Clim. Past Discuss.,
doi:10.5194/cp-2016-60-AC1, 2016

Climate of the Past Discussions
Open Access
EGU

Interactive comment on "Jens Esmark's Christiania (Oslo) meteorological observations 1816–1838: The first long term continuous temperature record from the Norwegian capital homogenized and analysed" *by* Geir Hestmark and Øyvind Nordli

Geir Hestmark and Øyvind Nordli

geir.hestmark@ibv.uio.no

Received and published: 16 September 2016

The Editor Climate of the Past

Re: Cp-2016-60, 2016 Date: 16 September 2016, Oslo Dear Editor

Thank you for the positive and constructive referee reports on our paper on Jens Esmark's early temperature observations from the Norwegian capital. We hereby submit our replies to the comments of the three referees. More generally we have altered the

Printer-friendly version

Discussion paper

structure of the Results chapter to make it more logical and readable, and changed several figures for the same reason. For this reason, and because many of the answers to points raised by the referees refers to new amendments of the text, we also include as a pdf the proposed textual revision of the paper.

We thus hope that the paper is substantially improved and will approach publication in Climate of the Past.

Sincerely yours

Geir Hestmark Øivind Nordli

Replies to particular points raised by referees:

Ad ref. Przybylak: - The original fig. 4 has been deleted as superfluous, and figures after no. 3 renumbered in accordance with this.

Ad anonymous ref. 1: - The many suggestions for improvements of language and clarifications have all been adopted.

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).

- We think it is important not to use reference stations too far away, in particular when they do not represent the whole circle around the station under testing. In this way wrong conclusions might be drawn as spatial temperature differences could be inter-

preted as inhomogeneities. For the Esmark series there exist contemporary stations mainly to the east (Stockholm/Uppsala) and south (Copenhagen). From Bergen original data are lacking, and daily data exist only in these intervals: 1818.01-1818.05, 1823.08-1824.07, 1824.08-1824.09, mostly printed in newspapers, and mean values, of questionable quality. We will not use Bergen as reference for the homogeneity testing as it could "contaminate" the results rather than improving them. In the discussion part we have already compared Esmark's observations with other series from Oslo. Now this section is extended by binging in the Stockholm/Uppsala and the Copenhagen series. However, in the result part we exclude relative testing over those long distances (350 km and 450 km for Stockholm and Copenhagen respectively), also because (as mentioned above) the narrow sector available. Taking into account some further work (see text) and also your consideration we have dropped the correction for overheating of the midday observation. We have also changed formula for calculating monthly mean temperature. Now, we use Mohn's formula, which is more robust when there is not complete knowledge of the observation times. When correcting for the inhomogeneity in the evening observation, 1816-1821, a trivial sign error entered the original table submitted. This has been corrected.

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.

-Same point as above. If Esmark's series was given the same pattern of variability as the stations far away, important spatial climate variability could be hidden. However, we now open for using the available stations Stockholm/Uppsala and Copenhagen for evaluation of the final result, but detecting inhomogeneities by internal testing is our focus. It should be kept in mind that we do not know the exact observation times for the whole period of Esmark's observations. Here internal testing is the only possible

way for detecting changes over time.

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

-A reference to Table 2 is added.

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).

-We do not disagree, but prefer to keep this correction as simple as possible. The aim for the article is to homogenise a series of monthly mean temperatures only. Also, the correction is applied to minimum temperature only, so the amount of the correction on the monthly means will be much less than 2oC, cf. Fig. 7.

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

-New text has been added.

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.

-There are certainly good arguments for correcting the midday observations, but as you say we have limited knowledge of the microclimate at the station area. We have now made a new comparison with the station Oslo II, where thermometers were well

protected by the Astronomical Institute building, and this supports no correction, Fig. 9

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?

-No, the comparison was done before Esmark’s data were corrected (However, now they are not corrected)

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

-Yes, an important reason for dropping the correction of the midday observation, see above.

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.

-We now note these points in the Discussion

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?

-Bergen is situated in quite another climate region than Oslo – oceanic vs. semi-continental.. The distance between the cities is about 300 km. See new comment in the text and reference to literature.

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.

-Our conclusion concerns temperature only. However, it is important not to forget that Europe is larger than southern and western Europe. In the media we often see the summer of 1816 reckoned as the coldest one in our newest history. The present article contributes to a more nuanced view.

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.

-Good point. We have done this. See new figure and text.

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

-See new text

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

-Done -Relevant new references have been included.

Ad anonymous ref. 2.

While the authors have made extensive efforts to account for data quality and to homogenize the readings for long term climatic analysis in the face of sparse metadata, I am particularly uneasy about the lack of information concerning the observation times, and it is my opinion that further analyses may help reduce this uncertainty. In particular, the authors could make use of frequency analysis as exemplified by the work of Bergström and Moberg (2002) and Slonosky (2014) to compare Esmark's daily morning, afternoon and evening observations to the nearly 25 years of modern hourly observations mentioned in Table 1 and possibly obtain an approximate idea of the times of observation. It may be necessary to sub-divide the historical record for suspected

changes in observation time derived from the SNHT analysis and to consider the possibility of observation times, especially in the morning, changing with the season, if this is supported by other metadata (e.g the statement of observation times quoted on lines 188-189). If probable times of observation can be established, the entire analysis will stand on much firmer ground. As it stands, there are many adjustments made on a statistical basis which add to the uncertainty of the final values of the observations, particularly given the differences seen when compared with other nearby observations.

-We think that the difference between a relatively stable midday temperature and the temperature at times of the day when temperature are changing most rapidly is the most efficient tool for detecting changes in observation times. On the other hand the shape of the frequency distribution does not change much from one hour to the next, so it will not help us much. The main problems of Bergstrøm and Moberg (2002) and for Slonosky (2014) are not observation times, but the environment of the thermometers. We agree that for those purposes (and in particular for the data quality control) frequency distribution analysis is very well suited. See also our comments to reviewer No. 1.

The accounting of the adjustments due to inhomogeneities detected by the SNHT and other intra- series comparisons is extremely thorough and to be commended, but as is presented leaves the reader confused. A plethora of monthly adjustments is proposed in Tables 2-5, but it is not clear which adjustments were finally applied to which observations, the sequence of the adjustments nor whether the adjustments were applied to the daily data or to the monthly means. If daily, there will be artificial jumps between the end of one month and the beginning of the next - see Vincent et al (2002). In general, more use might be made of the advantages gained by having daily, rather than monthly, observations to analyze; much work had been done in the field of historical climatology in the past decade or two on analyzing daily observations directly.

-The Results chapter has been restructured to improve readability. We think that before starting the corrections of the observations it is important to have detected the

inhomogeneities. This is what we have done, cf. Table 2 and Table 3. We agree that to interpret the tables might be challenging, but we want to give the reader the opportunity to be able to better judge our work. You claim that "it is not clear which adjustments were finally applied to which observations". We think we have now made this easier to follow, as we have first a detection part (4.1) and then the corrections are discussed for each of the three shifts (4.2, 4.3 and 4.4). We also made it clear that the homogeneity testing was done with monthly and seasonal values, so we have no aim of adjusting daily values. From this article the outcome will not be a series of homogenized daily values. However, we agree that it is important to specify what data and how the results will be stored in the database at the Norwegian Meteorological Institute. See new text.

The fairly large differences shown between these data and other nearby stations, less than 1 km away, also give reasons for concern about the final quality of the data. Comparisons with other series, such as Uppsala -Bergström and Moberg (2002) - and Stockholm –Moberg and Berström (1997), although a considerable distance away, may still give valuable indicators as to the character of each month and help decide which series in the comparisons are the more reliable.

-Yes, we have now proceeded further with this comparison, see new text.

Finally, all the data, including the raw data, should be placed in an online archive.

-Yes, they will. This is easy in Norway, where the entire network of stations is freely available for everybody.

Introduction, line 45 and thereafter: "protocol" usually refers to a method; it would be less confusing if the authors could use a word like "logbook" or "weather registers" if they mean the actual physical records of Esmark's weather observations.

As Webster's Dictionary defines a protocol as 'an original draft, minute or record' and as neither editor nor any of the other referees have had objections to our use of 'protocol' here, we think the term is appropriate as used.

Line 350: The authors should take note of Gauvin's 2012 article on the Réaumur thermometer:The authors should be aware from Gauvin's work that theoretical adjustment of 1.25 for Réaumur to Celsius may not be accurate. This could help explain some of the large differences seen when comparing Esmark's values to the nearby observations in section 5.

-The problems related to the 'Reaumur scale' vs. other temperature scales were already thoroughly discussed in Middleton's book A History of the Thermometer and Its Uses in Meteorology (1966), where the conclusion is that by the 1780s a 'Reaumur scale' and 'Reaumur thermometer' had stabilized with thermometer manufacturers that was in fact rather different from the several scales and designs proposed by Reaumur, who died in 1757. For instance alcoholic solutions had been substituted by mercury, and the zero point defined by the melting rather than the freezing of ice etc. etc. Gauvin's paper mainly concerns the confusion surrounding this scale before this period, and is not particularly relevant (and largely recapitulates Middleton's work). The R-scales used by Esmark were certainly of the late 18th century kind, where notably the work of Swiss savant Jean-Andre De Luc had cleared up most of the confusion. It is today a standard convention to convert R-values to C-values by the formula R x 1.25 = C, and although some R-thermometers may not exactly exhibit this calibration with the C, such nonlinear deviations would be much too small to explain the large value differences in Sect. 5.

Section 4.1: The SHNT results seem somewhat ambiguous. Can the SNHT be run on all the 7665 days of observations, rather than dividing up into months and seasons? This might give a clearer indication of the actual break date. If this is too large a number for computational purposes, the series could be tested on running sub-portions (i.e first six months, move forward three months and test next six month period, and so on). Testing on other variables such as pressure might also give a potential indication of a change in the positioning of the instruments. It may help to further divide section 4.1 into subsections dealing with all of the adjustments to each of the three observation

Printer-friendly version

Discussion paper

times separately and consecutively.

-Barometric data have yet to be digitized, and as stated in the metadata we have no reason to believe that barometers and thermometers were situated in exactly the same place.

Line 380: A synopsis of the shifts and dates for each of the observation series would make these clearer. What were the final adjustments made to each series? A table summarizing the actual adjustments applied and the order in which they were applied would be helpful.

-See new text.

Line 381: The authors appear to be postulating a replacement of an hourly observation in the morning with a minimum thermometer. Hourly temperature observations and minimum temperature observations are not the same entity. If the authors think that a minimum thermometer was in use, a new series labelled “minimum temperature” should be analyzed. Rather than an inhomogeneity, this is a new variable.

-We think our arguments for the minimum thermometer are very strong. We put the minimum temperature to the ‘morning’ observation only because listed by Esmark as such. Of course information will be to find in the metadata file of the MET-Norway.

Line 390-2/871: This reasoning needs to be better explained, especially given the actual observation times are unknown. What does the description in the title of Table 4 “minimum temperature at 0800 UTC” mean?

-We have amended the text to provide a better explanation.

Line 405: More specifics are needed to explain this conclusion: 26% of interpreted “minimum” values being higher than the evening temperature is a high proportion. This unusual temperature trend is a situation which could occur with the passage of frontal systems overnight. What is the proportion of such unusual diurnal temperature trends in the modern record?

-In a modern record the minimum temperature comes from the same sensor as the ordinary temperature so this percentage will be 0%. On manual stations it is difficult to say because the meteorological institutes correct there minimum thermometer with a so called “formal test”. In practice this means that the minimum temperature will be corrected. In the case of Esmark one should expect more violation of this formal test as he seems not to have used a screen for the thermometers. For modern manual stations both thermometers are located in a common Stevenson screen.

Table 4/ Line 872: What is the authors’ interpretation of the negative summer differences for 1816-1828 and 1822-1828, compared to the modern differences? How are these differences changed with the selection of different observation times in the modern period (e.g. 0700, 0600, and sunrise?). Why is the period 1816-1821 corrected but not 1822-1828? Are these results from before or after the application of the adjustment of the 1821 inhomogeneity?

-Our first hypothesis was that changed observation time could explain the inhomogeneities, but this is hardly probable, see discussion. The answer to the second question is yes. See new text.

Line 442: How does this weakened diurnal temperature wave affect the reasoning section 4.1 concerning the minimum thermometer?

-The effect is zero as the problem with the minimum thermometer only concerned the winter. See new text.

Page 14, lines 447-451, Figure 5: Adjusting from one postulated unknown time to a second unknown time is a procedure beset with uncertainty, particularly as the linear trend does not appear to apply as well in the middle of the period, 1833-1836, when the points would give a much less steep slope. Have the authors explored regressions and residuals for other, finer time resolutions than the three-month period shown in Fig 5? What is the value of the sum of squares error? If better estimates of actual times of observation can’t be made, some portion of the data may just have to be classified as

unusable.

-The same procedure is used as for the other inhomogeneities, detection and then correction. For correction linear regression analysis is used with regression lines shown as equation (2). The standard error of the estimates is 0.15áţŠC.

Section 4.5 Again, how and why are these adjustment values derived? This is not clear.

-See the last answer and text.

Line 457, Figure 8: These adjustments should be presented in a Table separate from the Figure.

-We think tables and regression equation is sufficient in addition to Fig. 7 that gives an overview.

Line 482: Section 4.6 should be in the discussion section, while section 5.1 would perhaps be better placed as a summary in section 4. The comparisons with other observers and discussion of the thermometer error would be better placed in a data quality and comparison section, with the climatic discussion in a separate section.

-As these errors are not used in the analysis, they have been moved to Appendix B.

Line 541: Again, if we don't know the observation times, it's impossible to attribute the difference between the observers to a specific cause such as instrument location.

-For Hansteen we know the observation times. (for Esmark we only know the exact observation time in 1833).

Line 550: is 2100 UTC after sunset in summer?

-Yes

Figure 12: This would seem to suggest that the unadjusted values for Esmark are closer to the Observatory than the adjusted values.

-Yes, this is one of the reasons for not having applied correction of Esmark's midday observation in this new version.

Please also note the supplement to this comment:
http://www.clim-past-discuss.net/cp-2016-60/cp-2016-60-AC1-supplement.pdf

Interactive comment on Clim. Past Discuss., doi:10.5194/cp-2016-60, 2016.

Printer-friendly version

Discussion paper