

REPLY TO THE REVIEWER COMMENTS

(Reviewers' comments are in blue, replies in black, changes to the manuscript in red)

REVIEWER #2

The thematic focus of the article — a global inventory of "historical documentary evidence" — is highly relevant. And it is certainly desirable to create such an inventory. However, there are serious methodological issues to solve in order to create a useful and reliable inventory of this kind, and I doubt that these problems have been addressed appropriately in the reviewed version of the article. Methodological problems start with the problem of definition. In the context of historical climatology, "historical" and "documentary" evidence are synonymous. So, what is the meaning of "historical documentary evidence" in the first instance?

Thank you for that comment. When we talk about historical evidence, we can distinguish between historical instrumental (measured) evidence and historical documentary (written) evidence. In this inventory, the focus lays strictly on the latter, which is why I explicitly specified "historical documentary evidence" throughout the manuscript. However, I agree with the reviewer and will omit "historical" altogether starting with the title and throughout the manuscript to avoid redundancy. To specify that I am talking about historical documentary (written) evidence I adjust the sentence where "documentary evidence" appears for the first time in the following way:

"It combines information on past climate from all around the world, retrieved from many studies on documentary (i.e., **written**) sources. Historical evidence range from personal diaries, chronicles, administrative/ clerical documents to ship logbooks and newspaper articles."

Furthermore, I add the following sentence to the section where I define the criteria for the inventory:

"First and foremost, only written historical evidence are considered in this inventory. Other historical evidence such as (early) instrumental measurements are excluded since they are measured observations."

Moreover, the term "observations" is used in more than one way in this article, which requires clarification. In some of the earlier passages, observations are distinguished — at least implicitly — from the documentary record. In these passages, "observations" refer to measurements. Later, Pfister is quoted affirmatively for subsuming measurements under the broader distinction between "direct" and "indirect" information. Observations in his understanding — and I agree with him — not only include measurements, but also non-quantitative (in other words: qualitative) descriptions. These distinctions need clarification in order to produce a coherent explanation regarding the type of information gathered in the inventory.

Thank you for pointing out the lack of clarity around the term "observations". If I understand the reviewer correctly, he/she refers to the fact that observations can be used as a general term for "(early) instrumental measurements" as well as to refer to "direct observations" i.e., weather and phenological observations recorded in documentary evidence defined by Pfister et al. 1999.

The term "observations" is used ten times in the manuscript. In eight out of ten instances, the manuscript clearly states "which kind of" observation is meant e.g., "weather and phenological observations"; "early instrumental observations"; "the assimilation of surface pressure observations from (ISPD)"; "Sources containing historical documentary evidence related to climate [...] can be divided into direct observations and indirect (proxy) data (Pfister et al., 1999)."; "Direct observations include narrative reports on daily weather, climate anomalies, weather-induced hazards, and non-weather-related events such as famines and epidemics".

There is, however, one instance in the manuscript where the term "observations" is used to refer to instrumental measurements which might be misleading (page 5, line 3). Therefore, I omit "observations" here and instead specify as follows:

"Here, I only include data with a minimum record length of 30 years (necessary for statistical analyses, e.g., allowing meaningful standardization). Out of those, a minimum of 20 years need to be before 1880, otherwise, the value is questionable given the availability of **instrumental measurement-based** data sets from that period onward)."

There is one additional instance where "observations" is not explicitly defined: "[CLIMWOC] contains observations from [ship] voyages for the pre-industrial period 1750-1854." From these ship voyages, both observations of the weather

(precipitation, clouds, winds) and early instrumental measurements are available. To be unequivocally clear that I am referring to the weather observations in this context I rephrase the sentence as follows:

“In terms of marine data, the climatological database on the world’s oceans (CLIWOC) compiles documentary evidence from European ship logbooks (García-Herrera et al., 2005). It contains aside from temperature and air pressure measurements, **direct weather observations** (e.g., wind direction, wind force, present weather, sea conditions) from voyages for the pre-industrial period 1750-1854 from all over the world. The non-instrumental observations can be transformed into quantitative data and are thus available for climate reconstruction.”

Moreover, I change the wording “instrumental observations” to “instrumental measurements” in all the above-mentioned instances to avoid misunderstandings.

The greatest challenge in creating a global inventory of non-quantitative documentary information, which is also much more difficult to solve, is related to language.

There are no qualitative records in this inventory. Every single record included is numerical which means that it is either an index, a time series of e.g., phenological dates, a calibrated temperature/precipitation reconstruction, or an event chronology. If the focus of the inventory were on qualitative documentary evidence, the issue related to language raised by the reviewer would be indeed very valid.

I agree, however, that fact that the inventory only includes quantitative records is not stressed out enough in the manuscript. To avoid misunderstandings in this context, I adjust the manuscript accordingly and emphasize throughout the article, that **only numerical/quantitative/derived records** are considered. For instance, I change the manuscript title and the abstract as follows:

“A global inventory of quantitative documentary evidence related to climate since the 15th century”

“Here, I attempt to compile the first-ever systematic global inventory of quantitative documentary evidence related to climate extending back to the Late Medieval Period.”

The author is obviously not in a position, nor is it her ambition, to provide a survey of all available archival records related to climate history in the given timeframe (late middle ages to the present).

Again, this is a global inventory of relevant documentary climate data in **quantitative** form that only includes numerical record time series. Primary documentary sources are hence not included.

This is clear enough in the article. However, even a review of existing reconstructions based on the historical record (the instrumental record not counting) is almost impossible to achieve for a single person. It requires a consortium of authors capable of screening through older as well as more recent bodies of literature in multiple languages. I nevertheless believe that the findings presented in this article deserve attention and should be published. Yet, the problem of limits to the scope of such an inventory created by a single person should be problematized and addressed.

It would certainly be worthwhile to have a consortium of authors work on this topic. Clearly, the workload could have been carried out more efficiently, and the extent of the output would be even more comprehensive. Since no research group has yet taken on that task, my inventory can be thought of as a first step in assembling and connecting the vast number of quantitative documentary records in a global perspective which is largely missing to date. At best, it might stimulate further research in this direction. To acknowledge this limitation, I will add the following sentence to the conclusion:

“The inventory presented in this study contains a comprehensive set of highly relevant document-based climate time series, however, is certainly far from complete. It may serve as a starting point for further research and potentially stimulate a community effort to e.g., compile a full inventory of documentary evidence, including qualitative archival records”

Further problems: (1) The case studies in part 4 are unnecessary. Historical climatology is an established field of study. Hence, there is no need to demonstrate the value of historical climate information. It is unclear how part 4 relates to parts 1-3.

I disagree with the reviewer on this point. Admittedly, historical climatology is an established research field. However, as demonstrated in the manuscript, the vast majority of the relevant research and its analyses and applications to climate reconstructions is confined to a local or regional scale and is thus lacking a large-scale perspective. To my knowledge, no hemispheric case study using solely documentary records exists. Therefore, it is very relevant to point out the potential of combining individual records in such large-scale perspectives. Furthermore, documentary climate records are only very marginally considered in global climate reconstruction to date. The case studies strikingly underscore their potential for climate reconstruction and are thus essential in conveying the core message of this article. Reviewer #1 agrees with that and even suggests adding more case studies.

(2) The explanations given for the decline of number of historical records in the 19th (in some places) and 20th century (in other places) are too general and somewhat superficial. For example, the Chinese case — and even more so the Asian — is much more complex than can be grasped in one sentence stating that Imperial China ended in 1911. In this context it is particularly important that the author addresses the changing relation between the instrumental and the non-instrumental record. (3) The English language requires considerable rephrasing and editing.

Thank you for this comment. Yes, I agree that my explanations for the decline of record numbers are too brief. In the revised manuscript I change the section to the following:

“The number of documentary series gradually decreases in the second half of the nineteenth and twentieth century for all regions. Partly this is simply because no new series are included after 1860. The reason for this is that many of the compiled series are used in a global climate reconstruction that after 1880 is solely based on instrumental measurements. More importantly, however, this correlates with the exponential development and expansion of instrumental measurement networks across the globe and a diminishing interest in non-instrumental records as a consequence thereof. The sharp drop in numbers for Europe around 1880 can be traced back to the availability of the ice phenology records from the Russian Empire published by Rykachev (1886). Out of these 119 record time series, 103 are not extended beyond Rykachev's publication from 1886. The large majority of these break-up and freeze-up series refer to rivers geographically located East of the Ural Mountains and are, thus, regarded as European series. They account for 85% of the European series ending between 1878 and 1882 and explain the sharp drop. Fourteen records from West of the Ural Mountains initially published by Rykachev (1886) are extended to the 1900s in the publication of Shostakovich (1909). The latter publication includes fifteen additional series from eastern Russia (regarded as Asian series) that also end in the 1900s. These records contribute to the visible drop in the available numbers of Asian series at the beginning of the twentieth century. A further reason for this drop is the fall of the Qing dynasty, China's last imperial dynasty, in 1911 since many Chinese documentary-based record series are based on the vast collection of institutional records from the imperial dynasties. There is an additional marked drop in the overall number of records in the twentieth century which coincides with the start of the Cold War in 1947. The gradual decline in numbers in the late nineteenth and twentieth century can pose a complication since the overlap with instrumental series (needed for calibration) is often limited.”