

Interactive comment on “Concepts for benchmarking of homogenisation algorithm performance on the global scale” by K. Willett et al.

B. Trewin (Referee)

b.trewin@bom.gov.au

Received and published: 26 June 2014

This is a sound paper and I have only fairly minor comments on it. There are a few areas where I would have liked to have seen more detail, but that might prejudice the blind nature of the benchmarking so I can understand if it has been omitted; take suggestions to that effect as suggestions.

Specific comments

• I think it needs to be stated earlier than page 20 that this version of the benchmarking is taking place with monthly mean temperatures only. This point may also be worthy

C72

of some further discussion – particularly as the paper makes reference on page 6 to diurnally varying inhomogeneities, and maximum and minimum temperatures respond in different ways to some inhomogeneities. • I am assuming that references to the ISTI databank, number of stations, maps etc. refer to the beta version. With the first ‘official’ version due for release on 30 June these references should be updated to the latest version if possible. • Page 4 lines 9-13: in this context, CDRs could include single stations, sets of stations collated for the purpose, or gridded products. This could perhaps be spelt out a bit further, as could the fact that these CDRs could exist at global, regional or national/local scale. • Page 5 line 2-4: ‘SI traceable manner’ – I wonder if this might need a bit more explaining for a climate audience? • Page 5 line 11: should be ‘inhomogenities’ for consistency with ‘changepoints’ in previous line. • Page 5 line 21: ‘affect entire networks’ – might want to give a specific example (e.g. change in observation time). • Page 6 line 22: in this context, it’s worth noting that station-level homogenisation has been found in various studies to have more impact on spatial coherence of trends, locally consistent info etc. than on large-area averages. • Page 11 line 13-17: there isn’t much information (perhaps deliberately) on how the analog world’s spatial cross-correlations are constructed. My experience is that spatial correlation structures can be quite complex (though I’ve worked more with daily than monthly data), with stations being more closely correlated with geographically similar stations than with geographically different stations which are closer (for example, it’s quite common for a site on the coast to be better correlated with another coastal site 100km away than with a site 20km inland). A purely distance-based correlation decay function probably won’t accurately represent real data. • Page 12 line 18-19: ‘benchmarks will not include random error’ – noting that in the real world, a rapidly increasing frequency of errors (or a rapid deterioration of the site) may be the trigger for a site move; from my experience, the last 1-2 years before a move may not always be representative of how the “old” site had behaved in the long term. • Page 14-15: note all the examples of well-understood inhomogenities relate to instrument shelter changes – is there anything else which could be introduced here? (e.g. time of obser-

C73

vation changes?).

• Page 15 line 13-14: note that lack of digitisation of historical metadata is an even greater issue than lack of digitisation of historical data.

• Page 15 line 17-18: there are also cases where an apparently substantial change produces no detectable inhomogeneity – this can happen, for example, in some cases with a site move of several kilometres in flat, uniform terrain.

• Page 15 lines 25-29: an aspect which might be worth exploring is changes in detectability of changepoints over time – from my experience, older changepoints can be harder to detect due to reduced metadata availability and sparser networks to use as reference series.

• Page 17-18: agree that being able to define a “hit” is important, and that how a “hit” is defined will itself affect the results, and may favour particular methodologies (for example, if it’s tightly defined, will penalise methods which are imprecise in their timing but good in assessing the size of adjustment required).

• Page 18 line 25-27: note that for some methods, especially more complex ones which incorporate seasonality, weather type dependence etc., adjustments may not be easy to characterise with a single number.

• Table 1 line 4: while such changes can be gradual, they can also be abrupt (e.g. removal of vegetation, construction of building or paving of surface nearby).

• Table 1 line 5: I’m not quite sure why correction is not desirable in this case but is in some others? Seems to me to be a similar situation to line 6 (where correction may not be desirable from a physical perspective but may be for low-frequency large scale monitoring and detection/attribution).

• Figure 1: confirm this is current.

Interactive comment on Geosci. Instrum. Method. Data Syst. Discuss., 4, 235, 2014.