Reply to the referee comments on "Extreme droughts/floods and their impacts on harvest derived from historical documents in Eastern China during 801–1910" by Zhixin Hao et al

Dear editors and reviewers,

Thank you for your valuable comments and thoughtful suggestions on our manuscript. Following your comments on the manuscript, we made careful revisions, and the point-to-point response (in red) of the comments (in black) is listed below. We hope these revisions would make this manuscript more acceptable for publication. Please feel free to contact me if you have any questions.

Many thanks again. With best wishes.

Sincerely yours,

Jingyun Zheng

To Anonymous Referee #1:

After we carefully read the comments, as our understanding, the most concern from reviewer is that the dataset of droughts/floods is not independent of the dataset of harvest grades, since they were reconstructed from the same historical documents. We thought that the spurious correlation induced by historical documents' recording and circulation could be limited for three reasons. First, the two datasets were not from the same documents, in which the reconstructed harvest grades were derived from the records in "Twenty-four histories" and "Qing history Draft", while the reported droughts/floods were derived not only from "Twenty-four histories" and "Qing history Draft", but also chronicles, miscellaneous historical books, local gazettes and others. Secondly, as our reading experience in the historical documents, we did not find large number of records reporting both extreme events and harvests simultaneously. Thirdly, during the reconstruction of droughts/floods and harvest grades, the record collection, explanation and calibrated parameters were not the same. Although as the reviewer suggested, this problem could not be completely avoided indeed. So we accepted all the comments, and deleted the historical inferences such as "occurrence of extreme drought in eastern China could lead to significant increase of frequency of poor harvest". The possible inferences according to statistics are now presented in the *Discussion* section, along with discussions on the data independence problem and confounding factors. Also, the "reported" droughts/floods and "reconstructed" harvest grades are highlighted explicitly throughout the text to make clear that all the results are focused on the two datasets, instead of

actual historical events. The detailed point-to-point responses are listed below.

1) The title should reflect the fact that this is primarily a study about patterns found in data in historical and climate databases, and only secondarily a study from which we might make inferences about history. In this case, I would accept, for instance, “Patterns in the Reporting of Droughts/Floods and of Harvest Grades in Historical Documents in Eastern China, 801-1910” or “Patterns in Data on Precipitation Extremes and Harvests in Historical Databases for Eastern China 801-1910”.

Accepted. The title has been changed to "Patterns in data of extreme droughts/floods and harvest grades derived from historical documents in Eastern China during 801–1910".

2) The article should clearly distinguish its identification and analysis of data and patterns in that data from any inferences about (climate) history that it makes based on those data and patterns. Those inferences should be more limited, in accordance with the uncertainties and methodological problems outlined above, and should probably appear in the discussion or conclusion of the article. That would also be a good place to discuss whether studies such as Hao et al. 2010 really enable researchers to infer actual historical events and causality from this data.

Accepted. The results (section 3.1-3.3) are now focused on analysis of datasets, while the possible inferences are presented in discussions. (P12, L22-P13, L3)

The statistical results from these two datasets indicate that regional extreme droughts might be closely connected with poor harvest in Chinese history, and this connection seems to be weaker in the warm period and stronger in the cold period. However, these inferences are purely based on those two reconstructed datasets, and insufficient to reveal actual historical connections. One of the reasons is that both datasets used Twenty-Four Histories and Qing History Draft as their record resources in the reconstruction, which might induce artefact in databases and lead to spurious correlations between extreme drought/flood and harvest. Another reason is that there existed major shifts in many social factors (such as political, economic, demographic or cultural change) in the study period, which means that the different pattern in association between reported extreme drought and reconstructed poor harvest between warm and cold periods might be co-created by those confounding factors along with climate factor. Although the dataset for extreme droughts/floods used more historical documents and were not limited to the Twenty-Four Histories and Qing History Draft as resources, and both datasets have been validated by other proxy data in previous studies, the independence and consistency problems could not be eliminated entirely. The solution for these two problems requires further research using more

independent datasets from multi-proxies and many other study fields, and could be of great importance in improving the understanding of the climatic impacts on agriculture and adaptation to future global warming along with higher extreme climate probability.

3) While discussing the data itself and patterns identified in that data, all mentions of historical floods, droughts, and harvest levels should be qualified as "reported" floods or droughts and "reconstructed" harvest levels. Instead of "co-occurrence" of hydrological extremes and poor harvests, the authors should refer to the "co-reporting" of hydrological extremes and poor harvests.

Accepted. In the revised manuscript, when the droughts, floods and harvests are mentioned, the word "reported" and "reconstructed" are added explicitly to make clear that they refer to reconstructed datasets rather than actual historic events. The use of "co-occurrence" are also revised as "co-reporting".

4) Please see the replies below for further specific points.

In the reviewer comments, all the comments and author responses from previous review are listed, along with reviewer's reply to each of them. Some of our previous responses are accepted, and some are still disputable to the reviewer. Here we only present the comments which need more revisions according to the reviewer's comment.

4.1.4) What is being measured by "harvest"? Yield per seed? Total yield per hectare? Food availability?

Accepted and revised. Harvest in records was a relative concept and represented the ratio of actual yield compared to the possible maximum yield. (P6, L1-3)

In Chinese historical documents, the yearly harvest was usually recorded as a relative level compared to an expected maximum yield, rather than crop yield per hectare, although some records also report impacts of harvest fluctuation on food availability, tax remissions, livelihoods, and so on.

Reply: If this is the case, then the authors need to be explicit throughout the text that what they have reconstructed is not absolute "harvest levels" but rather interannual variability in local production of staple crops.

Accepted. We replaced "harvest level" with "harvest grade" in the text, and emphasized the definition for "harvest grade" in the *Data* section. (P6, L13-14)

It should be noted that these reconstructed grades do not represent absolute grain yield, but rather

the relative percentage in production of staple crops and reflect their inter-annual variability.

4.2.1) How do the data control for the changing borders of Chinese empires? A priori, I would expect vastly different vulnerabilities and patterns of reporting between the Northern Song and Southern Song periods, simply based on the major geographical shifts in population and wealth between those two dynasties.

Accepted and revised. For droughts and floods, the historical records was transformed into graded data based on 63-stations, each of which was set as a local area consisted of about 20 counties and does not change in different dynasties (P3, L30-31). Although the available graded data was unevenly distributed spatially for different dynasties, it had been proved in the paper of Hao et al. (2010a) that the extreme drought/flood years recognized were mostly robust despite of the percentage of data-missing stations (P5, L5-11). As for harvest, the impact of changing borders on harvest grade should also be limited since the main grain product area had been relatively stable in the study period and the records in documents was about relative harvest rather than absolute yield as suggested by Yin et al. (2015) (P6, L28-31).

Based on these records, Zhang (1996) reconstructed a dataset of annual drought/flood grades at 63 stations from 137 BCE. Each station consisted of a local area of approximately 20 counties with the same climate.

Reply: This revision is appropriate, but again only if the authors consistently make clear that they are reconstructing interannual variability in local production of staple crops and not an absolute "harvest level." The absolute level of harvests in different regions and eras on decadal or longer scales would still have depended more on changes in political economy, population density, crop strains, and technologies than on weather and climate.

Accepted. It is the same with comment 4.1.4. We replaced "harvest level" with "harvest grade" in the text, and emphasized the definition for "harvest grade" in the *Data* section. (P6, L13-14)

It should be noted that these reconstructed grades do not represent absolute grain yield, but rather the relative percentage in production of staple crops and reflect their inter-annual variability.

To verify the rationality of this method and criteria, validation was conducted in Hao et al. (2010a), based on 10 extreme events identified from a series of precipitation observations in each sub-region according to a threshold of probabilities of 10% and 90% occurrence. In this validation, all or part of grade 3 stations were deliberately omitted, and only 40% or 60% of stations with disaster or extreme grade were reserved without changing the drought-to-flood ratio within the available data. The results show that, with one exception, years of extreme drought and

extreme flood, identified according to this method and criteria, closely matched those extreme events identified by precipitation data, demonstrating that the method and criteria were reasonable.

However, such social factors should have only limited influence on yearly harvest grade dataset, since the harvest in the documents was reported as a relative level rather than the absolute yield, also the main grain product area, the staple crop, and the cropping system have been relatively stable throughout the study period (Yin et al., 2015).

Reply: It appears that Hao et al. 2010 helps establish that the historical sources underlying the databases were not biased toward reporting floods versus droughts or to reporting them in some regions and not others. That study also appears to establish that these sources did not usually falsely report precipitation extremes. However, Hao et al. 2010 does not establish that the reporting of precipitation extremes was independent of the reporting of poor food production throughout the long and diverse history of today's China. This remains a major shortcoming of the study, which I will return to below.

Accepted. The paper of Hao et al. (2010) only introduced the reconstruction of droughts/floods, and did not write about harvest grades. As mentioned in the first paragraph of this response, we revised the expression in the results, and made discussion on the flaw of data reflecting the real facts. (P12, L22-L27)

The statistical results from these two datasets indicate that regional extreme droughts might be closely connected with poor harvest in Chinese history, and this connection seems to be weaker in the warm period and stronger in the cold period. However, these inferences are purely based on those two reconstructed datasets, and insufficient to reveal actual historical connections. One of the reasons is that both datasets used Twenty-Four Histories and Qing History Draft as their record resources in the reconstruction, which might induce artefact in databases and lead to spurious correlations between extreme drought/flood and harvest.

4.2.2) How do data on "harvests" control for changes in staple crops, introduction of New World crops including peanuts and sweet potatoes, changing cropping patterns, and the increasing commercial orientation of agriculture?

Accepted and revised. The records for harvests were usually about relative percentage compared to expected maximum yield rather than absolute yield, and thus it should not be influenced by these factors. (P6, L1-5)

In Chinese historical documents, the yearly harvest was usually recorded as a relative level compared to an expected maximum yield, rather than crop yield per hectare, although some

records also report impacts of harvest fluctuation on food availability, tax remissions, livelihoods, and so on. Therefore these harvest records exclude differences in absolute yield between sub-regions with different climates, soil fertility and types, crop varieties, etc., as well as difference between historical periods with changing agricultural centres, farming technologies, staple crops, and so on. (Su et al., 2014).

Reply: This revision is appropriate, but again only if the authors consistently make clear that they are reconstructing interannual variability in local production of staple crops and not an absolute "harvest level." The absolute level of harvests in different regions and eras on decadal or longer scales would still have depended more on changes in political economy, population density, crop strains, and technologies than on weather and climate.

Accepted. It is the same with previous comments 4.1.4 and 4.2.1. We replaced "harvest level" with "harvest grade" in the text, and emphasized the definition for "harvest grade" in the *Data* section (P6, L13-14).

It should be noted that these reconstructed grades do not represent absolute grain yield, but rather the relative percentage in production of staple crops and reflect their inter-annual variability.

4.2.5) Most importantly, how can we make up for the fact there are simply more records from the Qing period than earlier periods? I don't see that the methods used in this manuscript avoid the problem that more records will create a misimpression of a greater frequency of floods and droughts. The authors propose to ignore reports of "average" conditions in Qing records to make them more comparable to Song and Ming records. However, that would only work if the Song and Ming records still reliably reported all disasters and extremes and only left out "average" conditions. I don't see any reason to make that assumption. Perhaps the authors could experiment with methods of introducing "noise" into the data in order to reflect the events missing from the reports. Or else they could employ a Bayesian method to indicate that the presence or absence of certain descriptions in the records may be used to obtain updated posterior probabilities of actual conditions, without ever assuming that the records provide a complete account of events. In any case, the authors must come up with a way to handle these changes in the documentary record over time if they are to make a convincing case for stable long-term correlations between floods and droughts and harvests.

Accepted and revised. The method of ignoring "average" conditions is based on the hypothesis that the records on droughts and floods were omitted randomly and unbiased, which suggests that the relative drought-to-flood ratio in the available data would be close to that in actual history. As in the abovementioned revisions, the recognition for extreme drought and flood years was still

effective even if 40% or 60% of the available data with disasters was omitted deliberately. (P4, L34-P5, L13)

Extreme drought or extreme flood years were defined in this way, as the probabilities for omitting drought and flood records were random and unbiased, despite the greater frequency of missing data in the older records. In other words, if one period had a large number of documents, it was expected to be rich in both drought and flood records, and vice versa. Therefore, the amount of missing data should not have a significant effect on the relative drought-to-flood ratio within the available data. To verify the rationality of this method and criteria, validation was conducted in Hao et al. (2010a), based on 10 extreme events identified from a series of precipitation observations in each sub-region according to a threshold of probabilities of 10% and 90% occurrence. In this validation, all or part of grade 3 stations were deliberately omitted, and only 40% or 60% of stations with disaster or extreme grade were reserved without changing the drought-to-flood ratio within the available data. The results show that, with one exception, years of extreme drought and extreme flood, identified according to this method and criteria, closely matched those extreme events identified by precipitation data, demonstrating that the method and criteria were reasonable. The reason for the close match is that precipitation variability in eastern China is dominated by the East Asian Summer Monsoon (EASM). Therefore, when extreme drought or flood events occur, the precipitation variation for stations within each sub-region usually share similar relative magnitudes.

Reply: It appears that Hao et al. 2010 helps establish that the historical sources underlying the databases were not biased toward reporting floods versus droughts or to reporting them in some regions and not others. That study also appears to establish that these sources did not usually falsely report precipitation extremes. However, Hao et al. 2010 does not establish that the reporting of precipitation extremes was independent of the reporting of poor food production throughout the long and diverse history of today's China. This remains a major shortcoming of the study, which I will return to below.

Accepted. It is the same with second part of comment 4.2.1. The paper of Hao et al. (2010) only introduced the reconstruction of droughts/floods, and did not write about harvest grades. As mentioned in the first paragraph of this response, we revised the expression in the results, and made discussion on the flaw of data reflecting the real facts. (P12, L22-L27)

The statistical results from these two datasets indicate that regional extreme droughts might be closely connected with poor harvest in Chinese history, and this connection seems to be weaker in the warm period and stronger in the cold period. However, these inferences are purely based on those two reconstructed datasets, and insufficient to reveal actual historical connections. One of

the reasons is that both datasets used Twenty-Four Histories and Qing History Draft as their record resources in the reconstruction, which might induce artefact in databases and lead to spurious correlations between extreme drought/flood and harvest.

4.3.1) Drought and/or flood might have correlated with other climate variables (such as temperature) that caused harvest failure.

Accepted. As elaborated in previous study, the relationship between temperature and harvest had been investigated by Yin et al. (2015, 2016), which suggested that there would be better harvest in warm climate. And our study, in section 3.2 of the original manuscript, found that more occurrence of extreme drought in eastern China could lead to significant increase of frequency of poor harvest (grade 1+2) compared with non-extreme years. To further examine whether the drought and/or flood are correlated with temperature change, and if so, how the drought and/or flood are correlated with harvest failure under different temperature backgrounds, we presented a study in section 3.3, and found that there were slightly more extreme droughts in the warm period. However, the connection between extreme droughts and poor harvest was not significantly close in the warm epoch, while it was more significant in the cold epoch. These results suggested that warm period could weaken the impact of extreme drought on poor harvest during historical times. (P11, L18-22; P11, L28-P12, L2)

As found in section 3.2, more occurrence of extreme drought in eastern China led to a significant increase in the frequency of poor harvests (grade 1+2) when compared with non-extreme years. Since more extreme droughts occurred over eastern China in 920–1300 than in 1310–1880, the harvest in the warm epoch could be expected to be worse than in the cold epoch. However, as Yin et al. (2015, 2016) found, the harvest in warm epoch was better than that in cold epoch. This suggests that the effects of regional extreme drought on the grain harvest differed between warm and cold epochs.

The results show that, during the warm epoch of 920–1300, there was no significant connection between the occurrence of poor harvest and regional extreme drought, although the frequency of poor harvest in extreme drought years was slightly higher than in non-extreme years for each sub-region. In contrast, during the cold epoch of 1310–1880, the frequency of poor harvest in extreme drought years was significantly higher than in non-extreme years, which indicates that the connection between the occurrence of poor harvest and extreme drought was still significant. Moreover, similar characteristics were found for the latter half of the cold period from 1650 to 1880, which indicates that the shift of harvest grade distribution did not affect the connection between poor harvest and extreme drought/flood during the cold epoch. These results suggest that

the warm period could weaken the impact of extreme drought on poor harvests in historical times.

Reply: I do not find that this approach adequately disambiguates the effects of extreme precipitation and temperature on food production. It's a basic inductive method that simply takes all the data and then comes up with an explanation after the fact based on any observed patterns – patterns that could have emerged entirely by chance, or by some confounding third variable (e.g., population movements or changes in political economy that influenced agricultural vulnerabilities), or due to some artefact in the way events were reported or records kept. I'd be much more satisfied if there were some way to model the effects of both temperature variations and extreme precipitation on food production levels so that the authors could weigh relative contributions of each.

Accepted. In the revised manuscript, the idea of making simple historical inferences from patterns in the data has been revised. The inappropriate subjective expressions, e.g., "the impact of extreme floods/droughts on the harvest", "connections between climate and harvest", have been changed to more objective expressions, such as "association between reported droughts/floods and reconstructed harvest grades". The result sections focus only on the patterns from the datasets, and possible inferences and confounding factors have been put in the discussion section.

In addition, since the reconstructed climate datasets have different time resolutions (e.g., reconstructed temperature has a decadal resolution, while reported droughts/floods dataset has an annual resolution), and the response of food production to each climate factor has different sensitivity and time scale, so it is difficult to weigh the relative contributions of each climate factor. Even under the modern observation condition, it is not easy to answer this question, which needs large data sample size or model simulation method. Therefore the revised manuscript did not induce discussions on the separate effects of temperature variations and extreme precipitation on food production. This question might be addressed in our future studies.

4.3.2) Drought and/or flood might have increased the likelihood that officials reported problems such as poor harvests and other disasters

Accepted. As expressed in abovementioned revisions (P6, L1-5; Table S2), the records on harvests in historical documents was a relative level and focused directly on cropping in most cases, therefore it is reasonable to suggest that there was no tendency in harvest records.

In Chinese historical documents, the yearly harvest was usually recorded as a relative level compared to an expected maximum yield, rather than crop yield per hectare, although some records also report impacts of harvest fluctuation on food availability, tax remissions, livelihoods, and so on. Therefore these harvest records exclude differences in absolute yield between

sub-regions with different climates, soil fertility and types, crop varieties, etc., as well as difference between historical periods with changing agricultural centres, farming technologies, staple crops, and so on. (Su et al., 2014).

Reply: The authors' response does not address my concern. Any causal argument about precipitation extremes and low harvests relies on the assumption that both of these phenomena were consistently reported independently of each other. It appears that Hao et al. 2010 helps establish that the historical sources underlying the databases were not biased toward reporting floods versus droughts or to reporting them in some regions and not others. It also appears to establish that these sources did not usually falsely report precipitation extremes. However, Hao et al. 2010 does not establish that the reporting of precipitation extremes was independent of the reporting of poor food production throughout China's long and diverse history. There remains the strong possibility that officials were more likely to report either the occurrence of such extremes when the harvests were poor (e.g., by way of explaining or justifying those poor harvests) or to mention poor harvests when there were meteorological extremes (e.g., in an effort to take advantage of the situation to secure state funds or tax remissions). I do not make this as merely a theoretical argument. Although I am not an expert on imperial China, this is exactly the pattern I have observed in the records of other early modern empires. Moreover, certain features of this study make this problem particularly troublesome for their conclusions. First, the information concerning both types of events (precipitation extremes and poor harvests) comes from the same set of historical records. Second, the records from earlier eras come primarily from Ming and Qing recompilations and not originals. Third, the older records are very incomplete and biased toward extreme events. Fourth, and most important, the authors' basic inductive method—taking all the data and then explaining any observed patterns on the assumption that they are causally related— sets a very high standard for the independence of the different datasets. In other words, if the authors were simply making the case that over the long run we should see an impact of floods and/or droughts on interannual food production, then small changes in political priorities or record-keeping and transmission practices shouldn't make a big difference. However, because the authors are identifying shifts and patterns over time—such as possible changes in the impact of precipitation extremes due to phases of warming and cooling—then any shifts in the independence of reporting the occurrence of extreme precipitation events and of poor harvests are likely show up in their presentation of the data and to be explained as real historical changes in the impact of floods and droughts on food production. In fact, that is exactly what I think has happened in this

study, and that is why I am asking for major revisions.

Accepted. In the revised manuscript, the historical inferences with insufficient evidence has been revised, and the result sections focus only on the statistics and patterns extracted from the two datasets. Although several explanations should be made toward some aspects of this comment. For the first part of this comment, the two datasets are not from same documents, in which the reconstructed harvest grades were derived from the records in "Twenty-four histories" and "Qing history Draft", while the reported droughts/floods events were derived not only from "Twenty-four histories" and "Qing history Draft", but also chronicles, miscellaneous historical books, local gazettes and others. For the second part of this comment, the recompilation of the Ming and Qing histories was only a systematic documents reorganization, and did not add new subjective points of view or more contents related with droughts/floods or harvests. For the third part of this comment, the problem induced by incomplete and biased older records could be avoided to a large extent, because the statistics focus on the relative ratio between extreme droughts and extreme floods. This ratio should not be influenced by the percentage of missing data, since there is no tendency in omitting drought or flood records. In addition, we divided the whole eastern China into three sub-regions based on the movement of Asian Summer Monsoon, and the large number of sites with similar precipitation characteristics in each sub-regions could compensate the shortcoming of incomplete records in the early period to a certain degree.

To Anonymous Referee #2:

The author has made an effort to insert a paragraph describing data characteristics and basic statistics. This definitely helps to increase data transparency procedure. However their explanations about the nature of the written records, interpretation of the records, and the data statistics are not rigorous and likely to incur more problems at the epistemological level. In fact, from my perspective, the authors do not really face the real challenges that the reviewers proposed to them and just to repeat the problems in the paper or to ignore. Below are just some of the examples.

Accepted. Same as the comments given by reviewer #1, in the revised manuscript we discussed the flaw of the dataset, and especially emphasized that the results only reflect patterns extracted from two datasets instead of actual historical conditions. We weakened the description of "impact" or "connection", and only wrote about the existing phenomena about "reported" extreme events and "reconstructed" harvests.

In addition, the uncertainty of the historical documents indeed existed, which need more historical

document resources, more datasets from other proxies, and new methods such as model simulation. In order to keep the continuity, we expect to address this problem with more datasets and model simulation method in the future study.

On page 3 line 23-29, the authors strictly assert that drought records in the historical documents can be regarded as meteorological drought rather than hydrological or agricultural drought, and same declaration also appears for flood records that reflect more rain rather than river or lake overflowing etc. The declarations are hard to be accepted or digested literally because it is obvious that both drought and flood phenomena and the intensity can be largely influenced by the population size, biological and geographical conditions even today. Thus, when authors purely declare so without providing further justifications, this can lead to a non-negligible mistrust and suspicion from the readers to question the reliability of the record data and the analysis.

As showed in the data example (P3, L13-15), the documents at the 19th year of emperor Zhenyuan in Tang Dynasty (803 CE) record: "*no rain fell in the Guanzhong area (now Xi'an and surrounding areas) from the 1st month (January 27 to February 24) until now (In day 25 of the 6th month (of the Chinese lunar calendar, or July 17, 803 in the solar calendar). Hundreds of officials and many people are praying for rain. In day 26 of the 7th month (August 18), rain. In day 17 of the 8th month (September 6), heavy rain*". The record has explicit date, type of rain (intensity), duration for the rainfall process or for long-lasting drought. Many other records also recorded the amount of precipitation through the number of inches. These characteristics are not related with population, land cover and geographic condition.

Also it is not clear for me on page 4 line13-16, how the number of 70% of ….can be derived. It is like the authors' assertion without much logical reasonability to follow.

Accepted. We rewrote this sentence. (P4, L13-15)

Even so, it is reasonable to infer that a considerable proportion of extreme events was recorded in 800-1469, comparing the percentage of records reporting "disasters and extremes" in that period and in the ideal frequency (41.8% compared with 60%, respectively).

The added content for validating the grade method on page 4 line 5-10 is somehow not surprising since in this case only stations with extreme grade and disaster were kept (omit grade 3, 40% of data) for comparing with precipitation data (of what years?) that seem to show good match between two sets of data.

Yet, in the first referee report, a clear jump of bumper harvest records (Figure 4) was notified that

corresponds to the coldest interval of LIA and many famine records exist in other documents. All those are against the authors' findings in the study. But the authors only respond to this comment by repeating this 'jump point' in the paper(p13 in revision), without more in-depth reflections or discussions. Actually, from my viewpoint, the number of good harvest can simply be exaggerated by the dramatically increased number of official documents and local chronicle staring Qing dynasty (1644 CE), and the kind of the very basic record/data issues has not been seriously taken by the authors resulting in severe degradation of the research quality of the paper.

First, the reconstructed harvest grades do not represent absolute grain yield, but rather the relative percentage in crop production and reflect their inter-annual variability, which is not influenced by the number of documents. Secondly, as fig.4c shows, the "jump point" still exists when all "average" harvest (grade 3) and missing data are excluded. As addressed in *Data* section, the probabilities for omitting each side of extremes in records are supposed to be random and unbiased, despite the greater frequency of missing data in the older records. When the total amount of records significantly increased in Qing Dynasty, it is expected to find that amounts of records on poor harvest and bumper harvest both increased, and the relative ratio between poor and bumper harvest should not be influenced by the change in number of records. In addition, according to fig.4c, there's also a similar abundant period for poor harvest (grade 1+2) during 1260-1530, yet the number of records in that period did not decrease significantly, which also indicates that the jump in relative percentage of harvest grades is not associated with number of records.