Dear reviewer,

we thank you for the careful read and helpful comments to improve the paper. The comment aims to identify some points for improvement as well as more fundamental methodological errors. The latter point in particular seems to us to require explanation, justification or revision by us. For ease of reading, we will list each comment in italics as well as attach a response directly below each comment. Citations that are not listed in the paper are listed at the end.

The paper describes a meteorological record covering the late 17th and early 18th century for Paris. Given the length of the record for a period in which instrumental records are very rare, the data and results presented are relevant to the scientific community. However, in my view the manuscript presents major flaws in terms of how the data are analysed and the results interpreted.

General comments:

C1.1- The comparison with modern data is misleading, as the historical data are affected by unknown biases related to the measurement location, thermometer scale, and other factors. These biases can easily surpass climate variability in magnitude. I understand the appeal of comparing historical extreme climate events with modern climate, but this is not possible in a scientifically sound way without applying the necessary bias adjustments - which is very hard if not impossible in this case given the lack of metadata. This is particularly true for temperature and cloud cover.

In the case of temperature, we have made an effort to validate both internally and externally with proxy data such as wine harvest dates. The reason for this was that the calibration of the thermometer by Legrand & Le Goff (1987) was based on proxy data indicating a similar climate of a time period in the 19th century and thus the temperature measurements of this time period served as calibration. Thus, we tried (p.10-13) to show possible inhomogeneities. The results show satisfactory results considering that we deal with measurements in the 17th and early 18th centuries. Nevertheless, we will point out explicitly that biases will remain to a certain extent. Without metadata it is difficult to quantify those biases, but with our analysis it can be seen that the temperature measurements are consistent with the hypothesis of earlier studies. In terms of quantity the time series is mostly inconspicuous. We will display inhomogeneities and provide a time series of the homogenized data in a data file (See C2.9 and C2.13 of the second reviewer).

The graphical illustration of the mean value of the present time in Fig. 8 (Cloud cover) is only intended as a comparative value. A comparative value, which does not establish a statement of the climate variability of the different time periods, but shall only show, whether the subjective "measurement" of the cloud cover results in plausible values. This is true for temperature as well. It is for instance easier to identify cold winters showing anomalies (Fig. 5) as when showing absolute values. But again, we do not claim that for instance a difference of 0.5 °C shows a different in climate between those periods. But we want to show that cold winters, with anomalies up to -4 °C, appeared frequently in the analyzed time period.

Therefore, we feel justified in maintaining these comparisons, with clearer explanation.

C1.2- In addition to the previous point, the Meteoblue dataset used by the authors is a commercial product that, as far as I know, has never been evaluated in peer-review literature. If that is true, it should not be used for a scientific article. My suggestion is to drop the use of a modern reference period for temperature and cloud cover - except perhaps to assess data quality such as for the NDR calculation - and concentrate on the decadal variability of the studied period.

We adopted the ERA5 reanalysis instead of the Meteoblue dataset, because it is obvious that the former dataset is far more often used in the scientific literature. Nevertheless, as stated in C1.1 we want to keep our "comparison" with the reference period.

C1.3- The temperature record is clearly affected by inhomogeneities. The authors actually do a very good job in pointing them out, by mentioning relocations, changes in the temperature scale, and changes in the ventilation of the instrument (Tab. 1; also Fig. 1a and 1b point to at least two important inhomogeneities). However, this fact is completely ignored when analysing the data. There are some confusing sentences about this at the end of Sect. 3.3 that actually raise even more doubts about the quality of the data. I believe that some kind of statistical homogenization is necessary, even though reference series for this period are scarce. Beside the Central England Temperature series, there exist many temperature reconstruction that could be used.

We will be content to only show anomalies and provide both time series of the raw data and a homogenized time series in one data file. However, since the homogenization does not affect the main statements in the paper, we want to keep the graphs of temperature based on the raw data after calibration.

Regarding a statistical homogenization: Proxy data do not seem to us to be useful since they do not have the required temporal resolution. However, a comparison has already been discussed in the paper anyway with grape harvest dates. A homogenization with the CET is conceivable, but also connected with problems. First, the CET has only a monthly resolution and is based on mean temperatures, whereas our data have a daily resolution, and we are interested in the maximum and minimum temperatures. A homogenization with only one time series can further lead to spurious correlations and climate signals of the own measurements can be lost.

Furthermore, comparisons with other measurements show quite consistent results (See below for instance the winter temperature anomalies of De Bilt, CET and Paris).

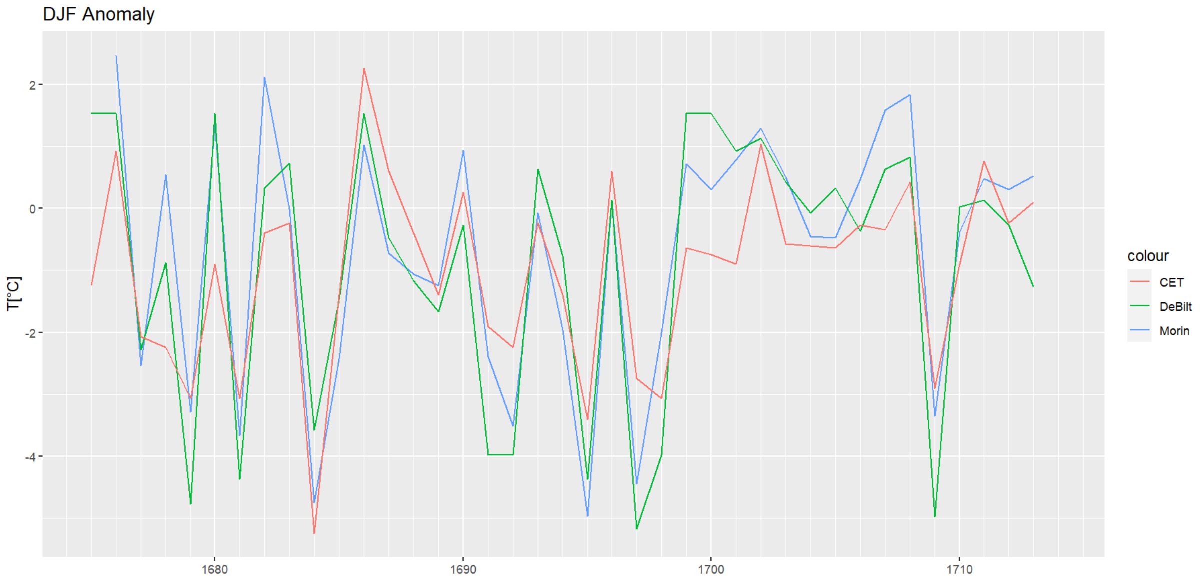

Figure 1: DJF Anomalies of Paris, De Bilt and CET (Manley, 1974)

C1.4- How does this record relate to the widely available long monthly temperature series for Paris? Is that series also based on Morin's observations? Are there any differences from your data?

If this point refers to the temperature series of Rousseau (2009, 2013) then yes. The difference is that we wanted to provide data on a daily basis.

C1.5- Many equations and definitions appear in the results. They should be moved to the methods section.

We will make slight modifications but would like to keep most of the layout due to readability.

C1.6- I am not a native English speaker but the quality of the language seems rather poor to me, to the point that I had difficulties understanding some sentences.

The article has been proofread by PRS. (https://www.proof-reading-services.org/en/)

Specific comments:

C1.7- The procedure to convert the temperature readings to Celsius need to be explained more in details, since the given references are in French. Besides, the conversion formulas (Eqs. 1-3) are not completely clear to me: I would expect that the TM in the three equations refer to different observation times, but this is not indicated. Moreover, it is often mentioned in the manuscript that the thermometer was filled with spirit: is this an assumption or a known fact? How do you explain that a linear conversion does not introduce a bias at high temperatures?

A more detailed explanation of the conversion to °C will be adopted. Also, the terminologies (TM) will be pointed out more clearly. TM basically means measured temperature and refers to the highest measured value per day when calculating Tmax and to the lowest measured value per day when calculating Tmin.

We do not know for sure if the thermometer was filled with mercury or spirit. Earlier literature assumes the latter. Nevertheless, we could not identify biases which can appear in temperature measurements using spirit (See Camuffo, 2016). We will make this point clear. However, the term "liquid" was used just once in the context of the calibration procedure, which is indeed an assumption. We will modify the text to avoid misunderstandings.

An unreasonably high bias at high temperatures due to radiation, liquid thermometers, etc. would be visible when comparing the GST (growing season temperature) with grape harvest dates and when, what we did but don't show it in the preprint, looking at qq plots with EOB-S data (Original units of Morin and temperature series of EOB-S). The comparison is not scientifically valid, but unreasonably high biases could have been seen there, if existing.

C1.8- Equation on page 10 (number missing): I believe the indices i,j,k here are in the wrong positions.

Thank you for making the effort to understand the equation and to point out the misleading indices. We changed that and added a number to the equation.

C1.9- P11, L207: Dai (2006) shows that the effect of pressure on snowmelt is negligible in the lower troposphere. Besides, increasing humidity cause the melting point temperature to decrease (i.e. a lower temperature is required for snow), not increase. More importantly, precipitation phase at the surface depends on the temperature profile above the station, of which surface temperature is merely a proxy (e.g. it can be significantly warmer 1 km above the surface than at the surface, hence it can rain with negative temperature). Another important factor is precipitation intensity (higher intensity implies higher melting point temperature).

We are aware that there are different factors which lead to a different melting point. Therefore, we see the calculation in Fig. 3 as a statistical approach to validate for plausibility. Or in other words: To check for inhomogeneities of the temperature measurement near the snow/rain threshold.

C1.10- *P11, L215: "So, if..." - something wrong with this sentence, snowfall frequency is not measured in °C.*

Thank you. We correct the sentence.

C1.11- *Equation on page 12 (number missing): What is the factor 2 for? The notation for the sums is confusing.*

The factor 2 will be deleted, because it appears in the denominator and in the numerator. But we will keep the notations of the sum (See also Camuffo, 2017)

C1.12- *P12, L231: What is a typical value for NDR for data measured indoor? How relevant is the change from 0.8 to 0.95 in 1688?*

For indoor measurements the NDR value is below 0.3 (See Electronic supplementary material Camuffo, 2017). Therefore, we can state that the measurements were performed outside. A change of 0.8 to 0.95 is of relevance, because it points out that the exposure or maybe the thermometer itself changed. (See also C2.13 for more detail)

C1.13- *P13, L263: The Maunder and Spörer Minima are defined by solar activity, not by climate, and the influence of solar activity on climate is still uncertain. This is mentioned briefly later in the manuscript, but I believe it should be clarified already in the introduction. The choice of LMM to describe the period covered by the data is perhaps not the best as it gives the impression that the climate anomalies were mainly driven by solar activity.*

We adopt a clearer explanation of the terminologies but point out once again that the terminology “Late Maunder Minimum” is well accepted in historical climatology.

C1.14- *P22, L349: "This means that..." - Circulation is an essential requirement for cold winters rather than an additional driving factor. Even a possible solar influence would mainly act through changes in circulation (e.g. Barriopedro et al., 2008).*

We totally agree to this comment and will modify statements, which are misleading or stating too strong hypotheses. What we have found in our study is just a lower frequency of WI (basically frequency of clouds moving from west to east) and that there is a correlation of a low WI with lower temperatures in winter. This can also be seen as a further validation of the temperature measurements.

C1.15- P22, L367: How is exactly the DI calculated from modern data? Are clear days excluded? If not, there would be an obvious bias with respect to Morin's observations. How dependent are the results from the choice of the levels? This comparison should be done using an open, peer-reviewed dataset (e.g. ECMWF ERA5 reanalysis), or dropped.

Thank you for coming up with this point. We did not exclude clear days in the first plot, but changed that and the results are still the same. A different weighting of the levels would result mainly in slightly different numbers of EI and SI. WI has in all levels of our interest higher values than our calculated mean for WI of Morin's measurements. We excluded the clear days (TCC<10%) and the result for ERA5 can be seen in the following below.

However, we will reconsider slightly different levels for comparison. In the plot below we took the mean of the 900 hP, 850 hP, 700 hP and 500 hP level for summer and spring and the same for winter and autumn with a weight on the 900 hP level of the factor of 2.

We would like to point out in particular the agreement of Morin's observations with the reanalyses, which is outstanding considering that the measurement was made by eye.

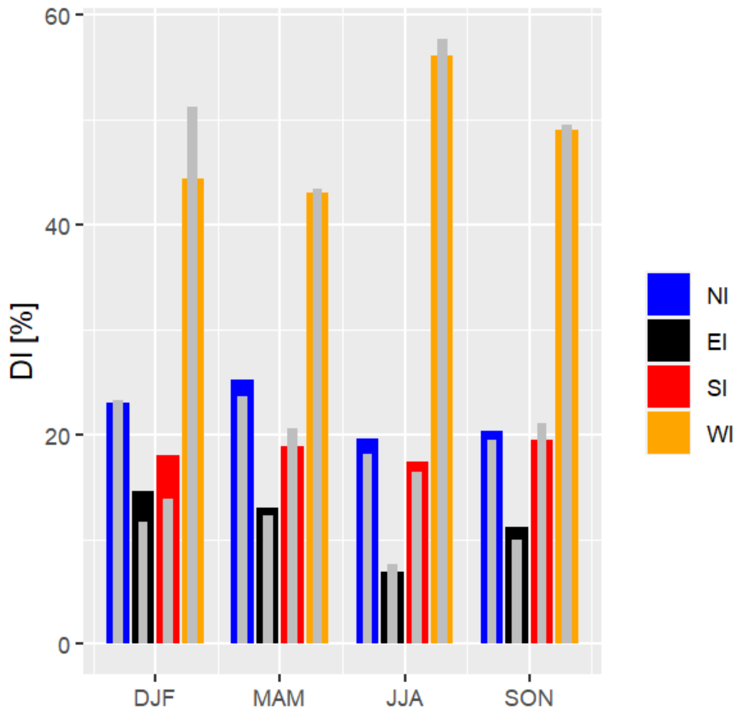

Figure 2: Seasonal mean of the DI of Morin's measurements (colored) and ERA5 reanalysis (1980-2020)

Brandsma, T. and van der Meulen, J.P. (2008), Thermometer screen intercomparison in De Bilt (the Netherlands)—Part II: description and modeling of mean temperature differences and extremes. Int. J. Climatol., 28: 389-400. https://doi.org/10.1002/joc.1524

Manley, G. (1974), Central England temperatures: Monthly means 1659 to 1973. Q.J.R. Meteorol. Soc., 100: 389-405. https://doi.org/10.1002/qj.49710042511

van der Meulen, J.P. and Brandsma, T. (2008), Thermometer screen intercomparison in De Bilt (The Netherlands), Part I: Understanding the weather-dependent temperature differences). Int. J. Climatol., 28: 371-387. https://doi.org/10.1002/joc.1531