

Reply to RC 2

Overview

The manuscript by Chen and Bronnimann develops a Bayesian framework to reconstruct winter temperature in Southwestern China since 1700 CE. The methodology combines multi-archive documentary evidence with prior distributions computed from climate simulation ensembles. The methodology and results are innovative, valuable, and fit very well into the scope of *Climate of the Past*. As detailed in my comments below, there are some methodological aspects that I ask the authors to improve. After resolving these issues, the manuscript could be suitable for publication in *Climate of the Past*.

Response: Thank you for your positive comment of the novelty and value of our manuscript, as well as for your many insightful comments and suggestions. We will revise this manuscript based on your comments to improve the clarity of the methodology. Below, we provide point-by-point responses to your specific comments.

General comments

1. Presentation of observations:

The manuscript provides a limited overview over the number and spatio-temporal distribution of observations included in the reconstructions. Could you make a visualization of the spatio-temporal characteristics of the observations (e.g., time series with the number of observations per year, a map with all observation locations and how many observations they contain) and the number of observations available for each observation type?

Response: Thank you for your comments and suggestions. Regarding the spatio-temporal distribution of the observations, this study primarily uses two types of documentary records. The first type consists of disaster and anomaly records mainly from local gazetteers. These records generally have a spatial resolution at the county- or prefecture-level, and each entry typically covers only a limited number of locations. We will provide the spatiotemporal distribution of the number of such records in the final Supplementary Material. The second type consists of precipitation records from *Yu-Xue-Fen-Cun* reports. For *Yu-Xue-Fen-Cun*, we have partly extracted from the original report and partly used the records provided by Yang (2006). And unlike local gazetteer, *Yu-Xue-Fen-Cun* reports of southwestern China often include detailed descriptions of precipitation conditions in provincial capitals, as well as more general summaries for other areas. Their spatial resolution is therefore more complex. We will provide the number of winters with precipitation information for provincial capitals.

Reference

Yang, Y.: Research on Monsoon Climate and Weather Disasters in Yunnan during the Qing Dynasty. Fudan University Press, 2006.

2. Construction of the likelihood:

The advantage of the Bayesian model to naturally incorporate uncertainties in the likelihood function into the posterior is only useful if the construction of the likelihood is reproducible and can be critically assessed by peers. It is currently unclear to what extent this construction follows objective criteria and to what extent it is based on a subjective assessment of the documentary evidence. Can you clarify how Table 1, which contains a deterministic association of CWI and observed phenomena, and Table 2 are related? Table 2 seems to be the most important to understand the likelihood construction. To make an assessment by peers possible, a similar table with information for each year should be included in the supplement or data/code repository.

It is stated that uncertainties in the documentary data such as source reliability, record quantity, and descriptive accuracy, have been incorporated in the likelihood (l. 215-216). Can you document how this is done? Potentially, the above mentioned table can clarify this if extended to all years. In general, consider moving Sect. 4.2.1 to the methods section as it clarifies several of the methodological questions I had while reading the methods section. In addition, it should appear prior to presenting the CWI reconstruction in Sect. 4.1.

Finally, some of the observed phenomena relate to short-term (extreme) events. To what extent are winter mean conditions correlated with these events? How strongly are anomalous conditions in one subregion associated with the regional mean winter temperature?

Response: Thanks for your comment and your suggestions. Our presentation of the process of estimating likelihood remains some unclear.

- ◆ Regarding the estimation of the likelihood, we took two aspects into consideration. First, if we do not consider the uncertainty associated with documentary sources and assume that all historical records are equally reliable, the likelihood only indicate the dependence of observations on climatic conditions. That is, how different climatic conditions shape the occurrence of different types of observations. For example, extremely cold winter is highly likely to produce corresponding records of extreme events (such as heavy snowfall), but this is not a necessary condition, since ordinary winters may also generate short-term extreme event records. For this part, we used observations and instrumental data during 1950–1999 to calculate the frequency of different observations under different winter temperature conditions, making this component relatively objective. However, it should be noted that documentary records from different sources vary in reliability. For example, a brief record from local gazetteer and a detailed report in *Yu-Xue-Fen-Cun* may both document a heavy snowfall event in Kunming, yet the reliability of the information they provide is not same. This is one of the characteristics

distinguishing documentary evidence from many natural proxies, and it also highlights the advantage of the Bayesian approach applied in documentary-based reconstructions. In this manuscript, this component follows the evaluation framework proposed by Ge & Zhang (1990) and is assessed based on expertise, and therefore involves a certain degree of subjectivity. The likelihood estimate takes both two parts into considerations.

- ◆ In response to your question, “to what extent this construction follows objective criteria and to what extent it is based on a subjective assessment of the documentary evidence,” we will add relative content to assess how the reconstruction results would change when only considering the dependence of observations on climatic conditions.
- Regarding the relationship between Tables 1 and 2, Table 1 presents the criteria used to generate temperature indices based on documentary evidence, which is same as the conventional index approaches. Table 2 presents the estimation of the likelihood based on documentary evidence, which is the core component of our methodology. We will include a complete table similar to Table 2 in the final supplementary materials.
- Regarding the Section 4.2.1, considering the Section 4.2.1 is a very comprehensive section, and the likelihood estimation could only be done after the CWI reconstruction, we will add more detailed explanation of the likelihood estimation in the methodology section, and we will also add a flowchart to present the whole procedure more clearly.
- In this manuscript, we use observations from 1950 to 1999 and classify these events by sub-region and intensity level to evaluate the extent to which extreme events in different sub-regions can indicate regional winter mean temperature (Fig. S3). The results show that the indicative power of extreme events for regional mean winter temperature varies across sub-regions and event intensities. For example, a mild frozen rain event occurring in subregion I corresponds to a standardized deviation of regional winter mean temperature between 0 and -0.5σ , whereas a severe frozen rain event corresponds to a standardized deviation below -1σ .

Reference:

Ge Q. & Zhang, P.: The Evaluation of Climatic Information in the Historical Literatures. *ACTA GEOGRAPHICA SINICA*, 45(1). <https://doi.org/10.11821/xb199001003>, 1990.

3. Construction of the prior

How are the winter temperatures in the simulations split into CWI classes? Are 20 members sufficient to create a stable prior or should there be some smoothing applied to ensure that the prior has a sufficient spread (similar to a variance inflation factor in ensemble Kalman filtering)? Regarding the time-independent prior, I’m unsure about its purpose. Should it create a reconstruction with minimal influence from the temporal structure of the simulations (in this case, it would be the most convincing to take the same prior for each year, which could simply be the class probabilities across all simulated

winters), or should it assess sampling uncertainties from the small ensemble size (in this case, I'd recommend to repeat the construction of CWI-ModE-Clim multiple times to quantify how results change for different realizations).

Response: Thanks for your comment and your suggestions.

- Regarding the comment, “How are the winter temperatures in the simulations split into CWI classes?”, we first calculated the temperature anomalies for all 20 members in the period from 1700 to 1949 in ModE-Sim. Then, we calculated the distribution of the temperature anomalies for all years of all members and the distribution of CWI. Next, we used quantile-based matching between the distributions of CWI and temperature anomalies from simulations to obtain the temperature anomalies range corresponding to each CWI. Based on the temperature anomalies range corresponding to each CWI, we converted the temperature anomalies of each winter in each member into the a CWI. To make this process clearer, we will modify Fig 5a in the revised manuscript and add explanation in Section 3.3.1.
- Regarding the ensemble size, a previous study using ensemble square root filtering (EnSRF) provides a useful reference (Bhend, J. et al., 2012). That study employed 37 pseudo-proxy temperature time series to constrain an ensemble of 29 atmosphere-only GCM simulations and investigated the impact of ensemble size on assimilation skill. The results indicated that increasing the ensemble size from 20 to 29 did not lead to a significant improvement in assimilation skill.
- Regarding CWI-ModE-Clim, our purpose is to compare the respective contributions of documentary evidence and the model-based prior in the assimilation results. When generating the CWI-ModE-RA, we used the 20 members of ModE-Sim as the prior for each winter. When generating each winter in CWI-ModE-Clim, we randomly selected 20 values from all members and all winters in ModE-Sim as the prior and this procedure is repeated 100 times. As a result, CWI-ModE-Clim no longer preserves the temporal structure of ModE-Sim and instead contains only information derived from the documentary evidence. Since the term “time-independent prior” may be misleading, we will no longer use this term in the revised manuscript.

Reference:

Bhend, J., Franke, J., Folini, D., Wild, M., and Brönnimann, S.: An ensemble-based approach to climate reconstructions, *Clim. Past*, 8, 963–976, <https://doi.org/10.5194/cp-8-963-2012>, 2012.

4. Computation of the posterior CWI and temperature

In my understanding of the Bayesian model, you first compute the posterior CWI distribution by multiplying the prior probabilities of each class with the likelihood of each class according to $p(\text{evidence}|\text{CWI} = n)$, and normalizing the resulting product. In a second step, the posterior mean of $p(T_{\text{rec}}|\text{evidence})$ is computed by computing the weighted mean of the mean temperatures of each class, where the weights are given by

the probabilities of the class. If that's correct, I would rewrite Eq. 4 such that it only contains $p(CWI|evidence)$ and only introduce the temperatures in the subsequent paragraph. Furthermore, the posterior distribution of T_{rec} should therefore be given by a mixture distribution

$$T_{rec}|evidence = \sum_{n=1}^6 p(T_{rec}|CWI = n) \cdot p(CWI = n|evidence)$$

where $p(T_{rec}|CWI = n)$ is the temperature distribution corresponding to the n-th CWI class. The posterior mean of that distribution is equal to Eq. 5 in the manuscript, but the credible intervals of that mixture distribution cannot be computed by simply computing the weighted mean of the quantiles of the temperatures of each class. In fact, this strategy will very likely underestimate the posterior uncertainty. Instead the posterior quantiles need to be computed from the posterior cumulative distribution function (CDF) $F(T_{rec}|evidence)$ given by

$$F(T_{rec}|evidence) = \sum_{n=1}^6 F(T_{rec}|CWI = n) \cdot p(CWI = n|evidence)$$

where $F(T_{rec}|CWI = n)$ is the CDF of the n-th temperature class. Alternatively, it can be obtained through Monte Carlo sampling from the posterior distribution of T_{rec} (Eq. 1). In both cases, $p(T_{rec}|CWI = n)$ needs to be specified, e.g., by using its empirical distribution or a parametric approximation of it.

Response: Thank you for this insightful and technically important comment. We agree with your interpretation of the Bayesian framework and the formulation of the posterior CWI distribution. In the revised manuscript, we will explicitly formulate the posterior probability of CWI $p(CWI | evidence)$, and modify Eq. (4) to focus solely on the posterior distribution of CWI. The calculation of temperature will then be introduced subsequently as a mixture distribution conditioned on CWI classes as you suggested. Furthermore, we acknowledge that the posterior Temperature uncertainty cannot be accurately estimated by computing weighted mean of the quantiles of the temperatures of each class. In the revised manuscript, we will therefore recompute credible intervals accordingly.

Specific comments:

1. Title: Given the wide range of available paleoclimate data types, the title would become more meaningful if the use of documentary evidence was mentioned.

Response: Thanks for your comment. We will revise our title to include “documentary records” to highlight the integration of documentary records.

2. Abstract: Using a Bayesian approach says fairly little about the underlying statistical model because almost any statistical model can be formulated in Bayesian way.

Therefore, a little more information about the statistical model would be useful in the abstract, in particular how documentary evidence and climate simulations are combined in the framework.

Response: Thanks for your comment and helpful suggestion. We will revise the abstract to better describe the methodological framework. We will briefly clarify how documentary evidence and climate simulations are combined within a Bayesian framework. Specifically, documentary evidence is used to update the prior distribution of CWI derived from climate simulations, yielding a posterior distribution of CWI.

3. l. 22: To what time period does the "data-sparse" refer? In general, it would be helpful to specify the period of interest in the first paragraph.

Response: Thanks for your comment and nice suggestion. The "data-sparse" is used in a general, comparative sense relative to eastern China. It refers to the overall lower availability of documentary records for climate reconstruction in southwestern China during historical periods, which is mainly due to its geographical distance from political centers and its lower population density. We will revise the sentence to clarify this expression.

4. l. 33: Can you provide an order of magnitude for how much more data is available in eastern China compared to southwestern China?

Response: Thanks for your comment.

- ◆ For local gazetteers, according to the *Integrated Natural Disaster Information System of the Qing Dynasty* (Fang et al., 2020), which includes records of all types of natural disasters, each province in southwestern China contains approximately 2,000 records on average throughout the Qing Dynasty (1644–1911), whereas provinces in eastern China such as Jiangsu, Zhejiang, and Anhui each contain more than 5,000 records. Considering that winters in southwestern China are relatively warm, the difference in the number of winter-related records may be even greater.
- ◆ As for the *Yu-Xue-Fen-Cun* reports, although they are distributed across 18 provinces in China, their spatial distribution is uneven. For example, Hebei Province in North China, which was directly governed by the central government during the Qing Dynasty, contains more than 13,000 records. In contrast, provinces such as Guangdong, Guangxi, Liaoning, and the southwestern provinces (Guizhou, Sichuan, and Yunnan) contain only about 4,000 records (Hao et al., 2018).

Reference

Fang, X., Zhao, W., Song, R., Chen, S., and Ye, Y.: The integrated natural disaster information system of the Qing Dynasty, *Soc. Hist. Res. J.*, 29–46, 2020.

Hao, Z., Yu, Y., Ge, Q., & Zheng, J.: Reconstruction of high-resolution climate data over China from rainfall and snowfall records in the Qing Dynasty, *WIREs Climate Change*, 9(3), e517. <https://doi.org/10.1002/wcc.517>, 2018.

5. l. 44: To emphasize the causal structure of evidence, consider replacing "differ from reality" by "reverse the causal structure of reality".

Response: Thanks for your suggestions. We will modify it in the revised manuscript.

6. l. 49-50: There are plenty of applications of Bayesian frameworks in paleoclimatology. Is there a reason to highlight the study by Camenisch et al. 2022?

Response: Thanks for your comment. There are numerous applications of Bayesian approaches in paleoclimatology. We highlight Camenisch et al. (2022) because it develops a Bayesian framework that integrates documentary evidence. In particular, the framework and methodology of our study are closely aligned with and build upon this work. We will clarify in the revised manuscript why this study is specifically cited.

7. l. 67/68: Can you also state the starting date of the compilations?

Response: Thanks for your comment. We now provide the starting dates of the datasets used in this study. The A Compendium of Chinese Meteorological Records of the Last 3000 Years and the Encyclopedia of Meteorological Disasters in China contain records that date back to as early as the pre-Common Era. The Integrated Natural Disaster Information System of the Qing Dynasty begins in 1644 AD. We will include this information in the revised manuscript. In addition, we will add a supplementary table summarizing the temporal coverage, spatial extent, and period of availability of all datasets used in this study.

8. l. 91: Replace "surface temperature" by "near-surface air temperature".

Response: Thanks for your comment and suggestion. We agree and will revise the term accordingly in the manuscript.

9. l. 92: Please provide the exact definition of winter that you use in the study.

Response: Thanks for your comment and suggestion. In this study, winter is defined as December of a given year and January–February of the following year (DJF). For example, winter 1892/1893 includes December 1892 and January–February 1893. We will add this definition in the Introduction section.

10. l. 93: I wouldn't call the difference between the ensembles "bias" but rather deviation, since bias would imply that there is a known truth from which they differ.

Response: Thanks for your comment and suggestion. We agree that "bias" may be misleading in this context. We will therefore replace it with "deviation" in the revised manuscript.

11. l. 106: What does "time-independent" mean?

Response: Thanks for your comment. The term "time-independent" refers to the removal of temporal structure in the construction of the prior, i.e., the prior does not preserve temporal order from the model simulations. In ModE-RAclim, this is achieved by randomly sampling 100 data points from all ensemble members and all years of ModE-Sim to construct a prior.

Valler et al. (2024) used the term “random stationary priors” to introduce it, and we will revise the term in the manuscript to make it consistent.

Reference

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

12. 1. 118: Why is ”frost” grouped into precipitation and not perception of temperature? How is frost observed?

Response: Thank you for your comment. In documentary records, most records of “frost” generally refers to frost damage to agricultural crops caused by low temperatures, which is recorded as an agricultural impact rather than a direct observation of atmospheric conditions. Therefore it should be grouped into biological. In contrast, the “perception of temperature” refers to people’s subjective perception and description of coldness or warmth. We will revise the manuscript to clarify this classification.

13. How are northern / middle / southern subtropical climates distinguished?

Response: Thank you for your comment. In this study, we directly adopt the climate regionalization results of Zheng et al. (2013) for 1981–2010. Therefore, the northern, middle, and southern subtropical climate zones are same as their classification, which is based on indicators including the number of days with a daily mean temperature $\geq 10\text{ }^{\circ}\text{C}$, accumulated temperature above $10\text{ }^{\circ}\text{C}$, and mean January temperature. We will clarify this in the revised manuscript to avoid ambiguity. The criteria for the northern, middle, and southern subtropical climate zones are presented in Table 1, based on Zheng et al. (2013).

Table 1 Classification criteria for northern, middle, and southern subtropical climate zones based on Zheng et al. (2013).

Temperature zone	Main indicator	Auxiliary indicator	Reference indicator	
	days with $T_{\text{mean}} \geq 10^{\circ}\text{C}$	Mean January temperature ($^{\circ}\text{C}$)	Accumulated temperature above $10\text{ }^{\circ}\text{C}$	Annual extreme minimum temperature ($^{\circ}\text{C}$)
Northern subtropical	210–225 (Yunnan–Guizhou Plateau)	0–4	3500–4500 (Yunnan–Guizhou Plateau)	–14 to –10 / –6 to –4
Middle subtropical	225–285 (Yunnan–Guizhou Plateau)	4–10	4000–5000 (Yunnan–Guizhou Plateau)	–4 to 0 (Yunnan–Guizhou Plateau)
Southern subtropical	285–350	9–10 to 13–15 (Yunnan Plateau)	5000–7500 (Yunnan Plateau)	0–2 (Yunnan Plateau)

Reference

Zheng, J., Bian, J., Ge, Q., Hao, Z., Yin, Y., Liao, Y.: The climate regionalization in China for 1981-2010. *Chinese Science Bulletin*, 58, 3088–3099, 2013.

14. Fig. 1: Can you comment on the reasons for the relatively large differences between DFDP and GSOD for rainy days?

Response: Thank you for your comment. The relatively large differences between DFDP and GSOD are mainly in the complex topographic transition zone from the Sichuan Basin to the Yunnan–Guizhou Plateau. GSOD is an instrumental dataset, whereas DFDP diagnoses daily precipitation types using multiple meteorological variables, including wet-bulb temperature, relative humidity, surface pressure, and elevation. These methodological differences can lead to discrepancies between the two datasets. However, precipitation-related records from these regions, where the differences are larger, were not included in our assimilation, and therefore their impact on the results is expected to be limited.

15. Fig. 2a: It would be more intuitive for me to use a moving window instead.

Response: Thanks for your comment. We will update the Figure 2a by using a moving window.

16. Fig. 2b,c: Consider visualizing this as a barplot containing the observed frequencies for each phenomenon.

Response: Thanks for your suggestions. We will update the Figure 2b,c by using a bar plot.

17. Fig. 3c,d: I would suggest to not connect neighboring years by lines since the Bayesian approach models each year independently.

Response: Thanks for your suggestions. We will update the Figure 3c,d accordingly by removing the connecting lines.

18. Fig. 4: Please add visualizations of uncertainties in Fig. 4 (for example using a density heatmap in Fig. 4a showing the posterior probabilities of each class for each year).

Response: Thanks for your comments and nice suggestions. The Section 4.1.2 only includes the results of index reconstruction before posterior reconstruction. We agree that visualizing posterior uncertainty would improve the interpretability of the results. We introduce posterior probability distributions in Section 4.2. We will add plots to present prior, likelihood and posterior probability distributions of the cold winter index in the Supplementary Materials.

19. 1. 320-321: Could you add uncertainties here to the numbers?

Response: Thank you for the comment and your suggestion. It is difficult to directly quantify the uncertainties associated with the frequency of each index. We will include the reliability of documentary sources for each winter in Fig. 4a to provide additional information on the associated uncertainties.

20. 1. 321-323: Is this a higher co-occurrence rate than expected by chance? I'm wondering because there are ~50 extremely cold / cold winters (i.e., ~20%) in the reconstruction. If

a similar number occurs in southeastern China than a co-occurrence rate of 4% or ~10 common cold winters among 250 years would be expected even if the two regions were statistically independent.

Response: Thanks for your comment. The frequency of extremely cold/cold winters in southwestern China in this study is around 21.6%, while the frequency of extremely cold winters in southeastern China is around 15.6% during 1700–1949 (Hao et al., 2011). As you pointed out, the observed co-occurrence rate is just slightly higher than expected by chance, but not statistically significant. However, winter temperatures in both southwestern and southeastern China are influenced by the Siberian High and the East Asian winter monsoon. Therefore, the historical co-occurrence of cold winters in these two regions may also reflect the connection associated with these large-scale circulation systems.

Reference

Hao, Z., Zheng, J., Ge, Q., Ding, L.: Variations of Extreme Cold Winter Events in Southern China in the Past 400 Years. *ACTA GEOGRAPHICA SINICA*, 66(11), 1479–1485. <https://doi.org/10.11821/xb201111004>, 2011.

21. Fig. 5a: I’m surprised that the distribution of CESM-LME for 1700-1949 seems to be warmer than for the period 1950-2000. Can you comment on this?

Response: Thank you for your comment. We also noticed this feature in Fig. 5a and further examined it at the global scale. We found that the differences between ModE-Sim and CESM-LME are much smaller in the mid- and high-latitude regions (Fig. S2). This suggests it may be related to some latitude-dependent factors.

22. Table 3: Do the p-values account for autocorrelation in the data, which reduces the effective degrees of freedom?

Response: Thanks for your comment. We computed the effective sample size of each pair of time series following the Equation (Xie et al., 2019):

$$\frac{1}{N_{eff}} = \frac{1}{N} + \frac{2}{N} \sum_{j=1}^N \frac{N-j}{N} \rho_{xx(j)} \times \rho_{yy(j)}$$

where the N_{eff} is the effective sample size, N is the real sample size (i.e. 250), $\rho_{xx(j)}$ and $\rho_{yy(j)}$ is the autocorrelation of two time series.

We used the N_{eff} to recalculate the p-values for each pair of time series. The results show that the effective sample size of CWI-ModE-Sim and CWI-ModE-LME is approximately 163. Their correlation remains highly significant after accounting for autocorrelation, whereas it was not significant before applying the assimilation procedure. We will revise the corresponding results in Table 3.

Reference

Xie, T., Li, J., Sun, C., Ding, R., Wang, K., Zhao, C., Feng, J.: NAO implicated as a predictor of the surface air temperature multidecadal variability over East Asia. *Clim Dyn* 53, 895–905. <https://doi.org/10.1007/s00382-019-04624-4>. 2019.

23. I. 414: Do you have any insights into which mechanisms are responsible for the warm winters?

Response: Thank you for your comment. We did not explicitly investigate the mechanisms of warm winters in southwestern China in this study; however, previous studies provide useful insights (Shu et al., 2021; Jiang et al., 2010; Wang et al., 2015; Tao et al., 2018). In general, winter temperature variability in southwestern China is dominated by two leading modes. The first mode represents a coherent variation across the region, while the second mode shows an east–west dipole structure. The dominant controlling factors of the first mode are the Siberian High and the East Asian winter monsoon. A weaker Siberian High and weaker winter monsoon are associated with warmer winters in southwestern China. The Middle East jet stream is the dominant factor influencing the second mode. In addition, local atmospheric systems such as the position of the Kunming quasi-stationary front and the winter monsoon over the Qinghai–Xizang Plateau may also play a role.

Reference

Shu, J., Jiang, X., Song, Y.: Interannual variation of winter surface air temperature over southwest China and its related influence factors. *Journal of Arid Meteorology*, 39(1), 15–27, 2021.

Jiang, X., Li, Y.: The spatial-temporal variance of winter climate anomalies in southwestern China and the possible influencing factors. *ACTA GEOGRAPHICA SINICA*, 65(11), 1325–1335, 2010.

Wang, Y., Li, D., Wang, H., Zheng, R.: New variation characteristics of winter monsoon over Qinghai-Xizang plateau and its influences on temperature over southwest China. *Plateau Meteorology*, 34(1), 11–20, 2015.

Tao, Y., Duan, X., Yao, Y., Duan, C., Ren, J., Chen, Y., He, H.: The climate correlation analysis between the cold wave in Yunnan and Kunming Quasi-Stationary front. *Journal of Catastrophology*, 33(2), 99–105, 2018.

24. Similar to my question above, is the agreement of warm winters in CWI-ModE-Clim and ModE-Sim actually higher than expected by chance?

Response: Thank you for your comment. We did not perform a statistical analysis on this aspect. We will add relevant information in the revised manuscript.

25. Fig. 7: Can you also plot a comparison of CWI-ModE-Clim and CWI-ModE-Sim which would show the influence of the simulations on the CWI-ModE-Sim reconstructions?

Response: Thank you for the comment and your suggestion. CWI-ModE-Sim includes information from both documentary evidence and simulations, whereas CWI-ModE-Clim only includes documentary evidence, and ModE-Sim consists purely of simulations. We will update Figure 7 to include CWI-ModE-Sim in the comparison.

26. I. 419: That information would be helpful much earlier, either in the data or methods section.

Response: Thank you for your comment and suggestion. The definition of cold and warm winters in Section 4.3 is introduced specifically for the purpose of constructing composite circulation patterns. The “extremely cold” and “cold winters” referred to in Section 4.1.2 are reconstructed directly from documentary evidence and are not defined using a statistical threshold. We will introduce this information in the method section in the revised manuscript to avoid ambiguity.

27. 1. 439, 443-444, 446, 453: Please add references here.

Response: Thanks for your suggestions. We will add relative references in these sentences.

28. 1. 459: Do you mean lower amplitude or lower temporal frequency?

Response: Thanks for your comment. Here we intended to refer to lower temporal frequency and we will modify the sentence in the revised manuscript.

Technical comments

1. 1. 32-33: Please check the grammar of the sentence.

Response: Thanks for your comment. We will carefully check the entire manuscript for grammatical issues and make revisions accordingly

2. 1. 38: Please check the grammar of the sentence.

Response: Thanks for your comment. We will rewrite this sentence.

3. 1. 385-386: Please check the grammar of the sentence.

Response: Thanks for your comment. We will rewrite this sentence.

4. 1. 439: Please check the grammar of the sentence.

Response: Thanks for your comment. We will rewrite this sentence.