

ISTI steering committee call

10th November 13Z (14CET, 13 GMT, 8EST)

Present on call: Peter Thorne, Blair Trewin, Victor Venema, Akiyo Yatagai, Xiaolan Wang, Jay Lawrimore, Antonio Possolo, Richard Chandler

Apologies in advance: Andrea Merlone, Kate Willett, Michael de Podesta

Actions arising

Peter to chase Colin Morice on issue of tagging stations with CCI land-use metadata and advise.

Jay and Jared will work with NCEI IT Branch to get the data loaded to the ISTI ftp site. (~10 monthly benchmarks more essential naturally than Rachel's Daily benchmarks, but shall try to mirror those also.)

Peter to chase Steve Worley for update on whether land ICOADS was raised at the ICOADS meeting.

Steering Committee members to send suggestions for any standard plots that may be useful to produce on a monthly basis in the RezaTeC effort to Richard Chandler.

Agenda

1. Review of actions from prior calls

1.1 All hands call

ACTION: Peter follow up on Madagascar data for POST

Latest this week from Tim Oakley who is en route to Madagascar:

- *Thanks Peter. The parallel data set from the 'HQ' location was only of marginal use for a comparison as there were a number of issues with the observations, both manual and automatic. I am hoping that follow this visit we will have several sites working, which are the official CLIMAT record, so should have something of benefit for ISTI.*

Victor: If the instruments are only at the HQ to show how they look like, this is okay. If they are used for national or international climate assessments, we would be interested, even if the quality is bad. The difference between two instruments can also be due to differences in the quality of the data from one instrument to the next. I was actually worried that the parallel measurements, which are often performed at meteorological offices, are of too high quality and that the differences in the wild are larger.

ACTION: Victor Venema to contact Albert Klein Tank regarding ETCCDI involvement in POST.

Victor: DONE. POST will use the ETCCDI indices where possible. ETCCDI would like to extend their indices for moderate extremes with indices for weather variability. We agreed in general how they should look like (decomposing the variability on different time scales (interannual, annual (seasonal cycle), intra-annual and daily). If POST is first with an implementation we will send it to ETCCDI for review.

ACTION: Colin Morice to follow up with David Parker and Jared Rennie on associating databank stations with land-use CCI data to provide land-use (modern) metadata per station.

No update. Peter to chase Colin Morice on issue of tagging stations with CCI land-use metadata and advise.

Blair is keeping an eye on emerging WIGOS metadata standard. Station identifiers discussion is the current focus and likely to be relevant.

ACTION: Kate Willett to include Jared Rennie in discussions on CF conventions for stations.

We have now had a lot of discussion in house about DRS structures, versioning (using ESGF and native version control), updating logistics and related file grouping structures, standard names, AMIP like variable names, use of CF tools such as attributes (cell_measures, status_flags, standard_error, ancillary_variables, comment), standard modifiers and global attributes. I'm writing up our plans and will share with Jay/Jared/Matt when these are complete. Ideally it would be good to do something that is common to multiple dataset builders. I've emailed Jay/Jared/Matt about this (only just) and hope to arrange a telecon shortly to

discuss. We are planning to develop improved netCDF files and dataset structures for both gridded and station data so much of this work could be useful for the ISTI netCDF format (and vice versa). I could give a 5-10 minute talk about this at the next call or just circulate the documents. Very happy to include others in these discussions...

ACTION: Kate Willett to contact Jay Lawrimore and Jared Rennie about linking daily benchmarking dataset on NCEI data homepage.

Rachels error worlds are currently here:

<http://www.metoffice.gov.uk/hadobs/benchmarks/>. I have the clean world data ready to post so will try and get that done today. I think it would be great if the data could be made available/mirrored from NCEI and or ISTI too. I have emailed Jay and Jared about this but only just.

Jay and Jared will work with IT Branch to get the data loaded to the ISTI ftp site. (~10 monthly benchmarks more essential naturally than Rachel's Daily benchmarks.)

XW: are the monthly benchmarks also available on the website? (not yet, maybe next June/July)

ACTION: Kate Willett to contact Peter Domonkos about his Spanish project to see if there are any useful linkages between the benchmarking groups.

This will be followed up at a later date with Enric Aguilar.

ACTION: Colin Morice and Victor Venema to discuss possible PhD student Victor: *DONE*. *During my meeting at the MetOffice we discussed the research proposal on a new homogenization method. If funded we will collaborate on it. If you are interested in details of the proposal, I guess Colin should disclose that.*

ACTION: Steve Worley to raise land ICOADS in upcoming ICOADS meeting. *No update Peter to chase Steve Worley for update on whether land ICOADS was raised at the ICOADS meeting.*

1.2 Last steering committee call

ACTION: Blair, Peter, Victor to draft abstract giving ISTI overview for the St. Gallen meeting

Done and presentation given - any feedback Blair? Not directly - with this audience (lots of homogenisation/QC type people), main objective was to build interest in taking part in benchmarking once it is up and running. Victor has reported more extensively below on POST aspects.

ACTION: Peter to contact Peter Siegmund of KNMI to initiate an entry for ISTI in the IDARE data portal.

Done see <http://www.idare-portal.org/content/dare-success-stories>

2. Updates from WGs and task teams

2.1 Databank WG - Jay

The latest version of the ISTI Databank was released on Oct 15. v1.1.0. This update contains data from 67 sources. The additional sources come from parsing of data from one large global source into the individual sources it was originally comprised of. In addition, one source that was in v1.0.0 was removed from v1.1.0 because it was determined the source had threaded station records (a join of two or more stations) and higher-ranking sources had these separate components already. The changes incorporated into the new version are considered a moderate update to the databank and as such is reflected by an increment in the second digit of the version number and the preparation of a technical report. The report is available on the surface temperatures website. It describes updates that include the changes in sources as well as changes to the merge program that ensure the most accurate data are incorporated in the merged (Stage 3) product. The primary change was an adjustment to the metadata threshold used during the merge process. More recent analysis showed that some merges were forced between stations that should have been retained as two separate unique stations. In v1.1.0 a stricter threshold of 0.75 was applied that reduced the likelihood of this occurring. A second change was a relaxation of the requirement for the minimum length of a data gap in the target station needed to accomplish a merge of a candidate's record. In v1.0.0 there needed to be 60 months. This was changed to 12 months, which ensured target stations with short gaps would be filled with qualifying data from candidate stations.

The progress plan for the past year is being completed and goals for the coming year established. This will include further improvements to the merge algorithm, incorporation of additional sources, and to establish a plan for multi-element data holdings.

Peter: Is plan still to release a v1.2.0 which incorporates as many of the 'new' stage 2 sources that have been queued as is possible using the current method in the interim?

Victor: What is the idea regarding mirroring?

Jay: At one point it was mirrored in Russia but has fallen off the radar.

Richard: The rezatec effort may enable some mirroring. Envisaged to be long-term.

2.2 POST - Victor

End of October POST had a side meeting at the Data Management Workshop. We now have 22 member and 27 associate members. Next to the large number of members, the interest in and support of our work is great. The WMO had already requested its members to help building a parallel database. The new Task Team on Homogenization (TT-HOM) of the Commission on Climatology now also endorses the project. The WMO Regional Association for Europe is interested in the transition to AWS and has offered to send out a POST survey to their members. WMO CIMO has also sent a survey to its members. The Status Report of GCOS includes text about POST.

Up to now getting data we know exists has been relatively easy. Data policy is not a big problem. It helps that parallel data is seen as experimental data, not as an operational product, but this is likely also because people see the importance of this research. Sometimes we could "only" get 5 years of data or could "only" publish the indices, but not the daily data itself. The problem is finding the datasets and their owners. To make it more attractive to help us find parallel datasets, data providers are offered to co-write our first articles and we will also produce quicklooks on the web with results for individual datasets so that data providers can gain insight to their own data.

We want to compute ETCCDI indices as far as possible for our short series. The ETCCDI is now planning to add indices for weather variability, which is an addition we also wanted. I have contacted Albert Klein Tank to coordinate. There are also indices from the WMO ET-SCI, which are "sector-relevant" focussed on users in health, agriculture and hydrology. Currently for temperature and precip; extensions are planned. On a first look these indices also sound interesting for us. Like for ETCCDI sometimes we will have to make changes for our short series.

At EMS we had our first POST talk about the transition from conventional observations to AWS. Please find below two more detailed presentations for temperature (POST-temp):

http://www.meteoswiss.admin.ch/content/dam/meteoswiss/en/Forschung-und-Zusammenarbeit/Internationale-Zusammenarbeit/doc/Slides/11_Aguilar_Slides_web.pdf

And for precipitation (POST-rain).

http://www.meteoswiss.admin.ch/content/dam/meteoswiss/en/Forschung-und-Zusammenarbeit/Internationale-Zusammenarbeit/doc/Slides/12_Stepanek_Slides_web.pdf

The first results for temperature show that the bias in the mean is very different from country to country. Thus we really need a large cross-country database to make meaningful estimates of biases in the global temperature record. It is already clear now that the screen is more important than the sensor. The countries with an automatic probe in a Stevenson screen had only small inhomogeneities. The automation of the precipitation measurements seems to lead to a decrease in precipitation, especially during winter. All conclusions are still quite tentative as they are still based on too few datasets.

Next to the above-mentioned two studies on the transition to AWS, we have a study on the transition to Stevenson screens, which is coordinated by Theo Brandsma (POST-early). And a study on the influence of relocations, coordinated by Alba Gilabert. Last week was the first telecon of POST-move. To study relocations we hope to be able to get data where the same instruments are used at both locations, but will also gather data with historical instrumentation. Depending on the data we can get we may study relocations to airports (including cases where the old station historically continued) or relocations from cities and villages to better sites. The latter would potentially have more datasets, but would also require better metadata and similar instrumentation.

There are plans/ideas for humidity, wind and sunshine duration observations. However, this would require additional people working on it or will have to be done later. Interested people please contact me, so that I can see if we have a critical mass for new studies.

Our software is not fully finished yet, which was ambitiously planned at this moment. (On the other hand we have progressed more with gathering data and analysing it for the mean.) We do now have a GitHub code repository for our processing software: <https://github.com/SurfaceTemp/ISTI-POST>

The main problem will be to obtain data from outside of Europe. If you have any contacts outside, please bug them, you never know what they know.

Peter: Long-term data policy?

Victor: Open upon publication.

2.3 data rescue task team - Peter

Nil by update. This has fallen off of the radar with other things that are going on presently.

2.4 Benchmarking WG - Kate

Not much to report.

Still waiting on co-author edits of clean world methods paper. These edits may result in some change to the methods but hopefully nothing major.

Progress report now written and submitted - aiming for Summer 2016 for release of the benchmarks but we still have quite a bit of work to do on the error world side of things.

3. Update on Rezatec activity - Richard

Rezatec have now developed a web-based interface to access and analyse the ISTI data. However, it is currently very slow (we know what the problems are, and they are fixable), and therefore probably not yet ready to share with the steering committee. At present, the capability includes:

- Display current ISTI stations (database updated monthly) on a map, with "standard" Google map tools
- Move around the world, select a rectangular (lon, lat) region of interest, download data for all stations in the selected region and carry out some simple visualisation / statistical analysis
- Analysis tools currently include time series plots, histograms, calculation of summary statistics, boxplots of monthly temperature distributions and the option to view and download the underlying data table.

The analysis and visualisation part of the interface is done using RShiny, which enables the development of web apps using R. I am working with Rezatec to resolve the speed problems, to improve the underlying R code and to develop

some further capabilities (e.g. calculation of different types of anomaly for each station as well as some more advanced statistical analyses). When it is a bit faster and more stable, I will share the link with the steering committee and invite feedback; we will deal with any (constructive!) comments and then open up the interface to the world at large.

Jay: Is it slow due to calculations on the fly?

Richard: A variety of fixable reasons.

Victor: Could it calculate standard plots monthly?

Richard: No reason why not in principle. Please send suggestions for standard plots should produce.

Steering Committee members to send suggestions for any standard plots that may be useful to produce on a monthly basis in the RezaTec effort to Richard Chandler.

Jay / Peter: Will we be able to look at the homogenised series down the line?

Jay: GHCN-Mv4 beta is available at

<ftp://ftp.ncdc.noaa.gov/pub/data/ghcn/v4/beta> .

Richard: At this point not planned to include but may be able to in future.

4. General updates - All

Peter: I provided a review of Kevin Cowtan's recent piece published this last week. The whole thing is at <http://www-users.york.ac.uk/~kdc3/papers/homogenization2015/index.html> . Kevin is interested in potentially further developing his method and participating in the benchmarking exercise. We should keep him informed and encourage further methodological development.

Peter: My 5 million euro grant application got write in reviews 2 of which were excellent, one was very good and one less so. We wrote a 2-page response and the decision now lies with the funding body. Given the apparent urgency of the review response expect to hear either way before Christmas but no guarantee that we shall do so.

Peter: I will give an invited talk providing an overview of ISTI in the Earthtemp session at AGU. Its on Friday am early slot for anyone who is planning to attend AGU.

Kate: I've been invited to speak on ISTI at a meeting in February in Cambridge (Computational and Data Challenges in Environmental Modelling, http://www.turing-gateway.cam.ac.uk/cdc_feb2016). They asked for a talk on the use of new technology to collect novel environmental data related to the ISTI. I said that I wasn't sure I quite fitted the bill but that I could talk about data

source and quality issues, archiving, rescue digitisation etc and they still said they were interested so we'll see what happens.

Richard: I know exactly why you received that invitation :-) Happy to discuss if this would help.

Jay: GHCN-M v4 which uses the Databank Stage 3 data, followed by QC and bias corrections will likely be released into NCEI operations in January. (Mean temperature only. Reconciliation with max/min/dtr later in 2016.) Mean temperature Beta currently available as noted above.

Victor: Will talk at MEDARE mainly POST, also some ISTI.

Xiaolan: Next IMSC will be in Canada June 6th-10th not far from Calgary. (Canmore, Alberta, Canada)

Antonio: Nov 2016 a workshop by NIST and PTB, maths stats and metrology - 2016 some talks on atmosphere - when call for papers is available shall circulate.

XW: As usual, there will be a session on data homogeneization at EGU 2016 (17-22 April 2016, Vienna, Austria). Open for abstract submission now.

5. Full year report

We have a report from benchmarking group to hand and shall have for databank shortly so no issues foreseen in completion of a progress report this year on time in full. As ever we'll have progressed somewhat less than hoped which is down to this being almost entirely voluntary effort with no dedicated funding.

6. AOB

We'll aim to have an all groups call in January.