RC 3

General comments

This paper combines a vast amount of paleoclimate information consisting of different proxies such as tree rings and documentary materials to reconstruct annual and rainy season precipitation across Asia in a grid of 2.5 degrees over the past 300 years. Reconstructions of precipitation in such a wide area has been done several times in the past, but in this paper, not only new data that has not been reflected in previous studies has been added, but also a method called GLDD, which is a new method of sorting proxy data, is effectively used, and a very accurate reconstruction is realized. Therefore, the usefulness of the obtained data set is very large.

On the other hand, with regard to the evaluation of the results of this paper and future issues, several points that should be added or improved were recognized as follows.

1) GLDD is considered a very good proxy sorting method. In fact, this method can be used to reconstruct paleoclimate fairly accurately even in areas where proxy data is not present, such as the eastern part of the Caspian Sea. This means that, in some places, the "searching region" of the grid is very large. In order to help readers better understand the effectiveness of GLDD, it would be good to have a map showing the size of the "searching region" for each grid.

Accepted and revised. We add the maximum and minimum distances from the boundary of the searching region to each target grid in section 3.3 with Figure 12a-c as following (Line 341-352) :

As shown in Fig. 17, there are evident difference in the maximum (Fig. 17a) and minimum distances (Fig. 17b) from the boundary of the searching region to each target grid across Asia. The maximum distance is 1000~2000 km for most grids in China, Mongolia, and central and northwest Russia, 2000~3000 km for most grids in India, central Asia, and southwest and eastern Russia, 3000~4000 km in Arabian Peninsula, and more than 4000 km for tropical islands. However, the minimum distance is only 250~750 km for most grids of study area, except very few grids in tropics. The difference between maximum and minimum distances (Fig. 17c) could reach 2000 km or more in regions with high topographic complexity, which means that the searching region is always in an irregular shape. Thus, for the area (e.g. the Tibetan Plateau and surrounding area) with complicated topography and multiplex hydroclimate variation, GLDD could identify the unique searching region (including shape and size) rigorously for each target grid. While for the area with homogeneous hydroclimate variability and rainfall regime, GLDD could capture the proxies far from the target grid, which could reconstruct well at the areas where proxy data are not present, such as the east of the Caspian Sea.

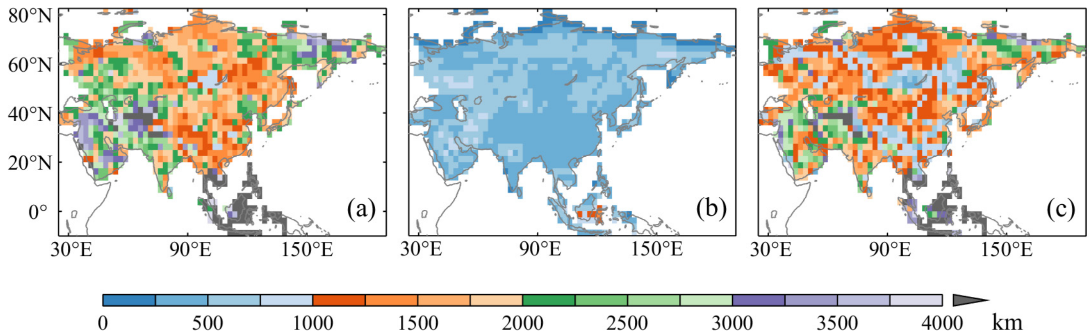

Figure 17: The maximum (a) and minimum (b) distance from boundary of the searching region to the target point, and their difference (c) for each grid.

2) In Figs. 5 ~ 8, the accuracy of the SPI reconstruction results in eastern China is very high. This is clearly because DWI is mainly used in eastern China. Using DWI is very important. However, unlike tree rings, the quality and quantity of the documentary material on which the DWI is based has been greatly improving as the times become newer, and from the 20th century, meteorological observation data itself should be included in the documentary materials. Since the calculations in Figs. 5 ~ 8 are based on the comparison of meteorological observation data and proxy data from 1948 CE, it is obvious that the SPI reconstruction accuracy will be higher in areas where DWI is used as a proxy than in areas where only tree rings are used, and it may not reflect the actual reconstruction accuracy of the SPI from 200 years ago or 300 years ago. There needs to be a mention of that possibility. In the same sense, since the observation and reconstruction data included in Fig. 10 are both affected by the meteorological observation data, it is obvious that the two coincide.

Accepted and revised. We discussed the issue in section 3.3 as following:

Thirdly, for documentary proxies, DW120 may use instrumental precipitation data to identify the dryness/wetness grades since 1951, especially after 1979, which might lead to overestimation of the calibration and verification metrics in eastern China. Fortunately, since the data of dryness/wetness grades before 1950 are completely derived from historical documents (Wang and Zhao, 1979), comparison between the Fig. 12-13 and Figs 6-7 by each site-grid could help us to assess the overestimation, which shows that the overestimation of R^2a in this reconstructions is about 10% in average over eastern China (Line 368-374).

3) In this paper, calculations excluding tree-ring data that show negative correlations with precipitation are performed in Figs 5 and 7. In arid regions, there should be certainly a positive correlation between precipitation and tree ring width, but in humid regions, there is usually a negative relationship of precipitation with temperature and/or solar radiation, so it is rather common for there to be a negative correlation between precipitation and tree ring width. In fact, as shown in Figs 6 and 8, results using tree ring data that show negative correlations with precipitation are much better than excluding it. In other words, I do not understand the meaning of calculating Figs. 5 and 7.

Accepted and revised. This is why we reconstruct the SPI data with version A (for only including the positive correlation tree ring width as candidate proxies) and version B (for including all tree ring data as candidate proxies), and we also rewrite the related section to show this issue more clearly as followings:

According to the principle of dendroclimatology, soil water affects the growth rate and formation of wood, both within a season and the longer terms, thus, tree-ring width is expected to be positively correlated with precipitation via this direct response (Vaganov et al., 2011; Wettstein et al., 2011). For tree-ring density chronologies, they are usually correlated with temperature variation and scarcely used in precipitation reconstruction (Briffa et al., 2002). However, due to multiple types of climate and complex topography in the vast study area, the tree-ring density chronologies and width chronologies with negative correlations to precipitation may also well indicate precipitation variation (George, 2014) and the use of such tree-ring predictors in hydro-climate reconstruction has been discussed in previous studies (e.g. Cook et al., 2020). Therefore, we reconstruct two sets datasets of SPI, one excludes tree-ring width chronologies negatively correlated to precipitation and tree-ring density chronologies (hereafter called as "Version A"), the other includes all tree-ring chronologies (here after called "Version B"). (Line 126-135)

4) In this paper, the accuracy of the reconstruction of SPI in eastern China was improved by using DWI based on document materials. This is a very good thing, but there are a lot of documents related to weather since 1700 CE in Japan and elsewhere. In this paper, authors also use the oxygen isotope ratio of tree rings, which are highly correlated with precipitation. Since the data on the oxygen isotope ratio of tree rings has increased rapidly in recent years, it is expected that the results of this study will be further improved if calculations based on GLDD are performed by incorporating such document materials from other regions and data on the tree ring oxygen isotope ratio. It might be good to have such a comment in the text.

Accepted and revised. (1) We add the series of wet-season (May to October) rainy days for 5-site in Japan for reconstruction. The information is added in section 2.3 as follow:

In addition, the series of wet-season (May to October) rainy day for 5-site in Japan are also included. These series were extracted from the historical diaries (https://www.ncei.noaa.gov/access/paleo-search/study/5412) and merged with instrumental data (Kamiguchi et al., 2010) by the method from Murata (1992). (Line 158-161)

(2) For tree-ring oxygen isotope data, we have included all available chronologies from WDC-P and add some chronologies from recent publications. The related information is added in the section 2.3 as follows :

Tree-ring data are mainly (2772) from the International Tree Ring Data Bank (ITRDB), maintained by the World Data Center for Paleoclimatology (WDC-P, https://www.ncei.noaa.gov/products/paleoclimatology), including 1854 tree-ring width records, 828 tree-ring density records, 67 tree-ring latewood percent records, 22 tree-ring stable oxygen isotope ($\delta^{18}O$) and 1 tree-ring blue intensity record. (Line 104-106)

Besides ITRDB, 17 tree-ring width chronologies and 3 tree-ring $\delta^{18}O$ chronologies that indicate local precipitation or drought from recently published papers are included in our study (Shah et al., 2007; Sass-Klaassen et al., 2008; Arsalani et al., 2018; Arsalani et al., 2015; Chen et al., 2016; Zhang et al., 2017; Pumijumnong et al., 2020; Xu et al., 2015; Buckley et al., 2017; Ukhvatkina et al., 2021; Akkemik et al., 2020; Kostyakova et al., 2017; Kucherov, 2010; Xu et al., 2013; Borgaonkar et al., 2010). (Line 119-125)

Specific comments

Lines 95-101: In this paper, authors utilize raw tree-ring data, instead of using published tree ring chronologies, and process them according to author's own method. However, the reader cannot judge the effect of the data processing on the results. I would appreciate it if authors discuss in some way whether or not there is an impact from the processing of data?

Accepted and added. We utilize raw tree-ring data because published tree ring chronologies are standardized by various methods and their expressed population signal (EPS) values are not available. We use running mean technique (Altman et al., 2014) to identify disturbance and age-dependent splines (Melvin et al., 2007) to remove growth trend, which are commonly used methods in tree-ring studies and we do not modify them. We add discussions on the potential impact of tree-ring processing on the reconstruction in section 3.3 as following:

Second, for tree-ring proxies, we use the same standardization method to build chronologies when raw measurements are available. However, about 4.5% of the tree-ring proxies do not have raw measurement file thus we have to use the processed chronologies with various standardization methods by different data providers. As the test for some sites, the difference between chronologies could reach 20% in maximum from different standardization methods (Li et al., 2011). This may also induce the uncertainty in the reconstruction. (Line 364-368)

Line 174: What is the specific reason for the decision to limit the number of proxies used for calculations to 5?

Accepted and added. This is because that the sample length to develop calibration equations for reconstruction is about 50 years usually, and the sample size should be preferably 10 times (or more) the number of variables for BSR according to the principle of statistics (Sekaran, 2003). (line 209-211)

Lines 177-178: 1942 and 1943 are mistakes for 1742 and 1743.

Accepted and revised.

Figure 9: In this figure, the meaning of indicating the area where the correlation coefficient is negative can be understood. But the meaning of classifying the area where the correlation coefficient is positive by the p-value is difficult to understand. In fact, India's low p-values only mean that the period of meteorological observations there is long before 1948 CE, and I don't think it is relevant to the discussion in this figure.

Accepted and add. We rewrite that paragraph with adding information as follows:

(1) Noted that the length of instrumental data before 1948 varies for different weather station, i.e., the degree of freedom for calculating these correlations are different station by station, we show the significant level for all positive correlations station by station instead of the value of correlation coefficient directly with levels of $p \leq 0.01$, $0.01 < p \leq 0.05$, $0.05 < p \leq 0.1$ and $p > 0.1$ (Fig. 12). (Line 284-287)

While the low p-values (i.e. $p>0.1$) of the positive correlation or negative correlation in part of stations (e.g., in India, southeastern Asia, etc.) might be induced by both uncertainties from reconstructions based on few proxies available and observations in early times, because the instrumental data from these station usually extend to the 1880 and before (e.g., several stations in India extend to 1836) with missing records and using defective rain gauge in early time frequently (GHCNm, 2022). (Line 294-297)