

ISTI Benchmarking and Assessment Working Group

5th Teleconference: Thursday 31st January, 2013 3pm GMT

Attending: Kate Willett (KW), Peter Thorne (PT), Claude Williams (CW), Ian Jolliffe(IJ), Robert Lund (RL), Mike Finney (MF), Victor Venema (VV), Steve Easterbrook (SE)

Not Attending: Lucie Vincent (LV), Lisa Alexander (LA), David Berry (DB)

Summary of ACTIONS from this call:

- i) ACTION KW: Kate to contact Robert/Ian/Victor/Lucie/Mike for help over the next week.
- ii) ACTION KW: send code/data to Robert and Mike and Claude (within 1 week)
- iii) ACTION KW : Circulate revised draft - all to sign off (or comment) by Monday please
- iv) ACTION KW: Reply to all in response to Ian's email
- v) ACTION KW: Send round a doodle poll.
- vi) ACTION KW: Redraft time lines and circulate with Prog Report
- vii) ACTION KW: Kate put date in diary to discuss WG meet up later this year.

- viii) ACTION CW: report back findings in the next meeting
- ix) ACTION CW: email round to kickstart this

x) ACTION RL: Robert to send Peter a link to the call. Contact NSF - start a proposal - feedback on the next meeting.

Agenda:

1) Progress report - Kate to update all on progress to date (of ISTI in general and benchmarking) and present the edited 2011 progress report - all members happy to sign off?

PT ISTI Overview: The Initiative overall has been progressing, although at a slower pace than envisaged in the first Implementation Plan. The databank is about to go to version 1 which of course represents the version that the steering committee expects this working group to spin their benchmark analogs of. Latest run (1/30) with a blacklist almost completely constructed and tested, is at 32K stations globally (+/- the odd hundred). More on progress is available in a slew of postings on the initiative blog at <http://surfacetemperatures.blogspot.com/> . Once the databank is released we will be trying to encourage groups to go in and create products and submit their algorithms to the benchmarks so the benchmarks do need to be available within some reasonable timeframe. We will have two posters at EGU. We also have a piece appearing in EOS next week led by Jay Lawrimore outlining the databank. A longer, more substantive, exposition the of databank is under preparation led by Jared Rennie.

KW WG Overview – progress against objectives:

- a) submit concepts paper to JAOT or similar – this is in first draft but needs work
- b) methods/software and paper for analog-known-worlds – the methods are now outlined and coded in open source software R using MCDW monthly temperature data for the USA from ISTI databank - issues with matrix algebra need working out.

ACTION KW: Kate to contact Robert/Ian/Victor/Lucie/Mike for help over the next week.

ACTION KW: send code/data to Robert and Mike and Claude (within 1 week)

c) methods/software and paper for analog-error-worlds - no progress

d) methods/software and paper for validation - no progress

VV: Shall we start with these activities, we cannot code yet, but we can already make up our minds what we would like to do, no need to wait for the final homogeneous worlds for that?

IJ: d) can go in parallel with c), rather than waiting until all of c is decided (IJ)

CW: There has been some current work at NCDC on Iceland data - analysis of errors

ACTION CW: report back findings in the next meeting

ACTION CW: email round to kickstart Team Corruption work

IJ: Team Validation can get going once Team Corruption start up.

e) create platform for guiding users - no progress

f) publicise the work of and need for benchmarking - on going

RL: Proposals on big data now part of an NSF funding call

ACTION RL: Robert to send Peter a link to the call. Contact NSF - start a proposal - feedback on the next meeting.

VV: What is big data? RL: daily, multi-variate?

VV: Homogenization of daily data is a great topic for a research project.

PT Assess progress against dedicated resources – resources are limited and so progress may not be as much as we had hoped.

ACTION KW: Circulate revised draft - all to sign off (or comment) by Monday please

2) *Timeline - Kate to present an estimate of deadline estimates for future objectives - all members happy to sign off?*

PT: I think its fine to concentrate upon concrete deliverables expected in 2013 and then more aspirational timeline for activities thereafter. The steering committee is likely to create a revised Implementation Plan for the Initiative as a whole in the summer so those longer-term details beyond 2013 can be fleshed out then.

VV: What is the timeline for second cycle with daily data? Proposal to homogenise ISTI daily data.

KW: Rachel - will hopefully have something towards the end of 2014

RL: working on that too - theoretical aspects of single changepoints in daily data.

Michael Robbins - U. Missouri.

ACTION KW: Redraft time lines and circulate with Prog Report

Questions outlined for Team Corruption:

VV: if there is a bias in the trend due to the inhomogeneities how well can this be corrected? Which worlds will be blind and for which ones the solution will be visible.

PT: What if there are no inhomogeneities? Ha ha ha

VV: We had one network in HOME without inhomogeneities, was no problem. PHA was extremely good for that case, very low FAR.

PT: What if breaks are nearly all small or nearly all large?

PT: What if breaks are steps, steps and slopes, steps and seasonal cycles etc?

PT: What if biases are the same in Tmax and Tmin or distinct in Tmax and Tmin? etc. etc. etc.

But we need to minimize the number of questions.

3) *Creation method overview (see TeamCreation_VARI_JAN2013.pdf) - Kate to present, Robert to say if that is the same as his understanding of the method - any thoughts? Kate to update progress in coding in R.*

PT: Main thought is not to let the perfect be the enemy of the good. What you have is far more detailed than anything I've seen before (although Victor's work is comparable in its realism). Therefore I would humbly submit that unless anyone cleverer than I (e.g. all the rest of you) spots gross errors I haven't you should aim to lock this down very soon.

VV: I would say that it is less realistic as HOME :-), but we could probably simplify and still be sufficiently realistic. For example modelling the lagged crosscorrelations may not be necessary, the cross correlation matrix and auto-correlations functions is likely sufficient.

VV: what about awkward stations that do not have close neighbours or overlaps?

PT: can make them temporally complete and then mask with actual missing data - part of Team Corruption?

KW/RL: Will try and do as many stations as possible but there may be some that are problematic.

ACTION KW: Reply to all in response to Ian's email with queries on Team Creation methods

VV: possible to simplify the method a bit more – this first time around doesn't need to be so complex.

KW: It could be run on regions/blocks rather than entire globe - presently running on gridbox level but can grow this out to the optimum size.

4) *Other team updates: (not expecting much here as I know you're waiting on Team Creation to some extent but a 'thumbs up' that you're still happy to lead that team would be good)*

Team Corruption (Claude Williams)

PT: I think we need very soon to have fleshed out what questions we wish to (primarily) address in the first benchmarking cycle and therefore what sort of 'corruptions' to add. The benefits accrued from benchmarking will be highly if not perfectly correlated with how well we set up this 'problem set'. With say 10 analog cases there are only a finite number of questions that we can address. The main thing, for credibility and sustainability is going to be winnowing down to a well defined set of analogs that together address systematically one or two questions? The alternative is a more 'random' set of choices but here the substantive risk to my mind is that we then don't really know what we learn / expect to learn when the algorithms are run across these benchmarks.

KW: Once ideas a formulating - circulate around a handful of key homogenisers for feedback. (May 2013)

Team Validation (Ian Jolliffe)

5) *Discussion of Benchmarking cycle concept - is this still workable/the best structure?*

PT: I don't see an alternative. Without a cycle with discovery and rebooting the problem set the value rapidly diminishes with distance in time from benchmark creation. Once NCDC for example run the PHA on the databank to create GHCNMv4 and submit to your analogs until we get feedback and a chance to confront our

algorithm with what challenges accrued we are fighting blind vis-a-vis making any further methodological advances.

KW: Main concern of 2.5 years being a long time to keep results a secret - maximising use to community verses preventing over-tuning

PT: keep timeline flexible at this time - will have a better feel for it when we have an idea of the known players

VV: May not have to keep everything secret - can have a mix: realistic ones secret, truth for difficult worlds intended for testing the limits can be published immediately.

SE: Staged release?

6) *Date of next meeting*

Suggest 6 weeks time - 11th-15th March

Doodle poll send out mid-Feb or set date now?

ACTION KW: Send round a doodle poll.

7) *AOB*

PT: I will be sending a link over the email distribution list within the week to a google spreadsheet that we plan to use to identify potential investigators who may submit algorithms to the databank and benchmarking and track who has approached who and what the response has been. There is a definite role for members of this WG in helping move this forwards. I like to think of it thus: The framework aspects that ISTI has direct control over are like constructing a really neat playground with large sandpits, cool slides etc. What ISTI cannot directly control is whether the scientists wish to come and play in the playground. But the value of the playground is going to be directly proportionate to the number of folks who come play in it. This is where we need our advocates. More later in this regard.

SE: Meeting up?

EGU2013: VV

South Korea: IMSC: KW, VV, IJ

September/October for a US meet up - code sprint could be co-run in USA and Europe.

Banff - 2015 Benchmark cycle wrap up - needs to be submitted September 2013.

ACTION KW: Kate put date in diary to discuss WG meet up later this year

Downscaling methodology:

MF looking into it

VV: recent paper on downscaling precip - Douglas Maraun 2012 - send around the list

IJ: EU project COST Action VALUE on downscaling

VV: Possible future project - Marlis Hofer - can we support?

;-----

OTHER USEFUL NOTES/UPDATES (please add):

KW invited to present at 12th International Meeting for Statistical Climatology, Jeju, South Korea, June 2013

IJ - forecast verification session coordinator

VV: <http://www.imsc2013.org/sub/Program.asp>

Rachel Warren (KW and IJ's PhD Student) is now underway looking at creating daily temperature benchmarks - she may also attend 12IMSC

Mike Finney (RLs PhD Student) to look at creating synthetic daily data.

New Members: Mike Finney, Rachel Warren and Giuseppina Lopardo

VV to present at EGU on ISTI but including work on benchmarking