Clim. Past Discuss., 7, C1828–C1833, 2011
www.clim-past-discuss.net/7/C1828/2011/

Climate of the Past Discussions

Interactive comment on "Benchmarking monthly homogenization algorithms" *by* V. K. C. Venema et al.

V. K. C. Venema et al.

victor.venema@uni-bonn.de

Received and published: 11 November 2011

Thank you for your kind words and valuable comments.

Referee #3: "One metric which is missing from the benchmarking study is one which assesses how well variability is reproduced at various timescales. I do not expect the authors to redo the study to include such a metric but it might be worth noting explicitly that this aspect was not considered. I also note that the error metrics don't consider extremes, but this is not so important for monthly mean data as it would be for daily observations."

That is a valuable suggestion. We have extended the sentence on further error metrics to study to: "Users may for instance be interested in the annual cycle, the cross-

correlations between stations, as well as secular trends for individual months, (interannual) variability, intermittence and long range dependence (Rust et al., 2008)."

We considered validation extremes as well, but rejected the idea as typically you do not have sufficient data to estimate extreme value parameters. Our COST Action is also working on a validation of daily homogenisation algorithms. In this study we do take extremes into account.

Referee #3: "Page 2665 line 11 – was the deseasonalisation only with respect to means? (i.e. anomalies from a reference period), or was some kind of deseasonalisation carried out on variability too? (and if so, how?)"

The mean annual cycle produces variance on the yearly scale (seasonal cycle). The IAAFT algorithm reproduces this variance. Deseasonalising the variance is not needed, it would not change the power spectrum or more specifically the variance in the spectrum at a yearly scale.

Referee #3: "Page 2667 lines 3-7 – it is not quite clear what you are doing here – do I understand it correctly that a seasonal cycle with an amplitude drawn from a Gaussian distribution with standard deviation 0.4 C is superimposed on an annual breakpoint with a standard deviation of 0.8 C? (i.e. if both 'random' points were at the one SD point, the mean annual breakpoint size would be 0.8 C, varying seasonally between 0.4 and 1.2 C). Also, what level of randomness (if any) is there in the time of year when the seasonal cycle of the breakpoint is at its maximum/minimum? (in general, section 3 would benefit from a couple of illustrative diagrams)."

You understood it perfectly. The seasonal cycle is implemented in a way that the minimum or maximum is in summer; see also our reply to referee #2 ("P2667 line 10). We have added: "The seasonal perturbations are computed by smoothing white noise and, if needed, shifting one of its extremes to the summer period." The requested illustrations are in the report on the benchmark dataset (Venema et al., 2011), for example, Figure 8 shows the (seasonal cycle of the) perturbations. An interested reader

CPD

7, C1828–C1833, 2011

or someone who would like to reproduce our benchmark can find more information there. As the manuscript is already quite long we did not want to add too much figures.

Referee #3: "Page 2668 line 26-27 – how do ratio-based methods deal with zero or near-zero monthly precipitation amounts? If there are no such amounts in your data set, then the paper should say so (but the issues of zeroes should be acknowledged, since, even if they do not exist in the test regions, they certainly occur in other parts of the world where the homogenization methods might be used)."

There were no zeros in our precipitation benchmark dataset. This is now mentioned in Section 2.2. Surrogate data section.

In praxis zeros can lead to considerable problems. Also in case of many values close to zero, the distribution of the ratio series is typically not well normally distributed. This is one of the fundamental problems of the homogenization of precipitation. The most typical pragmatic solutions are to set the zeros to 0.1 mm, or another small constant estimated by assuming that the distribution in lognormal, treat them as missing data or to ignore these values.

Referee #3: "Did any methods consider a combination of detection at the annual timescale and adjustment at the monthly timescale?"

Yes a number of algorithms did, sometimes directly as you describe it, sometimes making the date more precise on monthly data, sometimes by using the annual results as metadata for the monthly results, sometimes by mainly working on yearly data, but using an additional step using monthly data and sometimes by homogenizing all scales simultaneously. Please find such details in the report describing the contributions.

Referee #3: "Page 2679: It might be useful, if it can be done without too much difficulty, to present results on POD stratified by the size of the break – it would seem to me that an important aspect would be the ability of a method to effectively detect large breaks, with detection of small breaks being less important."

CPD

7, C1828–C1833, 2011

Indeed the most important aspect is to detect the large breaks consistently. One of the main advantages of being able to detect small breaks may well be to be able to determine the timing and size of large breaks better. This manuscript is intended as a first analysis of the benchmark results. We hope that a more thorough analysis of break detection as a function of size of the break, of the size of its seasonal cycle, of its timing, the length of the homogeneous subperiods and so on will be performed later.

Referee #3: "Page 2685 lines 14-15 – this wording suggests that the amplitude of the seasonal cycle of the breaks is a fixed 0.4 C, which contradicts the description on page 2667 (unless I have misunderstood it). Something somewhere will need to be fixed."

You understood it right. We have reworded is to: "the average seasonal cycle of the breaks inserted into the benchmark is 0.4 °C".

Referee #3: "Page 2685 lines 2-3: in addition to new technologies/measurement systems, there could be other reasons for inhomogeneities to have a tendency to be in one direction, e.g. a systematic pattern of station moves from town centres to airports."

This sentence has been reworded to: "This could, for instance, happen in case new technologies or measurement procedures are implemented; see introduction."

Referee #3: "Page 2688 lines 12-18: it should be noted that there may be some circumstances (although perhaps not in Europe) where absolute homogenization may be the only option because of a lack of reference series (e.g. stations in remote areas)."

We have avoided this discussion on purpose. Some people feel absolute homogenization may be applied if all other methods fail, like you. Not doing so would mean that we could not study the climate of much of the early period and many climatic regions.

Others feel that data is better flagged as being inhomogeneous in that case. Correction of a break known in the meta data is acceptable, but performing absolute homogenization both for statistical detection and correction may lead to results that are no longer independent of our subjective view of how the climate signal should look.

CPD

7, C1828–C1833, 2011

Referee #3: “Page 2689 lines 16-19: the comments on local trends are noted, which raises the question of how often local trends exist in the real world?”

Exactly, that is a good question. The estimate that 30% of the stations contain one trend was the best estimate of the homogenisation experts in the Action. As far as we know, there are no values in literature for this fraction. To get a more reliable number a careful study will be needed, which is difficult as the causes of local trends are often not recorded in the metadata and data analysis may fail because multiple breaks in the same direction are easily confused for a local trend and visa versa.

Referee #3: “Page 2659 line 24 – is there a reference for the Abbe criterion?”

The Abbe criterion is described in Conrad (1944), we have added this reference to the first sentence on the Abbe’s criterion. We have searched, but could not find a reference to a work of Abbe himself. Conrad and Schreier (1927), refer to Oppenheim (1925), which, however, does not give a reference and considers the Abbe criterion to be common knowledge.

Oppenheim, S. Über Schwankungen der Dichteverteilung der Sterne (On variability of the density distribution of stars). Astronomische Nachrichten, Vol. 224, p.283, Mai 1925.

Referee #3: “Page 2693 lines 22-23 – the meaning of this sentence is not clear to me.”

We wrote: “The use of metadata and reconstructions of past observation methodologies is preferred over statistical homogenization, especially in case sufficiently long parallel series are available and to precise the dates of the breaks.”

It now reads: “The use of metadata and reconstructions of past observation methodologies is preferred over statistical homogenization, especially to precise the dates of the breaks.”

Next to the changes requested by the reviewers, some further changes have been made that do not change anything scientific, e.g., changes in the affiliations, acknowl-

CPD

7, C1828–C1833, 2011

edgement, and units. Furthermore some obvious spelling errors have been corrected.

Interactive comment on Clim. Past Discuss., 7, 2655, 2011.