

**Response to the review by Anonymous Referee #2 on the manuscript cp-2020-57
"Unlocking weather observations from the Societas Meteorologica Palatina (1781-1792)"
by Duncan Pappert et al.**

We thank the reviewer for their effort in carefully reading and commenting on our manuscript. In the following, we reply to their comments point by point.

Review: *This study reported the digitizing and compilation of the Societas Meteorologica Palatina (SMP) weather observation, a network of 37 stations across Europe plus a couple in North America and Greenland covering the decade of 1781 and 1792. The quality of the SMP temperature and pressure observation data is then evaluated by using the C3S Quality Control software to identify outliers and visual inspection. The potential of the SMP reconstruction for climate research is demonstrated by two extreme weather events of the reconstructed period. The reconstruction is rigorously done and described in great detail. However, the manuscript, in its current layout, requires major revision before it could be considered for publication in this journal.*

Specific comments:

1. ***The structure of the manuscript is not logically sound.*** For example, section 2.1 on "Source material description" and section 3 on "Inventory" could be combined and shortened by removing the repeating or loosely relevant information. Section 2.6 on "Homogenisation" and section 2.7 on "Generation of daily and monthly series" should be parts of the "Data processing (section 2.3)" work. Section 2.7 on "Generation of daily and monthly series" should also be placed before section 2.5 on "Quality control" and section 2.6, since a number of discussions in quality control and homogenization refer to the monthly data.

Reply: As the structure is a concern raised by the first anonymous Referee as well, the revised draft of the manuscript will clearly need some rewriting as well as a reshuffling of the structure. To provide an explanation, the data and methods chapter was written trying to preserve the order in which the steps were performed. For instance, "Quality Control" was performed several times over at different stages of the data processing (e.g. before and after homogenization) but mostly at the subdaily resolution - that is, before calculating daily and monthly means; but in order to perform homogenization tests, monthly means need to be calculated from the quality controlled subdaily data. Your point that a number of discussions in quality control and homogenization refer to the monthly data calls attention to the back and forth nature of this procedure. The methodology section will therefore be revised to ensure fewer jumps in logic. As suggested, Section 2.6 on "Homogenisation" and section 2.7 on "Generation of daily and monthly series" shall be incorporated into a larger data processing section.

Parts of the "Source material description" could be shortened and included in the "Inventory". The "Inventory" section is actually part of the results and is an output of the study, as much as the rescued measurement series. We will more clearly explain this in the Introduction and Methods sections.

2. ***The study could be presented in a more constructive framework and its significance to the broader audience needs to be clearly emphasized.*** For example, it would be more informative for the general audience, if the uncertainties and errors of the observations in different stations (section 2.5 on "Quality control") could be categorized by common characteristics, by regions, or even by specific years. In the demonstration of the two extreme weather cases, the reference to the CAP7 weather type lacks necessary explanation and justification.

Reply: Agreed, the significance of the study to the broader audience and its contribution to different fields of research will be contextualized better in both the introduction and conclusion. Regarding the specific errors and uncertainties from each station may take up too much space; each type and occurrence of flagged values varies strongly from one station to another, for anyone wishing to know details, these flags are marked in the individual files in the data supplement. Section 2.5 “Quality control” seeks to highlight the main problems that were encountered by bringing up some examples, such as the `wmo_time_consistency` flags for Munich and Zagan, or the `subdaily_repetitions` for Moscow. We will evaluate how to better summarise uncertainties in this chapter.

A sentence will be added to clarify the reference to the CAP7 weather type classifications by Schwander et al. (2017).

3. *The authors should make a clearer distinguish between what’s available to the public (i.e., the work already done by previous studies) on the Ephemerides and what’s new from this reconstruction.*

Reply: We will better delineate how our work stands out from previous efforts that dealt with these observations. This is not a reconstruction but rather a rescue of observations contained in the *Ephemerides*, which includes: 1) the creation of an inventory that may serve future research as guide; 2) the conversion and correction of temperature and pressure measurements from 37 stations for use in modern climate research. It is the first time SMP data for one or more variables from all 37 stations is published. So far, studies have focused on the use of few series, and in most cases these have not been made publicly available.

4. *Please explain briefly why temperature and pressure, but not the other weather parameters, are specifically selected for the reconstruction. Could the quality of reconstruction on temperature and pressure be generalized to the other parameters? What implications does it have on the overall potential of the SMP dataset?*

Reply: The revised manuscript shall briefly justify the selection of the variables. The reason behind this choice lies in the fact that temperature and pressure are arguably two of the most fundamental parameters in any analysis of weather and climate. With regard to the *Ephemerides* this has its advantages, as thermometers and barometers were among the instruments issued by the SMP that were carefully calibrated and standardized to ensure more precise quantification. The same holds for the hygrometer, though it is not clear what the exact units being used are, so the data rescue for this parameter would require more effort. Precipitation is of course another important parameter; however, rain gauges in the SMP were designed for local construction and hence come in several different units, some of them based on specific regional subdivisions based on other weight units. This does not necessarily correspond to lower quality or reliability, but it does again mean the work involved to process them correctly would require tremendous effort. Even the descriptions of the state of the sky could be useful but would require time, care, and a different way of approaching data rescue, perhaps being processed as categorical variables or a set of indices.

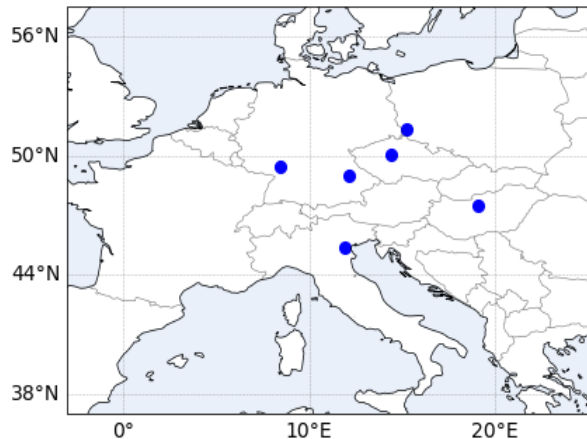
This explanation could be added to the manuscript either in the introduction when framing the scope of the study or in an outlook within the conclusion

Overall, it is hard to generalize the quality of temperature and pressure to other variables. The potential here lies in the sheer quantity of recorded data; the fact that these observations were taken following a set of agreed-upon rules already makes them more reliable than many other series from this time. Within this project, almost all variables measured by the SMP have been digitized on

excel sheets and are waiting to be used, processed and tested for their usefulness in climate research.

5. *Based on Fig. 5 there are more than 25 stations available in 1785, so please explain why only observations of six stations are presented in Fig. 8. What about the rest of the stations? Did they also record the cold spell in March 1785?*

Reply: True, more than 25 stations gave data to the Society in 1785 and yes, they do show the cold spell (to a different extent depending on their location). The selection in Fig. 5 shows a number of stations in Central Europe, the region with most SMP stations but also the region that felt this cold spell more intensely. The selection was made to represent this Central area ranging from Mannheim to Budapest and from Padua to Zagan, with two additional stations in between. Furthermore, plotting all 25+ series on top of each other in Fig. 8 would have been too messy; whereas here one can still clearly distinguish which line belongs to which stations. Nevertheless, a figure with more station series could be added to the electronic supplement.



Technical corrections:

1. *Ln 63, “In a first part” sounds strange to me, please consider use the common notation “In the first part”.*

Reply: Will be corrected.

2. *I don’t understand the exact meaning of the sentence “Overall, the quality of the temperature and pressure series recovered in this study is relatively high – due in no small part to the standardized thermometers and barometers made available by the SMP” (Ln 184-185).*

Reply: The homogenization test showed that most of the pressure series and virtually all of the temperature series agreed with each other, and therefore needed no additional corrections. This is not surprising given that the thermometers and barometers given to the stations were standardized and calibrated with each other in such a way that they could be comparable (see Ln 88-94).

“Relatively high” here means compared to other non-Palatina observations.

Instruments were in most cases quite good by the late 18th century, and problems with length units affected mainly precipitation (which the SMP did not solve). The most important thing that the SMP did is to spread best practices on how to use the instruments (e.g., isolating the outside thermometer from the wall with a wooden plate, measuring room temperature to later correct pressure readings). Providing standardized instruments with a common set of instructions to be gathered centrally was a good thing, though not because other instruments at the time were unreliable; it would just have been a lot more work to collect all information on each and every instrument and to make the necessary conversions/corrections (see Brugnara, 2015). Rescuing such

data for modern climate research requires a tremendous effort. These reasons make the SMP data even more valuable: the Society's insistence on precise standardized instruments and emphasis on coordinating observations to a common plan means that data rescue today can use these measurements with more confidence.

3. *Ln 287-288, Fig 5 does not give "which stations were more prolific", but how many stations. Please correct.*

Reply: True, it primarily shows how many stations contributed measurements, yet the last column "Period covered" shows the extent of each station's contribution, hence the use of the phrase "which stations were more prolific". Perhaps this can be expressed better.

4. *Ln 435-438, the whole sentence reads very confusing and needs clarification. What are "These differences" refer to precisely? What leads to "meaning that average March ..."?*

Reply: Will be clarified and explained in a more correct manner. The differences refer to the Budapest station in Table 4: the Palatina data show an anomaly of -5.6°C for the period 1781-92, whereas EKF400v2 shows an anomaly of -3.6°C , meaning a difference of -2°C . This means that the SMP observations presented in this study consider March 1785 to be more anomalously cold than the reanalysis EKF400v2. This could be due to: 1) the winter 1788/9 in Budapest is warmer in EKF400v2 than in the SMP observations or 2) average March temperature for 1781-92 (used to calculate the anomaly) are lower in EKF400v2 than the SMP observations.

Brugnara, Y., et al.: A collection of sub-daily pressure and temperature observations for the early instrumental period with a focus on the "year without a summer" 1816, *Clim. Past.*, 11, 1027-1047, <https://doi.org/10.5194/cp-11-1027-2015>, 2015.