

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

### **Anonymous Referee #3**

Received and published: 27 July 2014

First of all, I would like to congratulate the ISTI initiative for their efforts, and for their intention to take methods for homogenisation into account in such a systematic manner. As the authors state themselves, the proposed framework will not change the global picture of surface temperature increases, but might substantially improve the regional representation of temperature evolution.

Still, before publication, I would like to address some major issues:

1. What is the intention of the paper, and how far has the benchmarking working group come to set up their framework? For a mere layout of intended work, the paper is by

C90

far too long, yet for a full description of the proposed experiment, it is maybe too short and unspecific. Currently, descriptions throughout the manuscript are still rather vague. I would expect the authors have -planned the specific setup of analog worlds (based on GCMs, based on real world data surrogates...) -worked on a selection of validation measures for different aspects -discussed potential experiments with different types of benchmark data sets (e.g., open, blind, different scenarios, regional aspects,...) I would like to see at least some more detail on these issues in such a paper. Currently it seems a bit, well, premature. Of course not all detail can be given, but direct links to pages with descriptions of the experimental design should be given. In line with this comment, the authors might think of changing the papers title. Instead of concepts, one might use the term "a framework" or "an experiment".

2. The authors should clearly state the assumptions and resulting limitations of their benchmark set given by Eq. 1. (=the linear decomposition): -there is no interaction between the annual cycle and internal climate variability. This is a serious limitation, e.g., in the tropics. It is known that El Nino is coupled to the annual cycle. Such an example should be given. -there is no interaction between long term trends and the annual cycle - also this might be questionable (e.g., winters might warm faster than summers) -there is no interaction between long term trends and internal climate variability. ENSO will basically stay the same, as other modes will do. Currently, this information is missing. It might be that I have misread part of the equation. E.g.,  $c$  is given a time index. As it is called a climatology, I presume that  $t$  refers only to time of the year. There might be further limitations resulting from Eq1 - If aware of any, please state them explicitly!

3. What is actually the mandate of ISTI? On your webpage you state that a proposal had been submitted to the WMO. What was the outcome of this submission? Anyway, you should mention the status here, as an official mandate would give your initiative much more weight as community effort.

Further Issues:

C91

Is it useful to call the surrogates analogs? I don't have a strong opinion, but analog sounds very much like a downscaling method, and in similar contexts, the term "pseudo" has been used, e.g., pseudo proxy, pseudo reality.

Abstract, line 3: "at all scales." But this is not what you do.

p238, l22ff: sections don't discuss etc.

In your list of homogeneities, you might consider to add that inhomogeneities, at least at short time scales, might be weather dependent. Although I am not sure whether this is of relevance or just an academic question (e.g., a building might act as a wind shield from a particular direction only).

p241, l1. Add "most" before previous. You discuss the home COST Action afterwards, were a different approach has been taken.

p243 explanation of v: you should explicitly mention that different modes live on different scales (space and time)

p244 first sentence: "these terms are ASSUMED..." See also might detailed comment above.

p244/245, l13ff: these paragraphs are a bit too vague. If you only want to present a concept, well. But wouldn't it be much better if you had agreed upon how to construct your analogs? How do you estimate m and v without a GCM? How do you separate forced signal from large scale from local variability? Also you should be careful not to construct different error analogs from the same clean analog. Just by having the combined information available one might estimate the underlying clean analog much better than in a real world situation where one observes only one realisation of inhomogeneities.

l 8: stations are not long but their series. l 13ff: work out clearly that an analog needs to be plausible, but not perfect. Otherwise lots of people will criticise the approach.

C92

p247 l15 replace "stations" by "records"

p247 l25 "return" sounds a bit strange. Adjust?

p248 please be more precise in explaining the covariate effect. The examples you present do not clearly make the case for an inhomogeneity - for me this sounds like a real climate response. Maybe you imply that d is amplified when the variability is amplified, but this is not stated explicitly.

in your bullet points you should add "shift of" for each item.

p249 l13/14: the sentence is not quite logical. Maybe it is enough to delete the "digitally" in the second half, but it depends on what you really want to say.

p250 how do you ensure blindness? If your analogs are based on GCMs, it might be easy to detect some of the inhomogeneities by just comparing the signal with the CMIP5 data base.

p250 l17: shorter compared to what?

p251 l 1 "why e.g.? Is there more possible? p251 l 2 "across a range of space and time scales" very unspecific p251 l4 "This information" more precise! p251 l17 add "detection" after "location" p251 l18 sliding scale? does this term exist? p251/252: at no place you discuss level 4! As you do it with all others, you should also do it with level 4. p252 l4 is there a better term than "correct misses"? "correct homogeneous"? p253 l9 but then you should separately evaluate these? p254 l5-10 this explanation is not quite clear, at least not how the last sentence relates to the previous. Wouldn't the overtuning occur if the clean analog would be released? p254 l16 here you mention for the first time which time scales you want to consider. This should be much earlier in the manuscript!