Clim. Past Discuss., 7, C1813–C1816, 2011
www.clim-past-discuss.net/7/C1813/2011/

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: 10 November 2011

Thank you very much for your review and your helpful recommendations. All suggestions that were simply implemented are not listed below.

Referee: "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."

In the conclusions there was already a paragraph on the advantages and disadvantages of automatic algorithms. It is not easy to distinguish between automatic and manual; most contributions lie somewhere on a continuum between these two poles and many algorithms can be used in various ways, which are more or less automatic.

We have lengthened its last sentence to state: “The study showed that currently automatic and semi-automatic algorithms (ACMANT, MASH, USHCN) can perform as well as manual ones.“ These three algorithms would be the ones we would recommend for user needing automatic algorithms to homogenize large networks.

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

While technically right, this is often the case with historical references and we feel it is a good tradition to acknowledge the scientist on whose work we build.

Referee: “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?”

This is an important point, which should indeed be explained better. The five networks have not been omitted because the surrogate networks in the benchmark dataset were not homogenized well. They have been omitted because the input networks to generate the surrogate networks were inhomogeneous.

If a network is not homogenized well, by definition its difference time series still show breaks and the (long-term) variability of the difference time series is too high. A surrogate network will also have this too high (long-term) variability in its difference time series. This surrogate network is assumed to be the homogeneous truth, however due to its long term variability in the difference time series, a homogenization algorithm may insert breaks to make the difference time series more like constant white noise. After inserting breaks into the surrogate networks, the long term variability in the difference time series may lead to detections of breaks that were not inserted or, less likely, may mask some inserted breaks. For the ranking of the algorithms this problem is not too

CPD

7, C1813–C1816, 2011

important; all algorithms suffer from this is a similar way. For the values we report on the influence of remaining inhomogeneities on estimates of (decadal) variability or secular trends, this problem is more important. Yes, these remaining errors are now smaller; we expect that these estimates are more realistic. Therefore we have decided so remove the 5 networks, which were most affected.

The synthetic networks do not have this problem, as their difference time series are constant white noise. In other words, this simpler method is not able to reproduce the long term variability of the difference time series. In a future study it may be an idea to explicitly change the power spectra used to generate the surrogate networks to remove or reduce the long term variability. On the other hand, also real networks will contain some long term variability in the difference time series. Thus this variability should also not be reduced too much. To know how much variability is natural in a difference time series, one would need long homogeneous (not homogenized) pairs of stations. Historical records for which we can be sure that they do not contain inhomogeneities do not exist, however. Thus the natural variability of difference time series is something that is hard to study. At least, in a next study we should make sure that the input data is homogenized better.

This complicated story is now explained in more detail in the new manuscript. It now reads:

"During the analysis it was found that some of the input data was not homogenized well enough. Consequently, the (long-term) variability of some difference time series in these networks is artificially too strong. The algorithm used to produce the surrogate networks is able and has reproduced this (long-term) variability, which the homogenization algorithms may interpret as inhomogeneities. Consequently, these networks had to be removed and only the best 15 surrogate networks were used in the analysis. Selecting stronger did not change the validation metrics anymore. For the comparison of surrogate and synthetic data, a new dataset was generated using only well homogenized input networks; see Sect. 6.3.1."

CPD

7, C1813–C1816, 2011

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

This is apparently a typical German/Dutch mistake. We mean "is parameterized".

Referee: "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."

The table would be more readable, but we feel that the absolute values of the errors are also interesting to the reader.

Referee: "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 "

In the interactive-discussion version the figure has become rather small and indeed somewhat difficult to interpret. In the full size on an A4 page, the figure is much easier to read. The advantage of this presentation is that you immediately see the improvement in the trend estimates by homogenization, can see that homogenization does not produce any systematically other trends, which is often feared outside of the scientific community, and at the same time can compare the performance of the six contributions. It would have been highly desirable if all participants had homogenized all data. That would have made the analysis much easier and the results much simpler to interpret. However, the consequence would have been a much reduced number of contributions. Especially, the consequence would have been much less contributions by manual methods, which have been much less well studied in previous intercomparison studies before.

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