tree growth and instrumental records of hydroclimate (precipitation, soil moisture, drought) can be very strong (Yang et al., 2014) and such information has proved to be of use to water resource managers (Rice et al., 2009). This is especially the case in the western US (Woodhouse et al., 2016) and there is the potential for similar approaches in other arid and semi-arid regions. Even though East Anglia is one of the driest regions in England and Wales, tree growth is not strongly limited by moisture so they show only moderate skill in reconstructing past variations in hydroclimate (Cooper et al., 2013). Therefore, while we make use of the available tree-ring reconstructions, we must assess them cautiously and take into account reconstruction uncertainties.

In applying tree-ring reconstructions to the characterisation of low probability events two important issues must be considered, in addition to reconstruction uncertainties. The first is *stationarity*. The drought quantiles calculated from ~1000-year long reconstructions would not be applicable to current climate if the statistical properties of the climate in the earlier period were different to the current climate (i.e. there is significant non-stationarity). This is a valid concern at large-scales (e.g. continental to global) and for variables such as temperature where climate change driven by natural and anthropogenic forcing provides easily detectable non-stationarity. However, local precipitation is dominated by internal variability rather than external forcing (Deser et al., 2010; Hawkins and Sutton, 2009). We tested this by analysing the ensemble mean of a 60-member ensemble of externally-forced simulations using the ECHAM6 atmospheric general circulation model for the period 1420–2009 (the ModE-Sim dataset; Hand et al., 2023). For spring/summer precipitation over East Anglia, the ModE-Sim ensemble mean shows no significant response to either the strongest natural forcing (tested by a composite of 11 strong volcanoes) or to anthropogenic forcing over the last 150 years. So, although our estimated return values should be interpreted as statistical descriptors valid only under an assumption of stationarity, a model analysis suggests that the impact of this assumption may be small.

The second issue is *seasonality*. The tree-ring reconstructions are for spring/summer precipitation whereas water resource managers in East Anglia are particularly concerned with winter recharge and would ideally like year-round information. Nevertheless, prolonged spring-summer drought is an important component of drought hazard and 2026 provides a timely example where winter recharge was not notably below average yet subsequent dry conditions, high evaporation and reduced soil moisture led to temporary water use restrictions during summer. Therefore, our analysis of spring/summer extreme droughts is of direct relevance to water-resource management even though it is insufficient to provide a direct assessment of water-supply resilience. It could also indirectly benefit multi-method approaches that use climate or stochastic models to obtain year-round hydrometeorological information by providing a test for whether those models can produce realistic spring/summer extreme droughts.

The paper is structured as follows. Section 2 reviews the available tree-ring hydroclimate reconstructions relevant to the East Anglian region and describes the methods used to analyse them. Sections 3–7 present the analysis undertaken on the tree-ring-based hydroclimate reconstructions and their implications for characterising drought severity and return periods. Section 8 is the summary and concluding remarks.

2. Data and methods

2.1. Review of the available hydroclimate reconstructions

Tree ring width is influenced by both the growth from the previous year and the climate conditions in the current year. Response functions are used to determine the climate-tree relationships empirically but other external factors can also affect growth (Mašek et al., 2021) and some can make it difficult to extract low-frequency climate information from tree rings (Cooper et al., 2013; García-Suárez et al., 2009). Age-related growth trends must be removed but the climate reconstructions can be compromised if low-frequency climate signals are removed along with the age effects (Cook et al., 1995; Cooper et al., 2013). This limitation is particularly important when studying extreme droughts, as long- term climate variability, if present, can amplify their intensity (i.e., droughts occurring within a variable long-term climate may be more severe than those arising under more stable conditions (Cook et al., 2015)). Regional curve standardisation (RCS; (Briffa & Melvin, 2011)) is a method to preserve low-frequency growth variability as it adjusts ring-width data by accounting for expected growth based on tree age, species, and regional conditions (García-Suárez et al., 2009).

Isotope data is advantageous as there is no evidence of age-related trends (Duffy et al., 2019), and so stable oxygen isotopes do not need statistical detrending which avoids the potential loss of low-frequency climate signals. As trees grow, they absorb carbon from atmospheric carbon dioxide, and hydrogen and oxygen from soil water. The stable isotope ratios of these elements reflect past climate

conditions, as they are regulated by climate factors which influence the tree's water and gas exchange processes (Gagen et al., 2011; McCarroll & Loader, 2004). Although rainfall amount does not directly control the variation in isotope ratios, the isotopic signals do differ between wet and dry summers (Barbour, 2007; Saurer et al., 1997; Treydte et al., 2014). These signals are incorporated by the tree regardless of whether the tree is experiencing environmental stress (Young et al., 2015). The use of latewood alpha-cellulose ensures that the tree-ring data reflects environmental conditions from the late spring and summer of the year in which the ring was formed, making it a good indicator for seasonal climate analysis (Loader et al., 2020; McCarroll & Loader, 2004).

The oxygen isotope ratio ($\delta^{18}O$) can provide insights into historical changes in precipitation patterns, humidity, and atmospheric composition (Gagen et al., 2011; McCarroll & Loader, 2004; Roden et al., 2000). Compared with tree-ring width reconstructions, δ^{1} ^{8}O records generally integrate hydroclimate signals over larger spatial scales because the isotope composition of precipitation is influenced by regional atmospheric circulation and humidity (Gagen et al., 2011; McCarroll & Loader, 2004). In contrast, tree-ring width is more strongly controlled by local site conditions, including soil moisture availability. Consequently, isotope and ring-width reconstructions may differ in their representation of extreme droughts while still recording valid, but complementary aspects of hydroclimate variability.

A number of hydroclimate reconstructions (Table 1 and Fig. 1) have been made from UK networks of oak tree-ring width data (living trees and timbers from historical buildings). For southern/central England, Wilson et al. (2013); hereafter "*Wilson*"); reconstructed spring/summer (March–July) precipitation from 980 CE to the present. This captured 33% of the variance seen in overlapping instrumental data, although there is a potential deterioration in the quality for 1800–1920 that might be related to pollution. For East Anglia, Cooper et al. (2013); hereafter "*Cooper*") reconstructed spring/summer (March–July) precipitation from 900 CE to the present, also capturing around a third of the observed variance at annual to decadal timescales.

Information on past drought can also be obtained from broader geographical reconstructions based on tree-ring width. For example, the "Old World Drought Atlas" Cook et al. (2015); hereafter "*Cook*") uses a pan-Europe network of tree-ring width records to reconstruct a summer drought indicator (the self-calibrating Palmer Drought Severity Index, scPDSI) from about 900 CE to present. The *Cook* reconstruction is based on the *Wilson* and *Cooper* series along with a wider network of ring-width reconstructions, some of which have not been standardised using RCS. This has multiple implications. First, the *Cook* reconstruction could have increased reliability from climate information in the wider network used (and by combining *Wilson* and *Cooper* together). Second, the *Cook* reconstruction may retain less multi-centennial variability than reconstructions based exclusively on RCS-standardised series. Third, the inclusion of *Wilson* and *Cooper* leads to a degree of mathematically pre-conditioned shared variance, i.e. *Cook* cannot provide fully independent support for findings based on *Wilson* or *Cooper*.

Other work shows a notably stronger (>40% variance) correlation exists between UK precipitation and the δ18 O composition of tree rings in UK oaks (Young et al., 2015). Moreover, the tree-ring $\delta^{18}O$ rainfall signal appears to be more stable over time than the tree-ring width signal. Loader et al. (2020); hereafter "*Loader*") reconstructed summer (May-August) precipitation from 1200 CE to the present using $\delta^{18}O$ measurements from oak trees across southern and central England.

The two key issues that determine the utility of tree-ring data for better quantifying the characteristics of 1-in-500-year droughts are their reliability and seasonal sensitivity. The type of tree-ring measurements used in these reconstructions cannot capture individual short-term events; instead, they integrate these into an overall seasonal response. Therefore, a key question is which seasonal average best represents the environmental conditions influencing growth. The studies chose their target variables based on the processes which are understood to control tree growth in the region, primarily soil moisture during the growing season. The final selection of variable and season is made empirically, by testing different options and choosing the strongest correlation. While this approach improves fit, it also increases the risk of overfitting by choosing the variable and seasonal window which maximises the correlation during the calibration period, but risks compromising the reconstructions prior to the calibration period.

Loader is potentially more reliable at reconstructing summer precipitation than the ring-width records because the ratio of $\delta^{18}O$ in the tree cellulose is a more direct signal of precipitation than is ring width and is less affected by tree growth and age, as well as non-climatic factors. Indeed, Loader et al. (2020) shows that the *Loader* reconstruction has a stronger correlation with observed precipitation and a better representation of extremely dry summers than the ring-width-based reconstructions. However, the much higher analytical costs and expertise limit the spatial coverage and sample size of the available isotope data. *Cook's* gridded European scPDSI reconstruction uses a network of tree ring width records. Although we extract a grid cell containing East Anglia for this study, it could

Table 1
key information about the paleoclimate datasets.

Dataset	Period	Season	Geographical focus	Proxy variable	Calibration period
Cooper	900–2009	March-July precipitation	East Anglia (52–53°N, 0–2°E)	Oak tree ring width	1901–2009
Wilson	950–2009	March-July precipitation	Southern England	Oak tree ring width	1901–2007
Loader	1201–2000	May-August precipitation	South-central England	Oak tree ring oxygen isotope	1766–2000
Cook	900–2012[a]	June-August scPDSI drought	52–53°N, 0.5–1.9°E grid cell extracted from Europe/Mediterranean	European network of tree rings	1928–1978
Bebchuk	1710–2010	April-July scPDSI and precipitation	Southern England	Yew tree ring width	1901–2020

[a] Note: (Cook et al., 2015) data available from year 0 but it is not based on any tree-ring data within the East Anglian region prior to 900 so the period 900–2012 is used here.

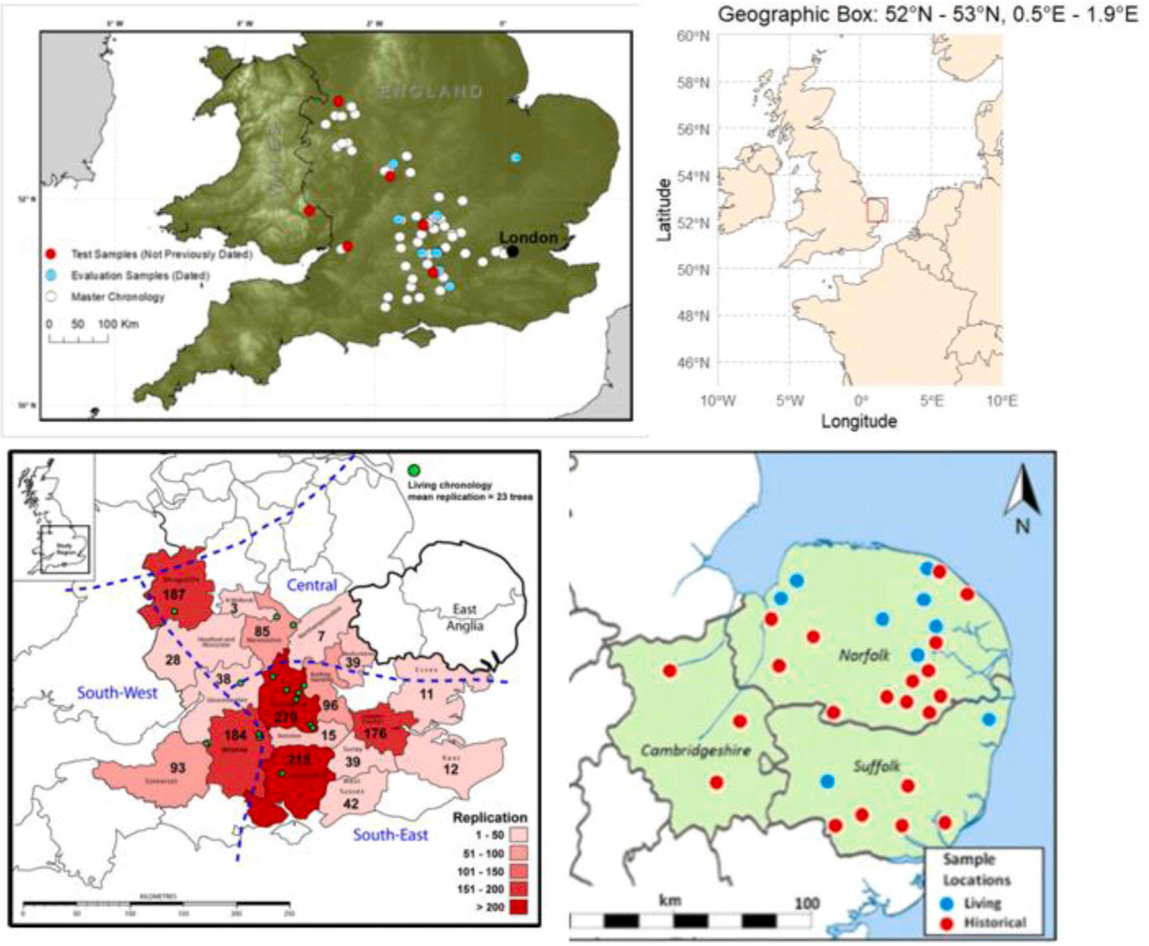


Fig. 1. Maps from the respective papers of the paleo datasets showing where the trees are located. A) *Loader* (Loader et al., 2020) B) Old World Drought Atlas (Cook et al., 2015) where the grid cell mostly containing East Anglia is located and extracted C) *Wilson* (Wilson et al., 2013) D) *Cooper* (Cooper et al., 2013).

nevertheless smooth out local signals which are specific to East Anglia. Summer scPDSI is a widely used metric that reflects both precipitation and evaporative losses (which are dependent on temperature) to represent soil moisture levels and summer drought conditions that incorporate the influence of precipitation deficits from previous seasons.

2.2. Data

Four tree-ring-based hydroclimate reconstructions were used for the main analyses (Cook et al., 2015; Cooper et al., 2013; Loader et al., 2020; Wilson et al., 2013). Table 1 summarises the key information from each dataset. All datasets have annual resolution and are calibrated to represent hydroclimatic variables (precipitation or drought index), but differ in scope, methodology, and specific climatic signals.

Fig. 1 shows maps of the tree sampling sites used in each of the studies from their respective papers. The geographic range of each dataset is broadly similar, mainly focusing on East Anglia and Southern England (Fig. 1). Even so, differences in the localised geographic range can lead to differences between the datasets, these may include localised differences in topography, soil type, land use and coastal proximity which can influence precipitation patterns due to maritime influences. Conversely, further inland may be more influenced by convective rainfall (Darwish et al., 2021). These effects lead to different mean precipitation, but temporal variability of monthly precipitation is significantly correlated over distances of around 450 km (New et al., 2000) so we expect these regions to be well correlated because they are separated by much less than 450 km. Moreover, the vegetation types, land use history, and regional hydroclimate drivers are broadly the same across the regions meaning that tree ring growth is likely to reflect similar climatic influences. As a result, although there are localised and site-specific differences, the overall trends and precipitation signals represented in the tree rings are likely to be consistent across the datasets.

A fifth hydroclimate reconstruction Bebchuk et al. (2025); hereafter *Bebchuk*) was used as an additional comparator (Table 1). This is based on Yew tree ring-width chronologies from southern England. It is not used for the main analysis of drought return levels

because it is shorter than the other reconstructions and therefore can add little insight beyond what can already be gained from instrumental data. However, it is useful as an independent indicator of the potential reliability of the other reconstructions for the period since 1700, especially for the question of whether the *Loader* isotope series may be more or less reliable than those based on oak ring-widths.

2.3. Methods

Some analyses were undertaken in the original units of each reconstruction (mm/season or scPDSI units) so that droughts could be assessed in physical variables. However, some analyses were done after each reconstruction was z-score standardised so that they had the same mean and standard deviation (SD) over a common period, to enable comparisons between the datasets.

Rank-based estimates of drought severity were derived by ordering each year's reconstructed values from wettest to driest. Expressing the rank as a quantile (between 0 and 1) allows datasets with different lengths to be compared on the same relative scale, facilitates return values to be calculated for each dataset and extreme years to be compared (but noting that differences in region, season, or hydroclimate variable mean that the absolute values cannot be directly compared). Thresholds for three different return period droughts are estimated: e.g. 50-year (2% annual probability of occurrence), 100-year (1% probability), and 500-year (0.2% probability) droughts. These thresholds (or return values) are relevant to current climate only under the assumption of climate stationarity; see Section 1 for further discussion and justification.

The calibration period, during which the proxy data is statistically matched to the instrumental records, varies (Table 1). Differing periods may change which climate variabilities are captured and influence how the extreme events are weighted which changes the model sensitivity and accuracy, but a reliable reconstruction should not be overly sensitive to this choice. Longer calibration periods capture more natural variability before industrial influences and human induced climate change. This may reduce bias towards the more modern climate states which show recent drying trends (Briffa et al., 2009), however, in the earlier years there may be less precise instrumental data.

We use instrumental precipitation to evaluate the expected level of similarity between the different regions and seasons represented by the reconstructions (Table 1) and for exploring reconstruction error (Table 3). For *Loader* we use the England and Wales Precipitation (EWP) series from 1766 they provided. To maintain consistency and extend the time period of analysis, HadUK-Grid (Hollis et al., 2019) administrative regional average precipitation for East Anglia and South East England were used for the *Cooper* and *Wilson* datasets, respectively, extending from 1836 to the end date of each reconstruction, rather than the shorter (1901 onwards) data these publications originally used. South-East England is a slightly different region to that used by Wilson et al. (2013) but they show

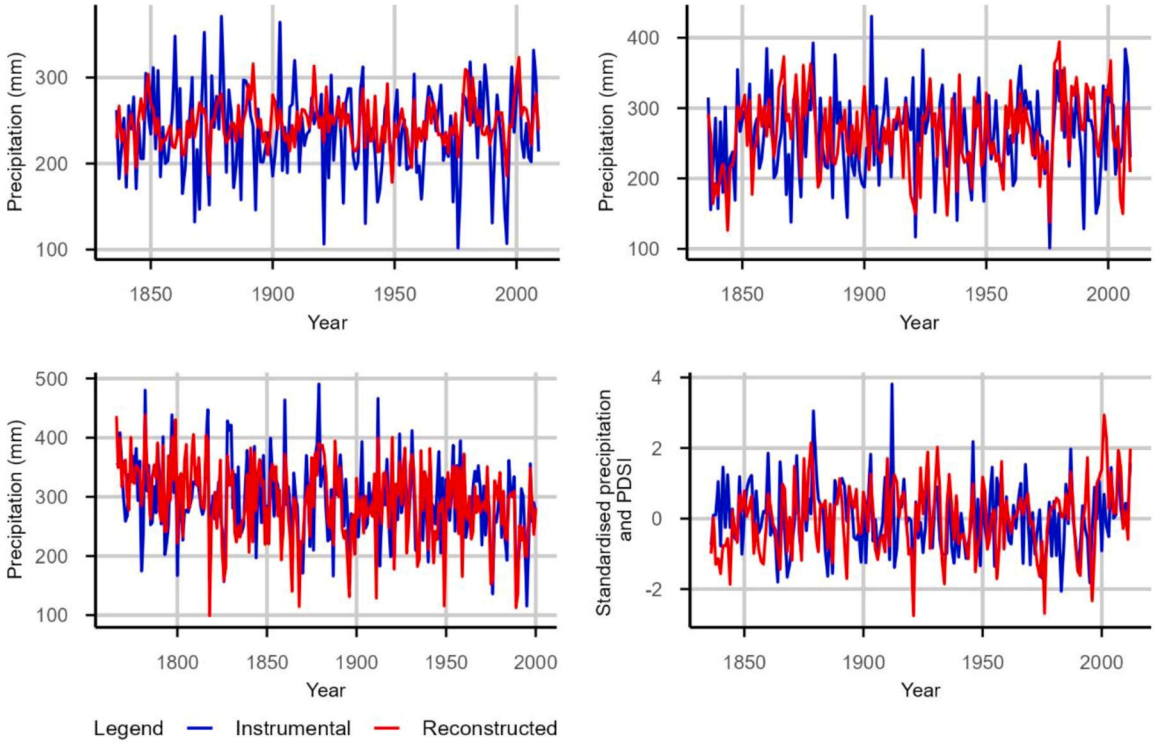


Fig. 2. The instrumental and reconstructed precipitation (mm/season) for the overlapping periods for each of the paleo-datasets; top left *Cooper,* top right *Wilson,* bottom left *Loader,* bottom right *Cook*. For *Cook*, the instrumental data has been standardised to enable comparison between the reconstructed PDSI (dimensionless).

(their Fig. 4a) a strong correlation with this region that is just as strong as the correlation with the gridded precipitation that they used. *Cook* calibrated against the scPDSI drought index, but to simplify comparison with the other precipitation-based reconstructions, four HadUK-Grid 60 km grid cells were extracted and averaged together as this most closely resembled the grid cell extracted from the *Cook* reconstruction over East Anglia. All monthly instrumental data were averaged to obtain the same seasons that the reconstructions represented (Table 1).

We evaluate one possible representation of reconstruction uncertainty and examine how it could influence the interpretation of extreme events in the reconstructed series for all cases except *Cook* (because the latter is not a precipitation reconstruction). The statistical adjustments ensure the error series represents the variance observed in the instrumental record, which we consider as a minimum requirement for producing credible drought return period estimates. We align each reconstruction so that its mean matches that of its comparator instrumental data over their common period by scaling by the ratio of instrumental and reconstruction means. An ensemble of 100 random errors is generated for the full length of each reconstruction using the methods described in Section 6 where the resulting error series ensures that the combined reconstruction plus error has the correct variance.

3. Comparisons during the instrumental period

In Section 4, we will use the level of agreement between the four reconstructions as one measure of their reliability. However, even if the reconstructions were highly accurate, some differences (and hence lower correlation) are expected because each record attempts to reconstruct slightly different regions, seasons and variables (Table 1). We can use instrumental data for these regions and seasons to assess how similar we expect each combination to be. We first compare each reconstruction with its target instrumental data to identify levels of agreement.

Fig. 2 shows the *Cooper* reconstruction follows the same general long-term trends with multi-year patterns tracked in the instrumental data, especially post 1900. Generally the reconstructed data has less variance than the instrumental, which leads to an underrepresentation of extreme wet and dry years, especially before 1900 but also in the 1920s, mid 1970s and late 1990s. This is partly a result of the regression approach used, which minimises the mean squared error rather than matching the variance of the observations. The *Wilson* reconstruction more closely follows its instrumental precipitation as both show high interannual variability, major fluctuations such as the wet periods of ~1900 and ~1950 appear in both series. The *Wilson* reconstruction data is smoother and occasionally misses wet and dry extremes, though not as noticeably as *Cooper*. The *Loader* reconstruction and instrumental series both show a general decline in precipitation and the reconstruction captures better the overall range and frequency of wet/ dry cycles, perhaps indicating greater reliability but also reflecting that the calibration was done by scaling the tree-ring series to match the variance of the instrumental record rather than by regression to minimise the mean squared errors. In the reconstruction some extreme wet years are underestimated rather than dry years. *Cook* shows a clear difference before the 1850s, perhaps because early instrumental data are less widespread or less reliable (Murphy et al., 2020). After 1850 the reconstruction follows a similar temporal pattern to the instrumental, with some occasional notable differences: e.g. a very wet year in the 1912 followed by a very dry year in 1913 in the instrumental data are not accurately captured by the reconstructed data. In general, each dataset's reconstructed precipitation generally follows its instrumental data quite well, suggesting reliability in the reconstructions, but with biases in variance that depend on how they were calibrated.

The strongest correlation between instrumental records is *Cooper-Wilson* (0.88; same season, adjacent regions), the others have moderate correlations with only *Wilson-Cook* showing a weaker (< 0.5; different but overlapping season, adjacent regions) correlation. From these correlations we would expect reconstructions that target different (though overlapping) seasons to be more different than those that target the same season but adjacent regions. Furthermore, we surmise that those that target both different seasons and regions will be the least similar. The comparison of instrumental series suggests that we would not expect really strong inter-reconstruction agreement if they target different (albeit overlapping) seasons, but targeting adjacent regions is less of an issue. This is relevant because we use the inter-reconstruction correlations as an indication of reconstruction reliability during the pre-instrumental period.

During the instrumental period the level of agreement *between* the reconstructions is quite mixed and most do not show good agreement. *Wilson-Cook* show a moderate to strong inter-reconstruction correlation which is slightly stronger than the correlation between the instrumental precipitation for these regions. This is perhaps because the June-August scPDSI is also influenced by precipitation from earlier months (spring and even, though to lesser extent, winter) which was not included in our instrumental comparisons, though *Cook* includes the *Wilson* data so it is not independent. *Cooper-Wilson* has the strongest instrumental correlation, but their reconstructed correlation is notably weaker implying that the reconstructions did not preserve all the shared signal. This is a similar case with the *Loader-Cook* correlations, likely due to the poor agreement between the $\delta^{18}O$ (*Loader*) and ring-width (*Cook*) records.

Instrumental data correlations generally show good agreement and are higher than the reconstructed data correlations (during the instrumental period) which suggests the instrumental precipitation patterns share more common variability than reconstructions. Next we compare the reconstructions over their full lengths.

4. How well do the paleo-datasets agree with each other?

One indicator of reliability is whether independent records are consistent with each other, so here we compare the reconstructions in more detail for different timescales (different levels of smoothing). Multi-year to multi-century variations are key because they represent the changing background state within which droughts occur, and stronger long-term variations can lead to long return period

droughts being more severe than if they occurred in a more stationary climate.

The data was z-score standardised to be brought into a common scale to enable visual comparison, then smoothed using moving averages. Fig. 3 shows all datasets have significant variability, with numerous peaks (wet periods) and troughs (dry periods). Datasets do not align, particularly at the 50-year timescale as some series oppose each other, for example, just after 1500, *Cooper* trends wetter,

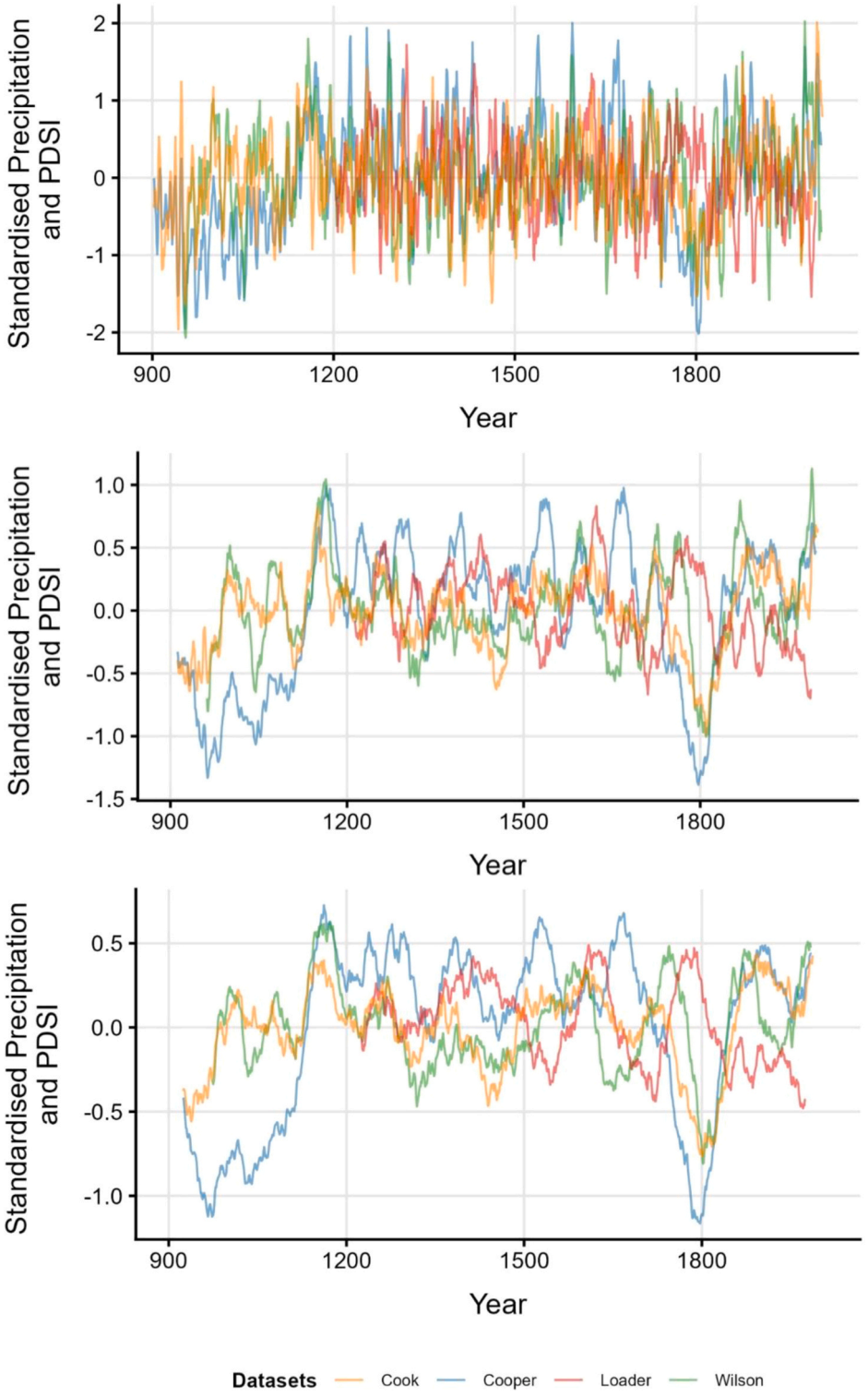


Fig. 3. Z-score standardised datasets where the reconstruction series are smoothed with running means of length 5, 25 and 50 years, respectively.

Loader trends drier, while *Cook* and *Wilson* remain fairly consistent. Periods with prolonged low growth – and hence inferred dry conditions relative to the twentieth century – are apparent in *Cooper* around 950–1150 CE and in *Cooper*, *Wilson* and *Cook* for 1770–1820 CE, but *Loader* is relatively wet during the latter period (Fig. 3). However, similarities are found throughout the series, for

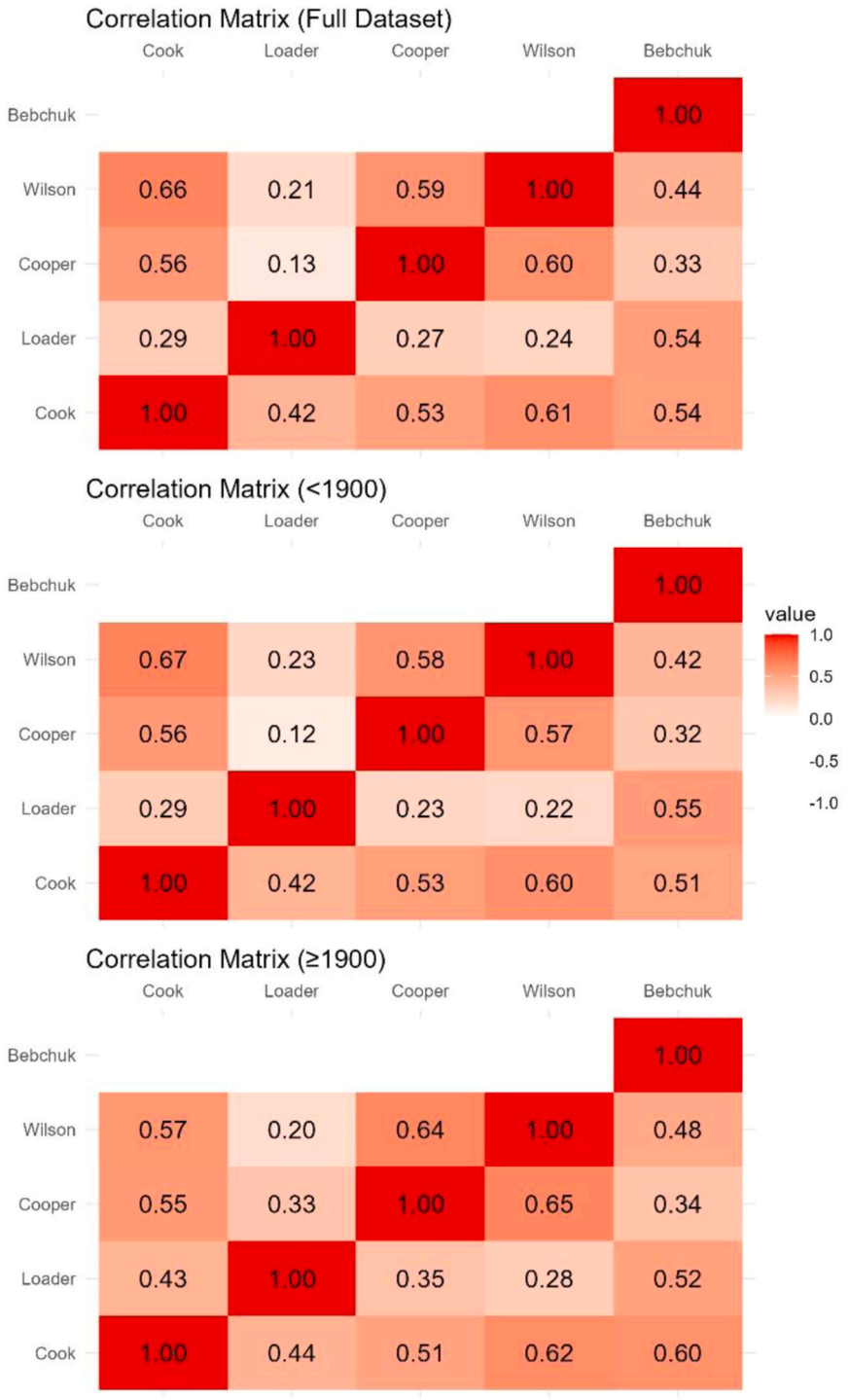


Fig. 4. Correlation matrix of each dataset pair calculated using the full dataset, pre-1900 and post-1900. Above the diagonal: correlations calculated from the original (non-detrended) series, excluding the *Bebchuk* reconstruction. Below the diagonal: correlations calculated after 130-year spline detrending, including correlations with *Bebchuk*. All correlations are significant at 95% confidence level although after 1900 above the diagonal *Wilson* and *Loader* are only marginally so at $p \approx 0.047$.

example, at the 25- and 50-year timescales, the *Cook* and *Wilson* series follow broadly similar patterns. *Cook-Cooper* differences are notably larger throughout time as *Cooper* remains wetter throughout most of the time series, however, from 1800 onwards *Cook* and *Cooper* are almost identical and *Wilson* diverges from the other two. Across timescales *Cooper*, *Cook* and *Wilson* shift towards wetter conditions and it is only *Loader* which becomes drier at the end of the record.

To explore the level of agreement between the reconstructions in more detail, each pairwise combination of datasets was evaluated using correlations (Fig. 4 above the diagonal). Even if the individual reconstructions were highly reliable, we would expect some difference because they represent different climate variables/regions and span different seasons (Section 3). Key findings are that the most strongly correlated reconstructions are *Cook* and *Wilson* ($r = 0.66$), followed closely by the *Cooper-Wilson* ($r = 0.59$) and the *Cook-Cooper* ($r = 0.56$) pairs, though recall that *Cook-Wilson* and *Cook-Cooper* are not fully independent. *Wilson* and *Cooper* both use RCS to retain climate-driven low-frequency trends in oak growth and they are based on independent data so the moderately strong correlation between them provides some confidence in their reliability. Overall, *Loader* has the weakest relationship with each of the other datasets. This could indicate that *Loader* is less reliable than the others, but Loader et al. (2020) claim that the closer agreement between their reconstruction and instrumental rainfall data, especially for extreme years, is an indication that their isotope-based reconstruction is more accurate than the others that are based on ring-width data.

Correlations before 1900 and after 1900 (Fig. 4) allow us to examine whether the inter-reconstruction correlations are artificially raised by over-fitting during the calibration period (which generally uses data after 1900) at the expense of creating reconstructions that are less reliable overall, indicated by substantially weaker correlations outside the calibration period (prior to 1900). Overall, the correlations tend to be stronger after 1900. However, the difference is not large (the average of the 6 pairwise correlations post-1900 is 0.45, only slightly higher than the pre-1900 average of 0.41). The correlations before 1900 remain very similar to the full dataset, especially for *Cook* and *Wilson* ($r = 0.67$) with *Cooper* and *Cook* remaining the same, *Loader* persists with weak correlations. After 1900 the correlations between *Cook* and *Wilson* weaken slightly ($r = 0.57$). *Loader* does show stronger correlations with two of the other series after 1900, perhaps suggestive of greater but artificial similarity during the calibration period, but the difference is not large: with *Cook* ($r = 0.43$) and *Cooper* ($r = 0.33$) the correlations are moderate, whereas with *Wilson* the correlation ($r = 0.20$) is slightly weaker than pre-1900.

Overall, this suggests that the reconstructions retain comparable reliability outside the calibration period and are not strongly biased by over fitting during calibration. Nevertheless, the correlations between instrumental series are still notably stronger than the pre-1900 inter-reconstructions correlations (compare Fig. 4 and Table 2 above the diagonal) in all cases except *Cook-Wilson* and *Cook-Cooper*, and for the latter the inter-instrumental correlations may be suppressed using JJA precipitation for *Cook* as noted above. Therefore, this analysis does not fully ameliorate concerns about the reliability of the reconstructions before 1900. In particular, there is poor agreement between the isotope-based and ring-width-based reconstructions, and even between the ring-width series the level of agreement is not particularly high (around $r = 0.60$). Further investigation is needed if we are to have high confidence in the set of reconstructions as a whole.

The Yew ring-width spring-summer (April-July) rainfall reconstruction of (Bebchuk et al., 2025) (Table 1) provides an independent comparison. Bebchuk et al. (2025) also created a July scPDSI reconstruction from the same ring-width measurements (but processed and calibrated differently) which shows drought conditions with a more prominent drying trend since 1990s. Here we use their rainfall reconstruction because it directly quantifies precipitation allowing a more direct comparison with the *Loader*, *Cooper* and *Wilson* series and their two reconstructions (rainfall and drought) correlate very highly with each other at 0.93 (Bebchuk et al., 2025). The primary value of this reconstruction is in assessing variability on timescales below ~100 years because variability on longer timescales was greatly reduced by the 130-year spline detrending applied by Bebchuk et al. (2025). For a fair comparison, each of the main reconstructions was also detrended with a 130-year spline prior to calculating their correlation with *Bebchuk* (Fig. 4) and undertaking a visual comparison (Fig. 5).

The *Bebchuk* yew-tree-based rainfall reconstruction is positively associated with all four of the datasets across their overlapping time spans (Fig. 4 below the diagonal), although its overall agreement is weakest with the *Cooper* series. Prior to 1900 when independent validation of the reconstructions is particularly important, *Bebchuk* exhibits a moderate correlation with the *Wilson* reconstruction but aligns most strongly with *Cook* and *Loader*. This lends some support to the hypothesis that the *Loader*'s $\delta^{18}O$-based reconstruction may be the most reliable, though the level of agreement ($r = 0.55$) is still not strong enough to warrant a definite validation. Furthermore, due to the *Bebchuk* detrending methodology, the comparison (Fig. 5) is unable to support or undermine the centennial-scale deviations (e.g. that centred on 1800 in the bottom panel of Fig. 3) between *Loader* and the ring-width series. Instead, the contrasting reconstructions illustrate the uncertainty associated with creating long-term hydroclimate variability from tree-ring records and underscore the importance of considering multiple reconstruction approaches when assessing extreme drought events.

Table 2
Correlations between the instrumental data above the diagonal and correlations between the reconstruction data below the diagonal. Correlations shown are for the period of overlap between the reconstruction and instrumental data.

	Cook	Cooper	Loader	Wilson
Wilson	0.47	0.88	0.57	1
Loader	0.82	0.59	1	0.11
Cooper	0.58	1	0.27	0.59
Cook	1	0.49	0.38	0.60

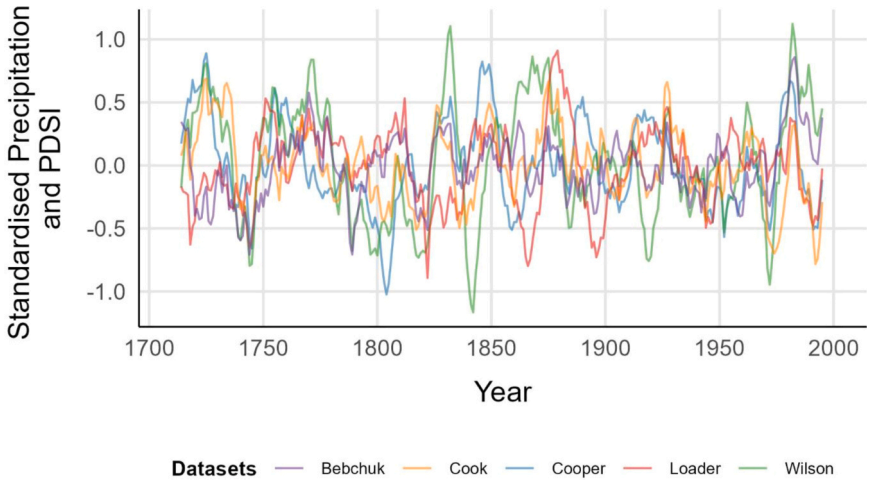


Fig. 5. Comparison of Z- score standardised *Bebchuk* data with the four hydroclimate reconstructions, detrended with a 130-year spline and smoothed with a 10-year rolling window. The period spans the overlapping years of the reconstructions.

5. What do the paleo-records tell us about when the driest years occurred and how dry they were?

We have seen in the previous sections that there is only moderate agreement between the reconstructions, especially prior to the main instrumental period, and that this can partly be explained by differences in the regional and seasonal representations, but also likely to be in part down to some unreliability, or at least uncertainty, in one of more reconstructions. There is some evidence that the *Loader* $\delta^{18}O$ reconstruction may be more reliable than the others as that reconstruction more closely matches the instrumental data especially for dry years (Loader et al., 2020) and shows a stronger correlation with the independent *Bebchuk* reconstruction, but enough caveats remain that we should consider all four as candidate reconstructions and evaluate what each of them indicates about

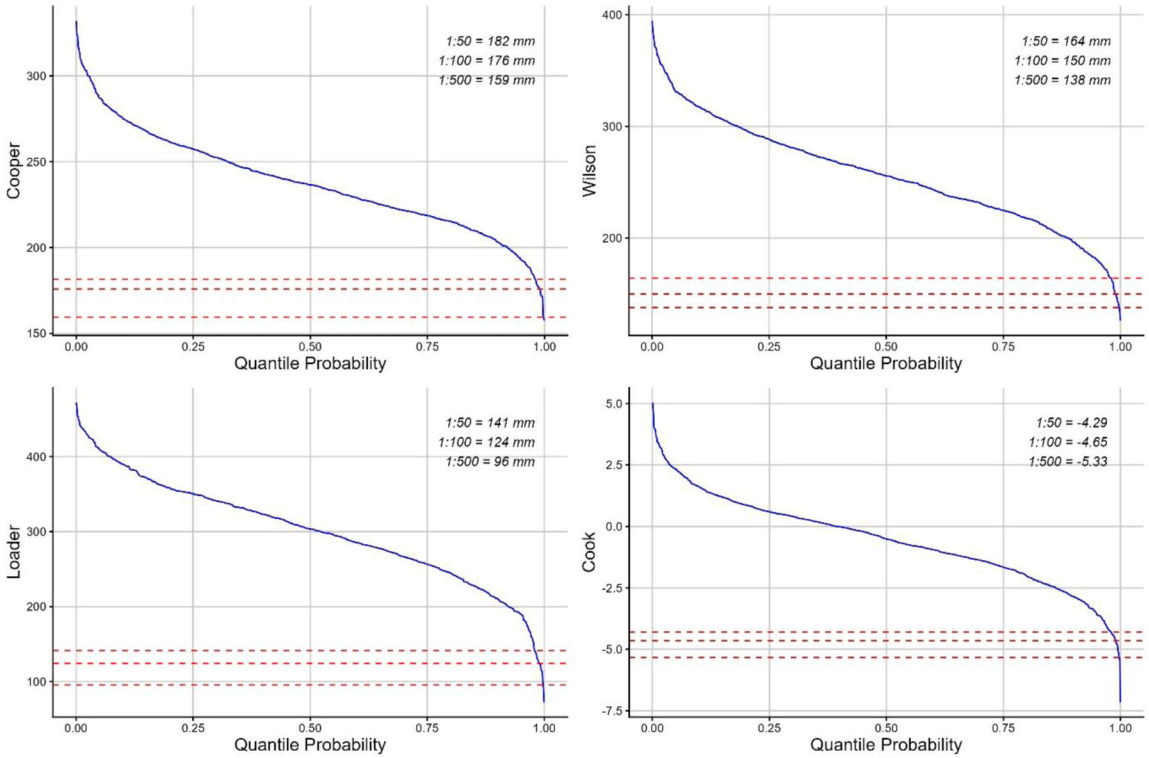


Fig. 6. Ranked values (mm/season and dimensionless *Cook*) of each paleo dataset, from wettest year to driest. Top left *Cooper,* top right *Wilson,* bottom left *Loader,* bottom right *Cook*. Each dashed red line (from upper to lower) marks the threshold for droughts with return periods of 1:50, 1:100 and 1:500 years, respectively. Thresholds (mm/season and dimensionless *Cook*) are recorded in the top right of each plot.

the severity of long return period, low probability, spring/summer droughts.

A rank-based quantile approach was chosen to show the rarity of extreme droughts because it estimates dryness directly from the reconstruction record, without assuming a particular distribution for extreme droughts. Given the uncertainties already present in the reconstructions this provides a simple method while avoiding additional uncertainty associated with fitting a more complete extreme-value model. Fig. 6 shows the four datasets display differences in both drought threshold values and distribution shapes. In terms of distribution shape, *Cooper* and *Wilson* show moderately curved distributions with gradual declines into the lower tail, implying more evenly spread precipitation values and fewer extremely dry events. *Loader* displays a steeper decline in the dry tail, suggesting a heavier skew toward fewer but more extreme droughts. *Cook* also exhibits a sharp decrease towards the extremely negative scPDSI values, reflecting the rarity but intensity of severe droughts. The uncertainty in rank-based estimates increases with the severity of the drought; consequently, the 1:500 empirical return levels are the most uncertain.

Although the overall aim is to use these reconstructions to estimate the intensity of these low probability drought events, regardless of when they occurred, it is nevertheless the case that we might expect the most severe events to coincide. There is multiple independent (i.e. not counting *Cook-Wilson* or *Cook-Cooper* alone) agreement on droughts that are at least as severe as 1:50-year events: 954, 956, 1054, 1326, 1742, 1793, 1818, 1949. The *Loader* dataset consistently records fewer high return period drought events that match the timing in the other series, with only two 1:50-year drought events coinciding with those identified by other datasets (*Loader* cannot corroborate droughts before its start date in 1201). When compared to a historic droughts inventory of supply droughts (spanning from 1890 to 2006) (Durant & Counsell, 2019), *Cook* and *Wilson* both identify droughts in 1921 (1:50-year) and 1976 (1:100-year), which are also recorded by Anglian Water (AW). *Cooper* and *Loader* both identify a 1:50-year drought in 1949 but this event is not recorded by any water company in the inventory. In addition to these events, *Loader* identifies 1:50-year droughts in 1896, 1911 and 1990, together with a 1:100-year drought in 1989. Using a two-year matching window these could correspond to 1898 and 1988 respectively in AW records. The 1911 drought is not recorded by AW but is identified by other water companies. *Wilson* also identifies a 1:50-year drought in 2006 recorded by other water companies but not AW and a 1:100-year drought in 1934 which may correspond to the 1933 drought recorded by AW. Partial agreement between the reconstructions and the historic drought inventory increases confidence that the reconstructed droughts represent hydrologically significant events that affect regional water supply.

6. Accounting for reconstruction error and testing return value consistency with instrumental data

Reconstructing long term hydroclimate variability relies on the assumption that proxy-based estimates can meaningfully reproduce the behaviour of observed climate records. However, it is essential to understand the impact of their uncertainties and calibration methods on the assessment of past drought severity and return period estimates. Two key aspects are how the reconstruction was calibrated and how the reconstruction error is represented, if at all.

There are two common approaches to calibrate an individual reconstruction timeseries. The first (used by *Cooper*: Table 3) is to use 'linear regression'. This results in an underestimation of variance (Table 3), as is well known, and clear from Fig. 2. This underestimation of variance consequently leads to estimates of return values that are substantially too moderate, as demonstrated in Fig. 7 when comparing *Cooper* return values against the instrumental precipitation over their common period. Here we demonstrate that this bias can be overcome by appropriate statistical modelling of the reconstruction error.

The second common approach (used by *Loader* and likely used by *Wilson* though there is some ambiguity in their description: Table 3) is to use 'variance scaling' (Esper et al., 2005), where the reconstruction is scaled to match the variance of the instrumental record during the calibration period. Variance scaling may appear to solve the issue of underestimated variance that hampers linear regression calibration, but it is important to understand that the reconstruction still exhibits error: there are differences from the instrumental data. Adding a statistically modelled representation of this error would actually inflate the variance of the reconstruction too high, so the error is often ignored. The approach also implies that the reconstruction error is correlated with the reconstruction itself, which is not justified.

Our preference is not to replace or reinterpret the original calibration methodologies used by each study, but instead to evaluate them and, where necessary, to statistically model the reconstruction error. It is worth noting here, that process-based models of reconstruction error are not available because, although they are guided by process understanding, the calibration of a tree-ring reconstruction is empirical in nature.

For *Cooper*, we statistically model the reconstruction term using a block-bootstrap (Vogal and Shallcross, 1996) of the residuals (i.e. the differences between the reconstruction and its instrumental target data) using a 5-year block size. We tested this method using block sizes between 2 and 10 years and find our results are not sensitive to this choice. Furthermore, the residual autocorrelation function (Box & Pierce, 1970) shows no lags exceeding 95% significance (ACF < 0.15 at lags 1–30), and Ljung Box tests (Ljung & Box, 1978) confirm no significant joint autocorrelation at any tested horizon. These results indicate that residuals are consistent with white

Table 3
Calibration and error characteristics of the reconstructions, including the SD (mm/season) of the instrumental, reconstructed and random error series with the chosen error modelling approach.

	SD Instrumental	SD Reconstruction	SD error	Error modelling approach	calibration method
Cooper	51	26	46.	Block- bootstrap	Regression
Wilson	60	53	38	Uncorrelated noise	Variance matching
Loader	65	65	44	Uncorrelated noise	Variance matching

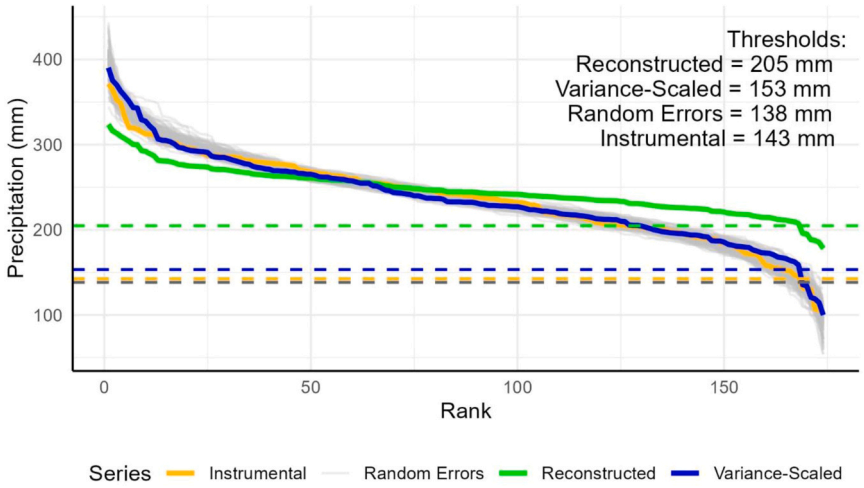


Fig. 7. Ranked values of the *Cooper* raw reconstruction (green), instrumental data (yellow), variance-scaled (blue) and an ensemble of 100 random errors (grey) for the overlapping years of the reconstruction with the instrumental data (all mm/season). The respective 1:30 year drought thresholds are shown by the dashed lines corresponding to their colour and recorded in the top right of the plot.

noise with no strong serial dependence for a block bootstrap to preserve, which explains why there is limited sensitivity to block length. The choice of a 5-year block size was intended to preserve any short-term persistence typically observed in annual hydroclimate reconstructions, while retaining a sufficient number of independent resampling units for robust uncertainty estimation. This statistical model is then used to create an ensemble of 100 realisations of the reconstruction error for the full length of the *Cooper* reconstruction that are added to the reconstruction.

Additionally, to isolate the effect of calibration and error representation on our results we need a case where the same reconstruction is used with both approaches. Therefore, we additionally create a variance-scaled version of *Cooper* by scaling it so that it matches the variance of the associated instrumental data and we compare the return values obtained from the two versions of *Cooper*.

Ordinarily *Loader* and *Wilson* variance-scaled series would result in a reconstruction whose variance matched that of the instrumental data over the calibration period, but their reconstructions show less variance than the instrumental data (Table 3). For *Loader* the difference is very small whereas it is slightly larger for *Wilson*. The reason is not entirely clear, hampered by the incomplete description of the variance scaling used by *Wilson*, but likely arises due to differences between the instrumental data we use for comparison with *Wilson* compared with the data they used (see Section 2.3) and due to differences in the period being compared (we used a longer instrumental record here to maximise the sample size for comparison).

This relatively small underestimation of variance would affect the return values for the drought events, so we have compensated for it by adding 100 realisations of a random error term with the necessary variance to raise the reconstruction variance to match that of the instrumental data. To be clear, we have not abandoned the original authors' preferred calibration method ('variance scaling'), we have kept it but added a small term to compensate for the variance being slightly lower than our instrumental comparator data.

The above approaches are evaluated for *Loader*, *Wilson* and *Cooper* but not for *Cook* (because it is not a precipitation reconstruction). Later in this section and in Section 7 we also use standardised data, which is another way to avoid the effects of reconstructions having different variance to their instrumental targets and there we also include *Cook* in our analysis.

Cooper is used (Fig. 7) as an example because this reconstruction shows the greatest reduction in variance when compared to instrumental records (Table 4; Fig. 2). The four series demonstrate that the raw reconstruction (green), systematically misrepresents both tails because regression-based calibration compresses extremes toward the mean where at the dry end the 1:30-year threshold is elevated (205 mm) compared to the instrumental data (143 mm). Variance scaling (blue) improves this by rescaling the reconstruction, so its overall variance matches the instrumental variance which pulls the threshold down (153 mm), however, since it's a

Table 4
Return period threshold values for a 1:30- year drought for the overlapping years of the Z-score standardised and original-scale data for the respective seasonal yearly average instrumental (precipitation, mm/season) and reconstruction (precipitation, mm/season, or *Cook* PDSI, dimensionless) data. These results do not account for reconstruction error.

1:30 year drought				
	Standardised Z-score		Original-scale	
Dataset	Instrumental	Reconstruction	Instrumental	Reconstruction
Cooper	-2.10	-1.44	150	210
Wilson	-1.87	-1.88	149	165
Loader	-1.76	-2.12	174	150
Cook	-1.64	-1.70	N/A	-3.08

single deterministic adjustment it does not capture how uncertain the reconstruction is at and does not fully close the gap to the instrumental. The random error ensemble (grey) adds error terms calibrated to the reconstruction instrumental residuals which corrects the bias and reproduces uncertainty at each rank and so its threshold (138 mm) is closest to the instrumental. This shows that simply rescaling variance improves on the raw reconstruction but statistically modelling the error term further improves the representation of the overall ranked value curve (the grey ensemble lies around the instrumental data). Including the random error ensemble illustrates the effect of restoring the variance lost through the regression-based calibration rather than providing a complete estimate of reconstruction uncertainty. There is some indication that the reconstruction ensemble produces values that are too high in the upper tail of wet extremes, but that is not a concern for our application.

These comparisons were repeated for *Wilson* and *Loader* (Supplementary Figures S1 and S2); again, this shows that ensuring the reconstruction has the same variance as the instrumental record by including a random-error term brings the ranked values closer to the instrumental series although the impact is smaller because the raw reconstructions do not exhibit the strong variance underestimate seen in *Cooper*. The 1:30 threshold for *Wilson* differs from the instrumental by only 2 mm with the random error term, compared with 20 mm for the raw reconstruction. The *Loader* plot also shows close agreement between the instrumental and the random-error series (5 mm difference), while the raw reconstruction differs from the random errors by 15 mm so although the bias is smaller than in *Cooper* and *Wilson*, a consistent systematic bias remains. Across all three reconstructions, extremes are underrepresented, with reconstruction lines lying below the instrumental and random-error series at the wetter end and above at the drier end.

The alternative approach to dealing with reconstruction error considered here is to use standardised (z-score) data. This avoids the need to assume an error model (e.g. uncorrelated noise or block resampling) and to estimate its parameters, but it does lose the physical units of mm of precipitation. It might be argued that the "variance scaling" calibration method is a form of standardisation that retains the physical units, though equally it could be argued that there is an implicit assumption that the errors are then correlated with the reconstruction. Table 4 shows these issues using the estimated 1:30-year drought severity values from both instrumental and reconstructed datasets, portrayed in both standardised and original-scale formats. The original-scale return values for *Cooper* and, to a lesser extent, *Wilson* show the raw reconstructions underestimate the dryness of the 1:30-year droughts, whereas *Loader* overestimates them (also visible in Fig. 2). Standardised values have mixed agreement between reconstruction and instrumental data. For example, *Wilson* has a very close match (–1.88 vs. –1.87), suggesting strong consistency in drought signal, but the *Cooper* reconstruction shows a less extreme drought than the instrumental record (–1.44 vs. –2.10) and the *Loader* reconstruction indicates a more extreme drought than the instrumental series (–2.12 vs. –1.76). These differences in return levels of standardised data reflect differences in distribution shape rather than differences in mean or variance (e.g. this is apparent in Fig. 7 in the shape of the ranked values curve and in the return value when we applied variance-scaling to *Cooper*).

7. Drought severity differences across drought probabilities

We now bring together multiple reconstructions, including a representation of their errors and using standardised data (to reduce the influence of their errors), to estimate spring/summer drought severity across events with different annual probability of

Table 5
Respective return values (standard deviation units) for different return periods (years) of each dataset for the of the standardised reconstruction without accounting for reconstruction error (upper table) and the median and 90% range of the simulated ensemble under the assumed error model when accounting for reconstructions error using an ensemble of 100 random error series (lower table).

Return Period	Cooper	Wilson	Loader	Cook
10	-1.21	-1.29	-1.31	-1.32
30	-1.73	-1.80	-1.90	-1.93
50	-1.97	-1.98	-2.32	-2.14
100	-2.18	-2.29	-2.57	-2.34
200	-2.32	-2.44	-2.72	-2.53
500	-2.75	-2.55	-2.99	-2.73

Return Period	Cooper		Wilson		Loader	
	Median	90% interval	Median	90% interval	Median	90% interval
10	-1.24	-1.38 to −1.02	-1.28	-1.36 to −1.18	-1.29	-1.38 to −1.18
30	-1.82	-2.19 to −1.55	-1.80	-2.00 to −1.68	-1.87	-2.04 to −1.74
50	-2.04	-2.37 to −1.66	-2.01	-2.22 to −1.83	-2.10	-2.36 to −1.95
100	-2.27	-2.61 to −1.84	-2.27	-2.49 to −2.00	-2.38	-2.81 to −2.14
200	-2.40	-3.02 to −2.00	-2.47	-2.88 to −2.17	-2.64	-3.05 to −2.29
500	-2.52	-3.18 to −2.06	-2.68	-3.18 to −2.37	-2.92	-3.43 to −2.49

[2]Note: *Cook* is not included in lower Table 5 because it is not a precipitation reconstruction so no uncertainty estimates were calculated here.

occurrence. The upper part of Table 5 summarises the drought severity for return periods of 10, 30, 50, 100, 200 and 500 years across the four standardised reconstructed series and indicates how rapidly severity increases with return period (the original scale results are in Supplementary Table S1). The median and 90% range reported in the lower part of Table 5 reflects the spread introduced only by our specific statistically modelled representation of reconstruction error. There are additional overall uncertainties in the climate reconstructions and these are partly represented by using four different reconstructions.

There is broad similarity between the paleo data reconstructions, with 1:10-year values in the range of −1.21 to −1.32 SDs, increasing to the range −2.55 to −2.99 SDs for 1:500-year droughts. *Cooper* and *Cook* show a mid-range increase in severity with return period, similar to *Wilson* until the final increase to 1:500-years (for which *Wilson* shows a smaller step up), while *Loader* displays return values that are more extreme than the others especially at higher return periods. Since the return period intervals get bigger this pattern suggests that growth in drought severity flattens more for *Cooper*, *Wilson* and *Cook* than for *Loader* (confirmed later in Fig. 8).

In contrast, the original scale reconstruction return values (Supplementary Table S1) show high variability in scale and magnitude between datasets, for example, *Cooper* return values decrease gradually from 203 to 159, showing modest sensitivity to return period, but *Loader* drops more steeply, from 211 to 96, suggesting greater amplitude in reconstructed extremes. This variability highlights how reconstruction methods and data standardisation affect the interpretation of extreme events, making cross-dataset comparisons difficult. The differences between the thresholds and number of drought years are also affected by the differing periods of year that the datasets represent and methodological differences in data collection.

Table 5 shows the standardised paleo reconstructions (Table 5 upper) broadly follow the individually standardised random error ensembles (Table 5 lower), and always fall within the ensemble range, often close to the ensemble median. The median error ensemble for *Wilson* and *Loader* has a larger increase between 1:200 and 1:500 empirical return level events than their paleo data counterpart, showing that the reconstructed data may supress extremes due to variance compression. In contrast *Cooper* shows the opposite pattern with larger differences at the lower thresholds. The original-scale random error ensemble median is much drier than the paleo reconstruction for *Cooper*, confirming the bias (seen in Fig. 7) due to the reduced variance of the reconstruction, but also now quantifying and correcting that bias across a wide range of return period events. This is the same case for *Wilson* but to a lesser extent, although the differences grow as the return period increases when the biases in the extreme thresholds become more influential. In contrast *Loader* ensemble produces wetter drought thresholds than the paleo reconstruction and the differences increase towards the more extreme thresholds (see Supplementary Table S1).

Finally, we show (Fig. 8) how estimated drought severity (return values) increases relative to the severity of 1-in-10-year events. The rationale for this is that we can determine the severity of 1-in-10-year events quite accurately from the instrumental record, and then we use the longer tree-ring-based records to estimate how much more severe longer return period droughts might be. To facilitate comparison between the series, we use the z-score standardised versions, and we compare the basic reconstructions with the results (median and 90% range) from the ensembles that incorporate reconstruction error. *Loader* has the largest difference in severity compared to a 1:10-year drought, the total change between a 1:10-year to a 1:500-year drought is −1.67 SDs, the steepest among all datasets. This suggests that extreme droughts are more severe in this dataset and that *Loader* may be more sensitive for capturing extremely rare and intense droughts, possibly due to higher variance or stronger signal resolution. In contrast, the other reconstructions show a shallower gradient with *Cooper* showing a total change of −1.54 SDs, *Cook* −1.41 SDs and *Wilson* −1.26 SDs suggesting a more conservative representation of extremes. In nearly all cases, the median of the reconstruction ensembles matches the basic standardised reconstruction drought severities, suggesting that z-score standardisation may be a suitable alternative to explicitly modelling reconstruction error when standardised severity is sufficient for the application at hand. Even then, the error bars show how sensitive each reconstruction's drought severity estimates are to reconstruction error. As expected, the uncertainty increases with increasing return period because extreme events lay to the tail distribution where there is less data and more dependent on

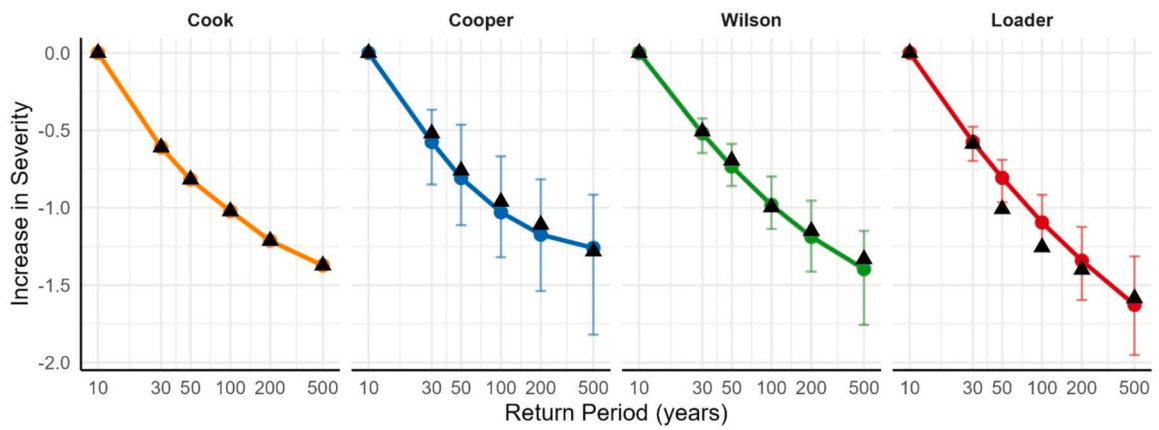


Fig. 8. Visual representation of Table 6 where drought severity for more extreme (longer return period) events is shown as the *increase* in severity compared to a common event severity (1:10 year drought), expressed in standardised Z-scores. Black triangles show the standardised single reconstructions, while the vertical bars for *Cooper*, *Wilson* and *Loader* show the 90% range (5–95%) of return values estimated from the reconstruction error ensembles and the coloured dots are the median of the error ensembles.

assumptions about the error model. Most of the increase occurs between 1:10-year and 1:100-year events, with progressively smaller increments at longer return periods.

8. Summary and discussion

Using four tree-ring based reconstructions of spring and summer hydroclimate for South and East of England, this study provides estimates of the severity of rare drought events beyond the range of the instrumental record, analysis of the standardised reconstructions indicate that:

- A 1:500-year drought is estimated to correspond to approximately 2.6 and 3.0 SDs below the long-term mean. When a representation of reconstruction uncertainty is considered, this range widens to between 2.1 and 3.4 SDs.
- Relative to a 1:10-year drought, the severity of which can be readily estimated from the shorter instrumental records, a 1:500-year drought is estimated to correspond to be approximately 1.4 and 1.7 SDs more severe. When a representation of reconstruction uncertainty is considered, this range widens to between 0.9 and 2.1 SDs.

These estimates of drought severity are conditional upon multiple assumptions including that non-stationarity (i.e. externally forced climate change) has a relatively small influence on precipitation of a small region, that individual reconstruction errors are well represented by our simple statistical models, and that the spread of results between the available reconstructions is a valid indicator of overall reconstruction uncertainty. They are, therefore, presented here as an illustration of our approach and should be considered only as tentative estimates of extreme drought severity.

These findings provide a framework for characterising the severity of rare spring/summer droughts beyond the length of the instrumental record, contributing to a better understanding of long-term hydroclimate variability and the uncertainty associated with estimates of extreme droughts. Such information may also provide useful context for drought risk assessments where instrumental observations alone are insufficient to characterise these rare events, particularly a 1:500-year return period drought. However, several important caveats and opportunities for future research should be considered.

First, the four reconstructions only show modest agreement with one another, reflecting differences in their seasonal and spatial representation, proxy type and reconstruction methodology. In particular, isotope records integrate regional atmospheric signals, whereas ring-width reconstructions are more strongly influenced by local soil moisture availability. While all datasets show broad agreement with the instrumental observations at decadal timescales, differences between reconstructions contribute to uncertainty of estimates of extreme drought severity. The analysis also demonstrates the importance of accounting for reconstruction error. Regression-based reconstruction tends to underestimate the magnitude of dry and wet extremes. Reconstruction uncertainty must therefore be explicitly incorporated into drought characterisation, either through error modelling or the use of standardised datasets as undertaken here, although the latter implicitly assumes that reconstruction errors are correlated with the reconstruction itself.

Second, the reconstructions analysed primarily reflect spring and summer precipitation as tree growth is most sensitive to moisture availability during the growing season. In contrast, water companies are often concerned with winter precipitation and groundwater recharge which governs reservoir replenishment and water availability during the hotter months. Nevertheless, they remain relevant as they provide the only long-term context for hydrological variability and extreme drought occurrence. Prolonged spring-summer drought remains an important component of drought hazard, as illustrated by 2026 when hot, dry conditions during the spring and summer months were the dominant driver of drought.

Alternative approaches are available for estimating the long records required to characterise low-probability drought events. Statistical models, stochastic simulations, and physically based climate and hydrological model ensembles can generate large samples of synthetic hydroclimate data and provide monthly data including the winter recharge period critical for water companies. However, these approaches introduce their own uncertainties, including model biases and limitations in representation of long-timescale variability.

A multi-method approach that combines paleoclimate reconstructions with climate and hydrological modelling might be a useful way forward. Tree-ring reconstructions provide an independent, long-term benchmark in which model simulations can be evaluated against. Their representation of spring and summer hydrological extremes could be assessed against the tree-ring-based results presented here. Greater confidence may be placed in models that successfully reproduce the magnitude and frequency of drought extremes than those that do not. Such an approach would combine the strengths of both methods and improve confidence in estimates of rare drought risk relevant to long-term water resource planning.

CRediT authorship contribution statement

Jane Thurgood: Writing – original draft, Conceptualization. **Timothy Osborn:** Writing – review & editing, Conceptualization.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Jane R Thurgood reports financial support was provided by Anglian Water. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

I am also deeply grateful to my wider supervisory team for their ongoing encouragement and support throughout my PhD journey.

I would like to acknowledge the University of East Anglia Faculty of Science and Anglian Water for funding this research and enabling this work to be carried out.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.ejrh.2026.104135.

Data availability

Datasets DOI shared in data availability in manuscript

References

Barbour, M.M., 2007. Stable oxygen isotope composition of plant tissue: A review. Funct. PLANT Biol. 34 (2), 83–94. https://doi.org/10.1071/FP06228.

Bebchuk, T., Moir, A.K., Arosio, T., Kirdyanov, A.V., Torbenson, M.C.A., Krusic, P.J., Hindson, T.R., Howard, H., Buchwal, A., Norman, C.A.P., Büntgen, U., 2025. Taxus tree-ring chronologies from southern England reveal western European hydroclimate changes over the past three centuries. Clim. Dyn. 63 (2), 108. https://doi.org/10.1007/s00382-025-07601-2.

Box, G.E.P., Pierce, D.A., 1970. Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. J. Am. Stat. Assoc. 65 (332), 1509–1526. https://doi.org/10.1080/01621459.1970.10481180.

Briffa, K.R., Melvin, T.M., 2011. A Closer Look at Regional Curve Standardization of Tree-Ring Records: Justification of the Need, a Warning of Some Pitfalls, and Suggested Improvements in Its Application. Dendroclimatology. Springer, Dordrecht, pp. 113–145. https://doi.org/10.1007/978-1-4020-5725-0_5.

Briffa, K.R., van der Schrier, G., Jones, P.D., 2009. Wet and dry summers in Europe since 1750: Evidence of increasing drought. Int. J. Climatol. 29 (13), 1894–1905. https://doi.org/10.1002/joc.1836.

Cook, E.R., Briffa, K.R., Meko, D.M., Graybill, D.A., Funkhouser, G., 1995. The 'segment length curse' in long tree-ring chronology development for palaeoclimatic studies. Holocene 5 (2), 229–237. https://doi.org/10.1177/095968369500500211.

Cook, E.R., Seager, R., Kushnir, Y., Briffa, K.R., Büntgen, 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., Čufar, K., Dittmar, C., Zang, C., 2015. Old World megadroughts and pluvials during the Common Era. Sci. Adv. 1 (10), e1500561. https://doi.org/10.1126/sciadv.1500561.

Cooper, R.J., Melvin, T.M., Tyers, I., Wilson, R.J.S., Briffa, K.R., 2013. A tree-ring reconstruction of East Anglian (UK) hydroclimate variability over the last millennium. Clim. Dyn. 40 (3), 1019–1039. https://doi.org/10.1007/s00382-012-1328-x.

Darwish, M.M., Tye, M.R., Prein, A.F., Fowler, H.J., Blenkinsop, S., Dale, M., Faulkner, D., 2021. New hourly extreme precipitation regions and regional annual probability estimates for the UK. Int. J. Climatol. 41 (1), 582–600. https://doi.org/10.1002/joc.6639.

Deser, Clara, Phillips, Adam, Bourdette, Vincent, Teng, Haiyan, 2010. Uncertainty in climate change projections: the role of internal variability. Climate Dynamics 38 (3-4), 527–546. https://doi.org/10.1007/s00382-010-0977-x. http://link.springer.com/10.1007/s00382-010-0977-x. (Accessed 31 December 2010).

Duffy, J.E., McCarroll, D., Loader, N.J., Young, G.H.F., Davies, D., Miles, D., Bronk Ramsey, C., 2019. Absence of age-related trends in stable oxygen isotope ratios from oak tree rings. Glob. Biogeochem. Cycles 33 (7), 841–848. https://doi.org/10.1029/2019GB006195.

Durant, M., & Counsell, C. (2019, October 7). *Historic droughts inventory of drought references to water resources 1890-2015* [Data Collection]. UK Data Service. https://doi.org/10.5255/UKDA-SN-853673.

Environment Agency. (2025). *Water company drought plan guideline, 2025*. GOV.UK. https://www.gov.uk/government/publications/water-company-drought-plan-guideline-2025/water-company-drought-plan-guideline-2025.

Esper, J., Frank, D.C., Wilson, R.J.S., Briffa, K.R., 2005. Effect of scaling and regression on reconstructed temperature amplitude for the past millennium. Geophys. Res. Lett. 32 (7). https://doi.org/10.1029/2004GL021236.

Esper, J., George, S.St, Anchukaitis, K., D'Arrigo, R., Ljungqvist, F.C., Luterbacher, J., Schneider, L., Stoffel, M., Wilson, R., Büntgen, U., 2018. Large-scale, millennial-length temperature reconstructions from tree-rings. Dendrochronologia 50, 81–90. https://doi.org/10.1016/j.dendro.2018.06.001.

Gagen, M., McCarroll, D., Loader, N.J., Robertson, I., 2011. Stable Isotopes in Dendroclimatology: Moving Beyond 'Potential'. Dendroclimatology. Springer, Dordrecht, pp. 147–172. https://doi.org/10.1007/978-1-4020-5725-0_6.

García-Suárez, A.M., Butler, C.J., Baillie, M.G.L., 2009. Climate signal in tree-ring chronologies in a temperate climate: A multi-species approach. Dendrochronologia 27 (3), 183–198. https://doi.org/10.1016/j.dendro.2009.05.003.

Hand, Ralf, Samakinwa, Eric, Lipfert, Laura, Brönnimann, Stefan, 2023. ModE-Sim – a medium-sized atmospheric general circulation model (AGCM) ensemble to study climate variability during the modern era (1420 to 2009). Geoscientific Model Development 16 (16), 4853–4866. https://doi.org/10.5194/gmd-16-4853-2023. https://gmd.copernicus.org/articles/16/4853/2023/. (Accessed 28 August 2023).

Hawkins, Ed, Sutton, Rowan, 2009. The Potential to Narrow Uncertainty in Regional Climate Predictions. Bulletin of the American Meteorological Society 90 (8), 1095–1108. https://doi.org/10.1175/2009bams2607.1. https://journals.ametsoc.org/doi/10.1175/2009BAMS2607.1.

Hollis, D., McCarthy, M., Kendon, M., Legg, T., Simpson, I., 2019. HadUK-Grid—A new UK dataset of gridded climate observations. Geosci. Data J. 6 (2), 151–159. https://doi.org/10.1002/gdj3.78.

Li, Y., Wang, S., Niu, J., Fang, K., Chao, Y., Li, X., Li, Y., 2016. Tree-ring-based reconstruction of drought variability (1792–2011) in the middle reaches of the Fen River, North China. Dendrochronologia 40, 1–11. https://doi.org/10.1016/j.dendro.2016.05.001.

Ljung, G.M., Box, G.E.P., 1978. On a measure of lack of fit in time series models. Biometrika 65 (2), 297–303. https://doi.org/10.1093/biomet/65.2.297.

Loader, N.J., Young, G.H.F., McCarroll, D., Davies, D., Miles, D., Bronk Ramsey, C., 2020. Summer precipitation for the England and Wales region, 1201–2000 ce, from stable oxygen isotopes in oak tree rings. J. Quat. Sci. 35 (6), 731–736. https://doi.org/10.1002/jqs.3226.

Martin, J.T., Pederson, G.T., Woodhouse, C.A., Cook, E.R., McCabe, G.J., Anchukaitis, K.J., Wise, E.K., Erger, P.J., Dolan, L., McGuire, M., Gangopadhyay, S., Chase, K. J., Littell, J.S., Gray, S.T., St. George, S., Friedman, J.M., Sauchyn, D.J., St-Jacques, J.-M., King, J., 2020. Increased drought severity tracks warming in the United States' largest river basin. Proc. Natl. Acad. Sci. 117 (21), 11328–11336. https://doi.org/10.1073/pnas.1916208117.

Mašek, J., Tumajer, J., Rydval, M., Lange, J., Treml, V., 2021. Age and size outperform topographic effects on growth-climate responses of trees in two Central European coniferous forest types. Dendrochronologia 68, 125845. https://doi.org/10.1016/j.dendro.2021.125845.

McCarroll, D., Loader, N.J., 2004. Stable isotopes in tree rings. Quat. Sci. Rev. Isot. Quat. Paleoenviron. Reconstr. 23 (7), 771–801. https://doi.org/10.1016/j.quascirev.2003.06.017.

Murphy, C., Wilby, R.L., Matthews, T.K.R., Thorne, P., Broderick, C., Fealy, R., Hall, J., Harrigan, S., Jones, P., McCarthy, G., MacDonald, N., Noone, S., Ryan, C., 2020. Multi-century trends to wetter winters and drier summers in the England and Wales precipitation series explained by observational and sampling bias in early records. Int. J. Climatol. 40 (1), 610–619. https://doi.org/10.1002/joc.6208.

New, M., Hulme, M., Jones, P., 2000. Representing twentieth-century space–time climate variability. part ii: development of 1901–96 monthly grids of terrestrial surface climate. J. Clim. 13 (13), 2217–2238. https://doi.org/10.1175/1520-0442(2000)013%3C2217:RTCSTC%3E2.0.CO;2.

Rice, J.L., Woodhouse, C.A., Lukas, J.J., 2009. Science and decision making: water management and tree-ring data in the Western United States. JAWRA J. Am. Water Resour. Assoc. 45 (5), 1248–1259. https://doi.org/10.1111/j.1752-1688.2009.00358.x.

Roden, J.S., Lin, G., Ehleringer, J.R., 2000. A mechanistic model for interpretation of hydrogen and oxygen isotope ratios in tree-ring cellulose. Geochim. Et. Cosmochim. Acta 64 (1), 21–35. https://doi.org/10.1016/S0016-7037(99)00195-7.

Saurer, M., Borella, S., Schweingruber, F., Siegwolf, R., 1997. Stable carbon isotopes in tree rings of beech: Climatic versus site-related influences. Trees 11 (5), 291–297. https://doi.org/10.1007/s004680050087.

Smerdon, J.E., 2012. Climate models as a test bed for climate reconstruction methods: Pseudoproxy experiments. WIREs Clim. Change 3 (1), 63–77. https://doi.org/10.1002/wcc.149.

Treydte, K., Boda, S., Graf Pannatier, E., Fonti, P., Frank, D., Ullrich, B., Saurer, M., Siegwolf, R., Battipaglia, G., Werner, W., Gessler, A., 2014. Seasonal transfer of oxygen isotopes from precipitation and soil to the tree ring: Source water versus needle water enrichment. New. Phytol. 202 (3), 772–783. https://doi.org/10.1111/nph.12741.

Valler, V., Franke, J., Brugnara, Y., Samakinwa, E., Hand, R., Lundstad, E., Burgdorf, A.-M., Lipfert, L., Friedman, A.R., Brönnimann, S., 2024. ModE-RA: a global monthly paleo-reanalysis of the modern era 1421–2008. Sci. Data 11 (1), 36. https://doi.org/10.1038/s41597-023-02733-8.

Vogel, Richard M., Shallcross, Amy L., 1996. The moving blocks bootstrap versus parametric time series models. Water Resources Research 32 (6), 1875–1882. https://doi.org/10.1029/96wr00928. https://agupubs.onlinelibrary.wiley.com/doi/10.1029/96WR00928. (Accessed 9 July 2010).

Wigley, T.M.L., Briffa, K.R., Jones, P.D., 1984. On the average value of correlated time series, with applications in dendroclimatology and hydrometeorology. J. Clim. Appl. Meteorol. 23 (2), 201–213. https://doi.org/10.1175/1520-0450(1984)023%3C0201:OTAVOC%3E2.0.CO;2.

Wilson, R., Miles, D., Loader, N.J., Melvin, T., Cunningham, L., Cooper, R., Briffa, K., 2013. A millennial long March–July precipitation reconstruction for southern-central England. Clim. Dyn. 40 (3–4), 997–1017. https://doi.org/10.1007/s00382-012-1318-z.

Woodhouse, C.A., Pederson, G.T., Morino, K., McAfee, S.A., McCabe, G.J., 2016. Increasing influence of air temperature on upper Colorado River streamflow. Geophys. Res. Lett. 43 (5), 2174–2181. https://doi.org/10.1002/2015GL067613.

Yang, B., Qin, C., Wang, J., He, M., Melvin, T.M., Osborn, T.J., Briffa, K.R., 2014. A 3,500-year tree-ring record of annual precipitation on the northeastern Tibetan Plateau. Proc. Natl. Acad. Sci. USA 111 (8), 2903–2908. https://doi.org/10.1073/pnas.1319238111.

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., Whitney, M., 2015. Oxygen stable isotope ratios from British oak tree-rings provide a strong and consistent record of past changes in summer rainfall. Clim. Dyn. 45 (11), 3609–3622. https://doi.org/10.1007/s00382-015-2559-4.