

GENERAL COMMENTS

It is a pity that the analysis focuses on temperature only, especially considering that a lot of metadata are also provided for the other variables and that pressure data were corrected by Esmark himself (therefore should be of good quality). I suppose that the reason is that only temperature data have been digitised?

Pressure observations would be particularly useful to the research community, for instance for reanalysis projects (e.g., Compo et al., 2011). If they have been digitised, I would like to see in the manuscript at least some information on the pressure series concerning data quality. The series of Stockholm from the Bolin Centre could be used as reference, or one could also use pressure reconstructions (e.g., Luterbacher et al. (2002); Küttel et al. (2010)).

Reference series are not used for the homogeneity testing. The authors justify this choice with the unavailability of contemporary temperature observations near Oslo (Lines 355-356). However, I expect temperature series such as Bergen, Stockholm, Uppsala, and Copenhagen (all available in public datasets such as GHCN-M: <https://www.ncdc.noaa.gov/ghcnm/>) to be correlated enough to be used as reference series, at least at annual resolution. The absolute tests carried out by the authors give valuable information but could potentially overlook inhomogeneities due for example to changes in the instrumentation, therefore I think that the choice of not using reference series should be reconsidered and an additional relative homogeneity test should be applied after monthly means are calculated (Sect. 4.5).

SPECIFIC COMMENTS

1.INTRODUCTION

Line 67: Formally Austria-Hungary was only created in 1867. I suggest to use “in todays Slovakia”.

Line 91: Replace “reanalyse” with “analyse”.

Line 92: Add “temperature” before “data” (unless other variables are analysed in the revised version).

2.METADATA

Line 102: Remove “a” before “garden”.

Lines 199-201: Please rephrase.

Line 214: What does the “average?” in parenthesis mean? You don't know if the published data are averages or you don't know if the corrections are applied to averages only? Either way, can this not be verified from the weather diaries?

Line 233: I think the translation for “hevertbarometer” is siphon barometer. Please verify.

Line 247: I think you mean “can” instead of “might”.

Line 253: What is the unit of the pressure observations?

Line 256: drizzle.

Line 257: Does “Flokker” have no meaning?

Line 259: cloudy.

Line 273: “and” should be “a”.

3.METHODS

Line 345: What is the significance level that you adopted?

Line 349: Replace “calculated” with “converted”

4.RESULTS

Line 355: “For much of Esmark’s period of observation there was no other nearby station in operation so internal testing was the only possibility” I disagree on this. Monthly mean temperature anomalies at stations 400 km apart are usually still strongly correlated ($r > 0.8$) (e.g., Auer et al., 2007). It should be possible to use data from Sweden, Denmark for reference and integrate the internal testing results with a relative test.

Line 367: “But this break in homogeneity was much less than that of the morning observation.”
Less what?

Line 398: Replace “notes” with “values”.

Lines 418-419: Correct title and first line of the paragraph.

Line 449: What about the other seasons? According to your formula, you adjust the minimum temperature of 28 February 1838 by 2°C, while the minimum temperature of the following day is not corrected at all! I think here a correction function should be estimated for each month of the year (with some smoothing to better represent the annual cycle if the correction parameters are too noisy).

You also ignore other significant breakpoints without explaining why (e.g., 1835 for III vs II in spring in Table 2).

Line 453: I am not convinced by the evidences for an overheating of the midday observations. You compare with a station with arguably a very different microclimate (different elevation, distant from the sea, etc.). I think that your conclusions should be more conservative, considering the limited information you have on the thermometer exposition and the surrounding environment. You could say that a correction might be necessary for some applications (e.g., analysis of extremes), but for the analysis carried out in this manuscript I don't really see the point of applying such a correction.

5.DISCUSSION

Line 544: “Also the midday observation is warmer by Hansteen than by Esmark. This is harder to understand.” Isn't this because of the overheating correction that you applied to Esmark's data?

Line 567: Here also the differences with the Astronomical Observatory in summer (Fig. 12) are in large part created by the overheating correction.

Line 589: A Section 5.4 should start here.

Line 590: Change “was probably the greatest one” to “was the largest one in terms of sulphur mass ejected (Oppenheimer, 2003)”.

Line 591: The role of Tambora on the climate anomalies in Europe and North America is still debated. Besides, the “paradigm” of the Year Without a Summer is related not only to temperature, but also (and probably more) to precipitation and cloud cover anomalies.

Lines 605-606: Isn't it somewhat surprising that in Bergen, only 200 km or so from Oslo, 1816 was one of warmest years? Is this consistent with the instrumental temperature series of Bergen? Can you comment on the uncertainty of the reconstructions for individual years?

Line 608-616: You cannot reach conclusions on “weather patterns, excessive rain, frost, snowfall” in the summer of 1816 just by analysing the seasonal mean temperature. You should rather answer the question: How consistent are Esmark's observations with the results shown by Luterbacher and Pfister (actually, the temperature reconstruction that they use is from Casty et al. (2005))? It would be interesting if you could add the series of the nearest gridpoint of that reconstruction in Fig. 10 and comment on the differences.

6. CONCLUSIONS

This section is incomplete and too synthetic, it should be much improved.

TABLES AND FIGURES

Table 8: It would be practical to have an additional column with the reference for each reconstruction.

Figure 4: This figure is useless. Remove it.

REFERENCES

Auer, Ingeborg, et al. "HISTALP—historical instrumental climatological surface time series of the Greater Alpine Region." *International Journal of Climatology* 27.1 (2007): 17-46.

Casty, C., D. Handorf, and M. Sempf. "Combined winter climate regimes over the North Atlantic/European sector 1766–2000." *Geophysical research letters* 32.13 (2005).

Compo, Gilbert P., et al. "The twentieth century reanalysis project." *Quarterly Journal of the Royal Meteorological Society* 137.654 (2011): 1-28.

Luterbacher, Jürg, et al. "Reconstruction of sea level pressure fields over the Eastern North Atlantic and Europe back to 1500." *Climate Dynamics* 18.7 (2002): 545-561.

Küttel, Marcel, et al. "The importance of ship log data: reconstructing North Atlantic, European and Mediterranean sea level pressure fields back to 1750." *Climate Dynamics* 34.7-8 (2010): 1115-1128.

Oppenheimer, Clive. "Climatic, environmental and human consequences of the largest known historic eruption: Tambora volcano (Indonesia) 1815." *Progress in physical geography* 27.2 (2003): 230-259.