

International Surface temperature Initiative steering committee call

4/4/12 12Z (8am EDT, 13 BST, 14 CST ...)

Dial in details distributed under separate cover

Present on call: Peter Thorne (PT), Richard Chandler (REC), Kate Willett (KW), Michael de Podesta (MdP), Jay Lawrimore (JL), Albert Mhanda (AM, on etherpad, unable to dial in), Antonio Possolo(AP), Blair Trewin(BT)

Apologies in advance: Chris Merchant, Rob Allan, Greg Strouse

1. Review of actions from the previous call - PT

a. Actions that were still open

ACTION: REC to discuss further with JL, PT and others on potential article for Significance on collection of papers on "large data", as and when plans are firmed up by editor.

REC has contacted the editor to ask about this. A response was received during the call: "Yes, the Big Data issue is indeed something I am actively pursuing, and pursuing now, because I am hoping for it to be the August issue; though it might slip to October on account of some leads drying up. But your ground-plan of a piece on organising, visualising, processing etc vast quantities of data, hooked on to climate change, would be ideal for us.

If we can still make it the August issue the deadline for a first draft would be the end of May. Would that be possible for you?". REC indicated that he could probably meet this deadline with appropriate support from a couple of members of the steering committee / Databank working group - to be followed up offline.

ACTION: REC to follow up with steering committee and databank WG.

ACTION: KW to discuss with AM membership from meteomet on the benchmarking WG and report back.

KW: Contact was made but then there hasn't been any collective BAWG action since last Steering Committee Call.

b. New actions

ACTION: PT to circulate a document to CCT members before ITS9.

The progress report was circulated to relevant CCT avenues.

'Richard writes: it would be helpful to know what the problem is here. If it has to do with the generation of Gaussian fields on the sphere (a problem I discussed with Victor Venema in this context a few months ago), I can probably find someone who can help - or, failing that, provide a reasonably robust "belt and braces" solution.'

ACTION: REC to contact KW about this.

REC: Ahem. Ooops, sorry, not done!

2. Updates on progress

The over-arching progress report was published in early February following iterations within the steering committee and given to our 'sponsors' for consideration. To date any feedback on this is still pending.

There will be a brief discussion at the Global Climate Observing System Atmospheric Observations Panel for Climate in late April at WMO headquarters given by Matt Menne on behalf of initiative participants.

KW: Follow on (to some extent) from Kate's presentation to GCOS Steering Committee. There will also be an item on a global climate surface network. Matt Menne will present this.

2.1 Databank working group - JL

Good progress is being made. Now over 40 sources collected - of daily and monthly temperature observations. Jared Rennie has been hard at work integrating new source datasets and also doing a lot of coding and testing associated with development of the monthly stage 3 dataset.

Please see the latest Stage 1 and Stage 2 source data at <ftp://ftp.ncdc.noaa.gov/pub/data/globaldatabank/>. There are more than 130,000 station records in the stage 2 monthly.

Peter provided an presentation on the Databank at the recent 9th International Temperature Symposium (along with several other presentations).

A certificate of appreciation - which was completed in November - has now been used in thanks to contributions of data from providers in Spain and Germany. On request we will be happy to print and mail this certificate to any contributors in thanks for their efforts to support the ISTI.

Regarding Stage 3 development -- We are working to develop a fully automated process for merging the sources of stage 2 data into a single stage 3 dataset. This involves 2 primary steps. Comparison of metadata and comparison of data between a 'Master' station and 'Candidate' stations. Before a merge is attempted a heirarchy of the sources is attempted - a prioritization - the most complete sources with the best provenance, of the

best quality along with other factors are given preference over data from sources that are of a lower overall quality.

The merge process begins with the source of highest priority (Master stations). The stations from the source of next highest priority (candidate) are compared to the stations in the master source. This is done to determine if the candidate stations can be simply added to the master source as a "unique" station (does not already exist in the Master) or whether the candidate station is already present in the Master and whether it can be merged to create a longer and more complete station record.

The 1st step - comparison of metadata between the master and candidate stations - was completed a couple of months ago. This is based on comparison of station id, location, and elevation. A probability is established from those 3 characteristics and any master-candidate pair with a probability of sameness of at least 50% are held over for a data comparison.

The past couple of months have been spent developing a method for testing overlapping data between the master and candidate stations. A Master station having data from 1910 to 1980 and a Candidate station with data from 1940 to 2011 - the data from 1940 to 1980 can be compared to help determine if the two station records originated from the same station. The Databank group is closing in on an approach that is based on Index of Agreement converted to a probability that a master and candidate station are the same (and should be merged) and a probability that the candidate is unique to the master dataset and should be added as a unique station.

It is also possible to test the non-overlapping data for a master and candidate station. The potential presence of inhomogeneities and the potential for real differences in climate between two separate time periods complicates a test of non-overlapping data. The Databank team is currently evaluating the option of developing a test of non-overlapping data after the release of Version 1 of the Databank.

There is no single correct method that can be developing for creating the stage 3 merged dataset. Because of the complex nature of weather and climate observations, reporting, collection, dissemination, and in many cases unknowns due to the passage of time and poorly kept metadata (or lack of availability) there will certainly be errors of omission and commission. We will likely provide several stage 3 merged datasets with a single "best" stage 3 dataset. Based on a recommendation from the Databank working group, the best merge process will err on the side of identifying too many candidate stations as unique versus incorrectly identifying master-candidate pairs as the same station.

We were hoping to release the 1st version of the Databank in April but we will miss that target by a couple of months.

PT: Note that the code will be included in the release and can be run open source. It will be modular in nature

KW: When do we need to think where the benchmarks will be posted?

JL: Discuss in May.

REC: Have you written up on what we are doing?

JL: There are some broad brush details in the ITS9 paper.

Information is available on a wiki

http://editthis.info/intl_surface_temp_initiative/Main_Page

REC: Reservations around non-overlapping record tests.

PT: We thought quite hard about this and agree with the concern about looking at the station in isolation, so we are thinking of using neighbors that are contiguous.

REC: Different perspectives on need for such merging whether a data supplier or user. User dependency.

PT: There are clear flags on where the source came from that will be pulled through. So, the user will know when they have been merged and which portions arise from where. The user will also be able to avoid the merge all together and simply look at stage 2 data if they wish to.

2.2. Benchmarking and assessment working group - KW

I've not been able to give any time to this since January really.

However - things are not all bad.

I have successfully set up a PhD jointly between the Met Office and Exeter University to investigate benchmarking of daily data. We now have guaranteed funding with Met Office sponsorship and an excellent student ready to start in October. While this won't help the initial release of monthly benchmarks it should really help to build the science behind the concept of benchmarking and make it workable at higher resolution.

I've just developed a pilot method for creating pseudo-station data from climate model output as a baseline but using real station characteristics. This will only work well where there are relatively dense stations (>1 per model gridbox). This is computationally and mathematically simple. I've written this up and will run it by Team Creation over the next two weeks so Team Corruption should have something to play with very soon.

I'm due to visit Robert Lund (Clemson) and NCDC (Claude Williams) in May for a week. My main aim here is to get the analog-known-worlds and analog-error-worlds up and running.

I'm also in the middle of setting up a meeting between the UK's National Physical Laboratory and the Met Office. Part of this visit will be to discuss all things ISTI and benchmarking.

BT: Arising out of Australian data some results on benchmarking to put in.

KW: What level of detail is wanted?

JL: Development wiki?

KW: We have the blog but this is possible.

3. Report from ITS-9 - PT / GS/ AP / M&P

ITS-9 - a once a decade meeting of temperature metrology - occurred in late March in California. PT, Matt Menne, and all metrologists (GS, AP, MdP) on the steering committee attended and GS and AP gave talks on the issues.

PT gave a plenary talk on the initiative as a whole. I can report that nobody walked out of the hall. Perhaps they were all just being polite.

The session on the analysis and understanding of the data was well attended with an update on NIST efforts to understand data issues and data analysis problem set. Matt Menne also gave a talk upon the NCDC algorithm and benchmarking.

MdP: The climate stuff came over quite well. There is a lot of very high level stuff too.

4. Updates on publications

Chandler et al is now in the pre-pub queue at Environmetrics. Can be found at <http://onlinelibrary.wiley.com/doi/10.1002/env.2141/pdf> (paywalled)

Metrologia piece is stayed pending clearance of time for GS / MdP.

Williams et al. paper describing benchmarking of USHCN (the US surface temperature record) has been published in JGR.

Papers on the initiative as a whole and the databank effort have been submitted to the ITS9 proceedings. As have several other relevant papers.

AP: Three NIST papers - GS on local observing network, 2 papers on the changepoint detection and homogenization.

KW: poster presentation at the WCRP 4th International Reanalyses meeting. A chance to discuss how reanalyses fit within the ISTI framework with the main reanalyses providers.

5. Check of progress against in-year items from the published Implementation Plan

Being midway through the year (ish) it seems worth taking stock of where we are against the things that we said we would do.

http://www.surface temperatures.org/steering-committee/Surface_Temperature_Initiative_Implementation_Plan_version1_release.pdf?attredirects=0

p18 forwards for the full list. I am just excerpting here the first column and then appending any notes.

Metrologia publication

See above

ITS-9 participation

See above

Release of version 1 of the temperature databank (monthly and daily)

Will not happen in April (see above) but is still in hand.

Benchmarking Position paper submitted for peer review

KW: not done. This will be pushed back until after July (IPCC deadlines). Hope to submit alongside release of benchmarks ~November.

Paper describing databank first version and underlying principles submitted to literature

Databank WG are currently of the mind to pursue a generalist paper and a technical paper. Importance of publication is recognized.

JL: Good basis from ITS9.

Creation and release of pilot benchmark analog-known-worlds and analog-error-worlds (monthly data)

KW: Planning on a release of the actual benchmarks ~6 months lagged from the initial Databank release with pilots released when they're ready. I think this is still achievable.

[PT notes that analog was pointed out to have a very specific meaning in metrology and some other circles and we may want to revisit the nomenclature of these benchmarks accordingly.]

KW: That's good to know - KW and MdP to discuss off line if there is no time within the call?

Creation of official cycle 1 benchmark analog-known-worlds and analog-error-worlds (monthly data) for official release and release of the analog-error-worlds

KW: see above

Instigate access and visualization working group

Nil to date. Thoughts on how to push this forwards?

KW: There are a lot of data portal and visualisation/user tools efforts going on in various different places. Can we piggy back to some extent rather than trying to repeat things?

PT: ACTION next call to discuss the five issues raised in the progress report to which steering committee members were assigned to work on.

6. General updates from the steering committee on ISTI relevant news

6.1 Any relevant conferences we should be thinking of attending?

PT will attend the first meeting of the EarthTemp initiative in June in edinburgh and will travel with as many posters as the organizers allow him space for. BT also plans to attend this meeting.

AM planning to attend Edinburgh workshop

KW will attend WCRP 4th International Reanalysis meeting

BT: Pacific islands workshop in New Caledonia in May. Will give presentation on the ISTI.

ACTION: JL send BT the certificate.

6.2 Any new developments of which we should be aware?

The new Australian daily homogenised temperature data set was launched on 23 March - we believe this is the first national-level data set of this type. Details and associated documentation at <http://www.bom.gov.au/climate/change/acorn-sat/>. A blog post about this is in preparation.

BT attended a meeting between WMO and BIPM while he was in Geneva in February. BIPM are considering a joint committee on measurement issues related to climate science and seeking WMO support for this, unclear at present what WMO view on this is. Note that BIPM already have a consultative committee for thermometry. Areas BIPM have an interest in include traceability, in situ calibration, quality assessment protocols and uncertainty evaluation of historical data (possibly through replicating historical instruments).

Hope to have launched or be close to launching the new RMS/Wiley-Blackwell online Geoscience Data Journal by the next call - RJA

7. AOB

Next call July 11th @8EDT