**Climate
of the Past
Discussions**

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

Anonymous Referee #1

Synopsis: This manuscript summarizes the most important results of COST-project ES0601. It is a well written paper with quite interesting new results and certainly worth being published. I have only a few minor comments regarding the presentation of the results.

-1): Title: Homogenization algorithms cannot be monthly. I would write " Benchmarking homogenization algorithms for monthly data"

0) There should be a small section that compares only the fully automated methods such as those from USHCN. They may be the most important for users who cannot devote as much time into homogenization as needed by manual methods.

1) Abstract last sentence: remove "currently", or do you expect that manual algorithms

will perform better in the future?

2) page4, line 14: to be consistent with the units of frequency, please write: at a frequency of roughly one per 15-20 years

3) p5, l3: systematically higher

4) One could omit the historical paragraph p5 l17-27, since it is unnecessary in this context.

5) p10 l14-15: This sentence is hard to understand, since the properties of data cannot match properties of a statistical test. I would rephrase it, e.g. as: While the surrogate data is most realistic, the statistical properties of the synthetic data are those assumed by most statistical tests used for homogenization.

6) p 10 l19-21: You should better explain why only 15 out of 20 networks have been used. The fact that five of them have not been homogenized well enough is not a valid explanation. Rather, the reader may suppose that 5 networks have been omitted to make the homogenization results look better than they actually are. Do the data of these five networks have some properties that cannot occur in reality? What is the reason why they are not homogenizeable?

7) p13, l1: Poisson process

8) p31, 8-9: What do you mean with "The size ... is operationalized...?". Please rephrase.

9) p36, l8: that it

10) Table 7 would be more readable if you stated the CRMS only for the inhomogeneous data and you used percentages instead of absolute values afterwards.

11) Fig. 5 is messy and very hard to interpret, at least for me. There must be alternatives for presenting improvements of trend estimates. It is also hard to interpret since results from different networks have been homogenized in different panels.

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

Interactive
Comment

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper