Benchmarking Working Group Call # 22

August 10th 2015 @ 1pm GMT (2pm British Summer Time, 3pm Europe, 9am Eastern USA)

Attending: Kate Willett (KW), Victor Venema (VV), Peter Thorne (PT), Robert Dunn (RD), Enric Aguilar (EA), Matt Menne (MM), Stefan Bronimann (SB)

Not Attending: Everyone else!

##
AIMS:
1) Catch up, its been a whole year! - where have we got to
2) Some specific points to discuss for Error Worlds
3) GitHub SurfaceTemp organisation for code
3) What are Kate and Victor doing for the rest of the week
4) AOB (next meeting)

ACTIONS FROM LAST MEETING:
- ACTION KW: Work on getting these distance/elevation/autocorrelation values from the real data.

DONE
- ACTION KW and IJ: Chat about Matern for future values possibly. Peter Challenor at Exeter may be able to help too?

Sort of done - got something working at least.
- ACTION KW: Try a better GCM with higher filtering

DONE
- ACTION KW: Just try it assuming MVN for stations we cannot model properly.

DONE - still in testing output mode
- ACTION MM: Look at a few more series and then make a decision.

PENDING
- ACTION MM: Give VV and RA the BFA/PHA results.

???
VV: I thought we already had this in Boulder, not? I cannot find in my mail, though. Would now be nice to study the properties of the inhomogeneities as a function of region. GHCNv3 is somewhat small for that; results are very noisy.
- ACTION IJ: Send newest version around - next week or two.

DONE - but Kate failed to follow up with any more calls to progress this further
- ACTION KW: Next call on this - 3 - 4 weeks.

NOT DONE
- ACTION KW: Update acknowledgements section and author names where necessary

DONE

ACTIONS FROM THIS MEETING:
ACTION KW: Add Robert Dunn to the official list of members on the website.
ACTION KW: Feed back to databank group on correlations / duplication and the issues of very short stations
ACTION: EA to send Kate some refs of field significance
ACTION KW/VV/RD: review the Benchmark world decisions to decide if there can be some mandatory and some optional worlds

##
AGENDA:

1) Catch up, its been a whole year! - where have we got to?
1.1) General (KW and PT)
Benchmarking framework paper published:
http://www.geosci-instrum-method-data-syst.net/3/187/2014/gi-3-187-2014.html
Talks presented at Edinburgh University, UK (KW), Bern University, Switzerland (KW), the WCRP
Grand Challenge on Data for Extremes, UNSW, Australia (KW), Australian Bureau of Meteorlogy,
Melbourne, Australia (KW) and CSIRO, Hobart, Australia (KW)
Posters presented at EGU (VV) and the Copernicus C3S workshop on Climate Data Requirements,
ECMWF, Reading, UK (KW)
Progress report submitted - November 2014
Kate to continue working on this at 5% over next 3-5 years.
Victor Venema visiting the Met Office in August 2015
ISTI Databank now v101
Robert Dunn has been in on a few of our calls now and is happy to become a group member. I think
he would make a very good addition and has good experience of working with climate data at a
range of scales. All on the call agreed.
ACTION KW: Add Robert Dunn to the official list of members on the website.

1.2) Clean World creation (KW)
Beta version now up and running for 32522 stations from ISTI databank v101 (downloaded July
2015)
- distance/elevation function to estimate correlations at lag0 and lag 1 in all cases
- pseudo elevation for ~600 stations from a digital elevation model so these can now be
  included
- not including anything looking like a ship ~30 stations (one is actually a New Zealand station
  with the wrong longitude!)
MM: Agree ships can be ignored
- not including anything with fewer than 36 months (~2000 stations)
- not including duplicate location stations - could possibly be included if location is nudged for
  purpose of estimating correlations but reported as original location
- assumption of exponential decay for lag 1 cross-correlations very poor but we're stuck with
  it because otherwise my matrices become non-positive definite
ACTION KW: Feed back to databank group on correlations / duplication and the issues of very short
stations
- assumption of Multi-variate-normal (MVN) for shock residuals where station is too
  short/gappy to assess in reality
- simulated autocorrelation distribution far too narrow, mean is a little too high, correlation at
  lag 0 too low - full testing still needs to be done. Hope its ok
- Code written to convert output to both ISTI and GHCN format
Still need to:
  - tweak distance/elevation function based on better initial data
  - test wiggliness of GCM smoothed fields added in
  - compare St Dev and autocorrelation of difference series between real and simulated ISTI
  - compare lag 0 and lag 1 between real and simulated ISTI
  - run PHA on output clean worlds to check they're 'ok'
  - possible rerun adding in the ~510 stations with duplicate locations - will take ages

Lund and Willett part 1 paper rejected from J. Clim but should be reasonably easy to rework the text
and resubmit somewhere else - not really novel/exciting enough for J Clim but still important enough
to publish - possibly GIMDS?

Willett, Lund and Chandler paper describing the Clean World creation (part 2) now mostly written pending final tests on output clean data - possibly GIMDS (Geoscientific Instrumentation, Methods and Data Systems: http://www.geoscientific-instrumentation-methods-and-data-systems.net/) or Advances in Statistical Climatology, Meteorology and Oceanography: http://advances-statistical-climatology-meteorology-oceanography.net/index.html

VV: It would be good it the clean world (and error world) match the ISTI dataset exactly. If it is so hard to generate clean data for short isolated stations, we could maybe also ask the databank group to remove such series in their merge. Can't be a big loss.
PT: Cautious about removal from the databank stage 3
VV: Could get that from stage 2
MM: They could be newly opened station - preference to keep in
VV: Very similar/close-by stations were also a problem, if I understood Kate correctly. They would be very valuable for homogenization. We should thus not simply remove them. If they are too similar it may also indicate that the merge does not work well enough yet. Maybe Kate could communicat the potential "problems" found with the database group.
PT: Same journal for the two papers
PT and VV prefer slghtly too difficult more conservative
EA: Field significance test for validation of clean worlds
ACTION: EA to send Kate some refs of field significance
MM: Looking for much less than 5%, a few hundred across a several thousand network
KW: Will run before mask with missing data before running PHA - should run before and after the mask.

1.3) Error World creation (VV)
We now have code so that the land use classification for ISTI stations should be able to be rerun for the new databank version (and when ever we need to rerun). This can be used to preferentially apply urban type gradual inhomogeneities (warming ones) to stations that are actually urban

VV: We still have to decide how to determine which stations are rural. Or decide where to find information to make that choice.

KW: The landuse classification comes from ESA CCI landuse dataset. This is a satellite product. They produce a snapshot of the globe every 5 years (three now available) where each pixel is classified by the predominant vegetation type. Pixels that are urban are classified as such, possibly even sub-urban too.

We have plans to use GCM net radiation at the surface and wind speed to inform some of the seasonal/interannual variability in the error worlds. These data cannot currently be accessed due to issues with ESGF - but we can get code set up ready.

Seasonal cycle in the inhomogeneities can be added by assuming a sine curve for most latitudes away from the equator. We can vary the size by pulling from a normal distribtion around zero. Mostly the phase will be aligned summer/winter but this could be varied if we wish. For tropical stations we plan to use a climatological curve based on GCM precipitation (interpolated to station) to try and match the monsoon (wet season/dry season).

PT: Careful with climatological precip which is the average of several wet years and several dry years and so not very informative.
KW: Good point but this is really just about getting some variability into the inhomogeneities, not necessarily realistic variability in terms of precise seasonal timing.

EA/PT: Maximum IH size should coincide with the dry season. Precip will dampen temperature effects.

KW/VV: We can make sure that this is the case - hopefully it may introduce some bimodality into some stations. We can go with something more sophisticated for future benchmarks.

KW: Does anyone currently use on information from other variables to apply inhomogeneity adjustments that vary intra-/inter-annually?

1.4) Validation (IJ) - not on call due to KW's choice of dates/times

1.5) Daily Benchmarks (RW) - not on call.
KW: Four USA regions and 3 to 4 worlds created, presented to homogenisors, and with response from 5 people (~9 tested algorithms). Rachel is now busy assessing the results and writing up. Aiming for completion by end of September.

2) Some specific points to discuss for Error Worlds (VV)
2.1) Gradual inhomogeneities:
Should they be linear, variable (changes given by normal distribution) or jumpy (changes given by thick tail distribution)?
See here for an example:
https://drive.google.com/file/d/0B16ocsUAaINSXzFmdTFJWHpET2M/view?usp=sharing
MM: Initial exploration suggests that a jumpy model is a better fit than a linear model.
VV: Could be linear with abrupt change points embedded though.

General feeling that its probably not linear but not really sure what it is. Best to explore across two worlds 0 one with jumpy-like and one with linear-like - could use world 8 for this. There will be cases where changepoints occur within a gradual trend anyway. If we specifically add changepoints then we need to be careful when assessing hit rate/false alarm rate whether these are counted or not - windowing of any other changepoints that happen to occur within the gradual inhomogeneity could become complicated. Having seasonal variability in the inhomogeneities will help especially when the interannual variability from the GCM goes on top - although this would mostly be removed by differencing with neighbours.

General chat about the optimal number of worlds and keeping the burden on assessment sensible.
ACTION KW/VV/RD: review the Benchmark world decisions to decide if there can be some mandatory and some optional worlds

2.2) Inhomogeneity size in gaps
Berkeley Earth found bigger inhomogeneities for gaps of more than one year than for breaks known in metadata. Should we increase the probability of breaks in gaps and maybe even size?
KW: If this is easy to do, and doesn't complicate validation too much then yes. I don't think it should complicate validation if we're sticking to basics (overall comparison of homogenised verses clean worlds and some broad scale detection efficiency
VV: What is the experience of the group with homogenized data? The results for Berkeley would suggest that most gaps of a year or longer contain a break. Giving all a break is probable not realistic and might help the homogenizers too much, would give them perfect metadata on our error worlds. Shall we say that 80 or 90% of breaks longer than 1 year have a break? And for shorter gap periods proportionally less likely?
VV: We could move the nearest break to the break, to implement these breaks. That would avoid changing the average break frequence.
VV: The breaks found by Berkeley Earth were on average about 1°C for gaps (less for breaks known in metadata). Shall we make the standard deviation of these breaks larger than the rest?

MM: Berkeley had a very high false alarm rate when tested agains the USHCN benchmarks of Williams et al. 2012 so we should be a little cautious over these results but in general it makes sense that inhomogeneities are often aligned with gaps.

PT: For the USA this is a fair assumption where all data have now been digitised so gaps in the record are where stations have been temporarily closed down. In other countries it could just be that the data for that period haven't been digitised yet

EA: This is the case in Spain where there were many gaps during/after the Civil War and stations resuming were not homogeneous with the earlier period.

General feeling that it is desireable to have a higher proportion of changepoints occur in gaps >12 months - 50 to 80% - this could be varied in different worlds.

2.3) Do we want to validate/rank homogenization algorithms or do we also have the ambition of trying to get uncertainties due to remaining inhomogeneities?

KW: Ideally the benchmarks would provide some additional information to help estimate uncertainties related to missed/adjusted inhomogeneities.

2.4) Bias in trends?

Best guess world would have a trend bias in the inhomogeneities of 0.2°C per century (from the average corrections of GHCNv3). Does my new research suggest that a higher value would be more realistic?

KW: I would be inclined to play cautious on this one - have some trend bias, but not so large. We can have a bigger one in version two if that is then the more realistic scenario.

VV: Would cautious mean erring on the side of a too large trend or a to small trend? Latter is harder to remove. KW: Too small.

PT: Make sure have worlds w/overall sign biases in both directions! Then can assess efficacy regardless of bias tendency and build confidence in what algorithm returns.

2.5) Outline of error worlds paper.

Combine that with the new GHCNv4 as a paper that is also about its homogenization results?

KW: Do you mean Error Worlds based on the findings from homogenisation of GHCNv4 rather than a combined paper with the GHCNv4 dataset paper? I don't think we want to be stuck on other's timescales - so if the info is ready to go then lets use it, if not, we already have enough to make a first version Benchmark I think without results from GHCNv4.

VV: I was thinking that a paper just describing our error worlds would not be terribly exiting. Especially if the statistical properties of the errors just come out of the blue air, our expert judgement. If we would study the statistical properties of the detected inhomogeneities in GHCNv4 to inform our estimates ("inform", not "set", because not all inhomogeneities are detected), that could make the paper more interesting to read. Would be too much detail for the general GHCNv4 paper, but still interesting for the homogenization community. We do not know much about inhomogeneities outside of Europe and USA.

KW: Fair point but again depends on timing. I think we have some interesting stuff to say - summarising the regional inhomogeneity info we have collected and some of the info from Berkeley and other earlier papers, how to go about implementing, getting the balance between reality and clarity of assessment - so how to design a set of worlds in a useful manner. It could be combined with the validation paper.

KW: I'm not opposed to an interesting exploration of GHCNv4 forming the basis - I just worry that we'll then take another 5 years to create the error worlds.

VV: I am also just putting up the question. Do not know whether it is worth the time. KW: Depends how efficient you are - more so than me I imagine!

PT: Careful not to bias towards PHA.

VV: Plan to use as a basis to build on - so no fewer IHs than are found by PHA at least.

PHA results on ISTI can be found here:
https://sites.google.com/a/surfacetemperatures.org/home/samsi_image_summer_program

4) GitHub SurfaceTemp organisation for code
Using git is very easy from R Studio.
To get R Studio working with GIT, I had to tell it where to find the binary of GIT in Tools|Global options|GIT/SVN. That was under Windows, may not be needed under Linux.

GIT software for your computer can be found here: http://git-scm.com/downloads
The Windows version adds contect menus to your Windows Explorer for most of the tasks you need to do: http://git-scm.com/download/win

The standard way of working with it is on the command line. Below is some information on how that works.
Short GIT info: https://training.github.com/kit/downloads/github-git-cheat-sheet.pdf
This lecture gives the basic ideas behind GIT: http://swcarpentry.github.io/git-novice/
From the same source there is also a video with the same content for the visual learners.

First also make a repository in your directory:
git config --global user.name "Kate Willett"
git config --global user.email "kate.willett@metoffice.gov.uk"
git init

Then link this git to the remote repository at GitHub. The normal command is:
git remote add origin git@github.com:SurfaceTemp/ISTI_Error_Worlds.git

I had to do (maybe because of a firewall):
git remote set-url origin https://github.com/SurfaceTemp/ISTI_Error_Worlds.git

Then it asks for your user name and keyword.

And pull the data from the remote repository to your repository:
git pull origin master

Later you will have to do the reverse:
git push master origin

Victor and Kate have set up a code repository in GitHub. We have set up an overarching organisation called SurfaceTemp that will contain projects for the Clean Worlds (ISTI_Clean_Worlds), the Error Worlds (ISTI_Error_Worlds), the Validation, and the ISTI-POST team. These should make it easier for us to collaborate on the code while its in development, and also to share/make public the code at a later date. All code should be written in an open language (R, Python) and well documented. Anyone who wants to can join the Organisation and contribute or at least keep an eye on what we're up to. VV The error worlds code can be found here:
https://github.com/SurfaceTemp/ISTI_Error_Worlds

5) What are Kate and Victor doing for the rest of the week?
VV: My ambition would be to get a basic version of the code running. Add more datasets, metadata, study values of parameters and complications later.

VV: Processing environment, break inhomogeneities, gradual inhomogeneities, seasonal cycle.

6) AOB (next meeting)
2 to 3 weeks

##
Notes: