Benchmarking and Assessment Working Group
Conference Call 4

Thursday 11th August 2pm GMT/3pm British Summer Time

Attended by: **Apologies in advance:**
Kate Willett (KW)   Steve Easterbrook
Peter Thorne (PT)
Victor Venema (VV)
Lisa Alexander (LA) (pending)
Ian Jolliffe (IJ)
Claude Williams (CW)
Robert Lund (RL)
Lucie Vincent (LV)

Actions from the last meeting:
KW: invite David Berry from the Marine Community to join - *DONE, awaiting reply* -
ACTION KW: chase
KW: publish Terms of Reference on website - DONE
KW: to suggest journals to submit to - *DONE* - *recommend JAOT**
Envirometrics? J. Clim? May be scope for a shorter concepts paper + a longer methods
paper.
CW: to chat to NWS colleague about downscaling issues - *PENDING*
Will follow up. VV may also be able to help (resource dependent)
ALL: to post on the blog one (or more if you like) world (Big Question to answer and
appropriate error world/s structure you would like to see with its justification. - *THANKS*
PETER, VICTOR, STEFAN
ALL: to email website improvements to Kate including links, references, documents etc. -
DONE, website updated and further plans are proposed
LA: There are a few too many colors. KW agrees, will tone down.
VV: I like the colors and they are not that bright.
VV: email example blogsite to Kate - *DONE*
KW: investigate page option on blogsite - *DONE and implemented* - *see 'Posts by Categories'*
under 'Pages' on the right of the page.

Actions for next meeting:
VV: put Kate in contact with anyone attending from COST HOME
RL - dates of TIES meeting in Hyderabad, India
PT and KW: get on with generic talks and posters
ALL: leader for validation - KW: if no-one appropriate I will take this on - I'm not a stats expert
but group members are. My preference would be to bring in someone new who is expert
though.
ALL: fill up teams with willing experts!
KW: setup email aliases: teamcreation@surfacetemperatures.org,

----

Purpose: Plan of action - building the error-free analog-known-worlds, the subsequent suite of analog-error-worlds, the assessment process and implementing the cycle of benchmarking.
----

1. Progress Update by Kate and anyone else who has anything (10 mins)
International Surface Temperatures Initiative:
- new logo - see www.surfacetemperatures.org
VV: The logo implies that the climate is warming. :-)
PT: could simply be max and min
LA: There might be a inhomogeneity in there too
PT: Logo is supposed to reflect the core aims of the initiative as a whole so multiple data products based upon observations, we tried not to show an extreme warming trend.
- Governance Document, Steering Committee Terms of Reference and Steering Committee Implementation Plan all now published on the website www.surfacetemperatures.org
- BAMS Meeting Report (from the Exeter ISTI kick off meeting) is available as early view online at: http://journals.ametsoc.org/toc/bams/0/0 doi: 10.1175/2011BAMS3124.1
- Kate to present ISTI update to GCOS Steering Committee - Reading, September 2011
- Meteomet collaboration - a metrology lead research campaign to improve traceability of observations to known standards. Includes 'harmonisation' of existing data which sounds like homogenisation to me. www.meteomet.org, kick-off meeting in Turin, October 2011.
- Surface Temperature Network - NERC funded effort to bring together all who work on surface temperature (satellite and in situ) - lead by Chris Merchant (Edinburgh University)
PT: I sit on the steering committee.

Benchmarking Specific:
- Benchmarking Terms of Reference now published on website
- Website updated - any thoughts/comments - further updates planned
- European Climate Data Management (ECDM) Workshop, Edinburgh, October 2011 http://www.metoffice.gov.uk/conference/ecsn-workshop/ - should we have representation there? Is anyone going?
ACTION: VV: put Kate in contact with anyone attending from COST HOME
IJ: New 'North American network for statistics in oceanic and atmospheric sciences' (Peter Guttorp) has proposed workshop on 'comparing numerical model output to data' which could possibly be relevant to validation phase.
RL: Meeting in Hyderabad, India - Jan 2012? - TIES - Robert Lund attending - could something Benchmarking related be presented
ACTION: RL - dates of meeting
RL: Joint statistics meeting last week, Richard Smith was cheerleading
PT: in process of preparing generic talks/posters for ISTI and Benchmarking

2. How to create the analog-worlds?

A first version of these needs to be available by November 2012 with a methodology paper submitted (shortly thereafter).

This will require some work-time from some group members. Therefore we need to try and align this as much as possible with our current research interests/plans. I have 10% of my time allocated for this so can do a fair amount of the coding work if necessary but ideally this will be a group effort - especially in terms of concepts and methods.

Questions for the group:**
Which part of this (if any) most interests you?
What (if anything) could you offer over the next 15 months: nothing - no time; ideas and expert advice; research/reading; designing/coding/building; analysis/testing?

Summary of intent:
Creation of at least one analog-known-world which is a replica set of stations in the databank created from downscaled GCM data with no systematic biases/inhomogeneities.
PT: I think you need more than one as some methods (all?) will have some sensitivity to phasing of natural variability and presence / absence of multi-decadal signal. In USHCN benchmarking (close to submission - Claude lead) we put the same error structure on four distinct runs and although its a second order effect to gross error structure it is still an impact.
Apply various error-models to the analog-known-world(s) to create a suite of analog-error-worlds exploring key problems for homogenisation methods.
Develop an assessment process.
Get the 3 year cycle of benchmarking and assessment up and running.

Basics:
- monthly land surface temperature (of daily means) but daily Tmax, Tmin and Tmean if time
PT: The databank will consist of Tavg, Tmax and Tmin but many stations and early records will just be Tavg. Tmax and min globally will only really kick in after 1950
VV: dont' know how to corrupt daily data yet - another project?.
- global set of stations but these could be a sub-set of the databank for now
PT: but this would not be any good for those looking at regional subsbsets - do need to include all stations
KW: build based on a small er subset but then do official release for all stations
VV: sub-divide whole world into ~100 regions with a bit of overlap for processing
PT: this could in some sense reproduce known regional governance of changes in instrument/prpractices
- ~10 error worlds
- code to create all of these should be legacy code that others can use: optimised*, well commented, open source (python, R?), with tunable parameters that are easy to tweak and flexible input/output so that different source stations can be read in/output.*

Idea 1 - TEAMS: All in favour

Team Creation: design and build the analog-known-world(s)
A team of people with knowledge in downscaling model data from grid box to point location and with knowledge of observational data features.
Kate
Robert LEADER
Victor
Steve?
Robert's PhD Student?
EMAIL:Peter
EMAIL:Lisa
EMAIL:Claude
Email: Ian

Team Corruption: research real-world inhomogeneity characteristics and design a suite of error-models to be applied.
A team with knowledge of causes and physical characteristics of land surface temperature inhomogeneities.
Claude LEADER
Lucie
Stefan?
Kate
Olivier?
Peter
EMail: Lisa
EMAIL:Ian
Email: Victor

Team Validation: design appropriate techniques to assess the skill of homogenisation algorithms in detecting and correcting for the correct location, sign and magnitude inhomogeneity.
A team with knowledge assessment statistics such as forecast verification and also analysis of climatological features (mean, variance, trends etc.)

VV: How to deal with homogenization algorithms which remove bad stations versus ones that do not. Must make sure we don't penalise removal verses attempts to homogenise. Is there a preference?

Ian MAYBE but really could do with someone else – has now volunteered (15/8/11)
Lisa
Claude
Robert
Lucie
Kate
EMAIL:Peter
Email: Victor

PT: Would it be worth considering whether there are funding proposal avenues that we could pursue? A number of us are in academic appointments and could put in funding bids. I wonder whether NSF might be interested and US folks could pursue? Also, PhD projects? Unfortunately, NOAA cannot pursue NSF funds... But fortunately CICS can :-)

IJ: Teams could involve people from outside the group.

Idea 2 - GROUP:
Creation phase: September-November 2011
- design and build the analog-known-world(s)

Corruption phase: December-March 2011/12
- research real-world inhomogeneity characteristics and design a suite of error-models to be applied.

Validation phase: April-July 2012
- design appropriate techniques to assess the skill of homogenisation algorithms in detecting and correcting for the correct location, sign and magnitude inhomogeneity.

Production phase: August-October 2012
- apply coding developed in prior stages to first version of the databank

3. A.O.B. / Next call (5 mins)
Next call proposed in two months (October).

PT: Just to remind you that a WG report is due October detailing progress. If you want a template we can set one up.
KW: Template sounds good - I will draft and then circulate early September so that we have group sign off by October.
ACTION KW: first draft of progress report - circulate by early September

PT: Would appreciate thoughts on whether at this point is appropriate to engage media over the progress of the initiative as a whole given the paper is published.
 VV: Good idea
 KW: very happy to do some media work on this

 CW: Google summer of code - Pairwise Homog code converted to python - code will soon be

available to anyone - availability of benchmarks should be timely - should be really useful.

 PT: paper on benchmarks for USHCN submitted to internal review - a conservative approach

VV: Validation group - what about terrible stations? What about algorithms that throw away stations verses algorithms that at least try. Have a compulsory validation? A compulsory subset? At least make sure we're not penalising against this.

4. Minutes agreed by:
PT
cw
IJ
LA
 LV
 VV

* Journal of Atmospheric and Oceanic Technology
JAOT covers research describing instrumentation and methodologies used in atmospheric and oceanic research,   including remote sensing instruments, measurements, validation, and data   analysis techniques from satellites, aircraft, balloons, and surface-based platforms; in situ instruments, measurements, and methods for data acquisition,   analysis, and interpretation; and information systems and algorithms. (up to 26 double spaced pages of text + figures and tables)

** Answers to Questions to the group:
KW:
Most interested in Creation and Implementation phases
Can contribute ideas, time to design/build, analysis/test, implementation

PT:
Prior experience in creation of these worlds - creation and corruption
Ideas and advice

VV:
All areas areas - main expertise in generating homogeneous data
No time to code, advice /expertise

LA:
Interested in all of it but time wise probably restricted to validation. Could provide some coding. Some interest in all - corruption but validation mostly. Could also potentially contribute PhD student

IJ:
Expertise is related to validation phase. Probably restricted to ideas/advice, but possibly a wee bit of reading/research. Need to keep in close touch with creation and corruption in order

to know how best to do the validation.

CW:
With the USHCN & GHCN work, we have many years of experience with Corruption! We have information about what corrections we have had to apply in both time and space. We also have suspected artifacts that we could give general information.

RL:
Creation, stats focus really,
army of PhD student and some time. Possibility of a student project with NCDC folks?

LV:
Creation/corruption/validation - corruption
advice, reading, testing - - no students so not much time for coding

SB:
Based on his blog post, I would say SB is mainly interested in corruption and its physical causes.