ISTI Steering Committee call

12th May @12Z (14 Europe, 13 UK, 8 EDT)

Present on call: Peter Thorne (PT), Victor Venema (VV), Blair Trewin (BT - possibly late joining), Kate Willett (KW), Jay Lawrimore (JL), Xiaolan Wang (XL), Michael de Podesta (MdeP), Andrea Merlone (Will leave before end), Richard Chandler

Apologies in advance: Rob Allan, Albert Mhanda

Actions arising

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

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

1. Updates on working group activities

1.1 Databank WG - Jay

Monthly Databank
The monthly databank has gone through a number of routine updates since the release of version 1. We have set up a near real time update system, which incorporates the latest data (including GHCN-Daily and CLIMAT streams). These updates occur on a monthly basis, and the newest data is posted on the FTP site no later than the 11th of each month. These updates have been the basis for testing and evaluation of the next version of GHCN-Monthly (version 4), which includes updates to its quality control and bias correction processes. We have an end to end process on our developmental server, and we are currently calling it version 4 alpha 1 (v4.a.1).

During the evaluation of version 4 alpha, we have taken a closer look at some of the specifics of the databank version 1.0.0 stage 3 merge and have identified a number of cases where the merge can be improved.

Increased the metadata threshold from 0.50 to 0.75, requiring matches on metadata to be stricter. This is to correct cases of station records merging with the wrong records from higher ranking sources. The merge problems resulted from the use a low metadata metric threshold for allowing data comparisons, which led to errors that favored data merges over adding unique stations.

Updated two stage2 sources (ghcnsource and russsource) to incorporate their ID's in the metadata matching part of the merge. Including the ID's helps match stations through the merge's ID test module. Also reduced the gap threshold from 60 months to 12 months in order to help piece together stations with small gaps, especially during the defined base periods.

We are evaluating the removal of some homogenized data sources with joined (threaded) station records when these records are also included separately as higher ranking sources. In these circumstances, the merge algorithm has difficulty reconciling the two sources.

Will also be updating new source data that has been in the queue for quite some time. The cutoff date for new sources has passed (2/28/2015).

Additional details and graphs are available on the ISTI Databank website in the document
http://www.surfacetemperatures.org/databank/WG_Update_20150506.pdf?attredirects=0&d=1.

Next version of the Monthly Databank Stage 3 will be version 1.1.0.
PT: Had discussed on WG call whether this might be v2 instead? Has this been decided against?
JL: After further discussion determined it's best to go with a moderate update given the nature of the changes.

Daily Databank/GHCN-Daily
The daily databank plan for this year had the addition of three new sources. The first of these was completed within the past month. Approximately 1200 stations from the Global Summary of the Day (GSOD) dataset were added to GHCN-Daily. This ensures that all of the RBCN stations that submit CLIMAT data are represented in GHCN-Daily. These data will be incorporated into the next Monthly databank merge and allow a direct id match between the CLIMAT WMO number and GHCN-Daily. It also permits the calculation of monthly mean temperature directly from daily data for all CLIMAT sites in real-time, which will help overcome CLIMAT transmission problems when they occur.

1.2 Benchmarking WG - Kate
This has been seriously parked while I have been busy with a few other things. I'm almost ready to get back to it all.

Meanwhile:

Rachel Warren has been very busy - daily benchmarks now completed, 5 methods have been tested in various formats on 4 regions and 4 different error worlds, analysis has begun. Rachel's work on adding inhomogeneity and evaluation will now be a great help to our work on the monthly time scale.

Victor Venema is visiting the Met Office in August to work on the error worlds - this is now pretty much confirmed - well, the £££ are definitely confirmed.

1.3 Data rescue task team - Peter
We attempted a call on April 29th but ended up with a lot of last minute apologies that meant we were sub-quorate again. Notes will be posted if only to record that an effort to have the call was made.

1.4 Parallel observations science team - Victor
VV: The Parallel Observations Science Team has been formed and had its first telecon. More members are naturally welcome.
VV: We have (active) member and associate members (people who would like to be informed and have the possibility to give feedback).
MdeP: Victor, could you please add me as an associate member?
VV: Naturally, welcome.

VV: The database will be build by working on papers for specific historical transitions in the climate networks.
VV: We aim to collect and anlyse dataset from many climate regions to that we can draw first conclusions on global biases in the raw observations.

* Enric Aguilar is leading the paper on the transition to AWS for temperature.
XW: my colleagues Ewa Milewska and Lucie Vincent just finished a manuscript on this topic. They found that the observing window change between manned and automated observations of daily max and min temperatures is the most

important cause of the biases associated with automation (if they are real paralell measurements, with instruments on the same instrument compound and very near each other). The biases vary by season but not much by wind condition.
VV: That was using Canadian data, right? Could we use that when your study is finished for our global bias study?
XW: no problem for me. VV: Great, thank you.

* Petr Stepanek is leading the paper on the transition to AWS for precip.
* Theo Brandsma is leading the paper on the transition to Stevenson screens.
* Kate Willett is leading the paper on changes in the measurement of humidity.
* Jenny Linden is leading the paper on relocations and possible improvements in station siting due to this. With a focus on villages.
More ideas for papers and leads are welcome.
AM: A paper on type B (only instrumental, not statistical) uncertainty in land ground based surface air temperature measurements, detailed list of sources, budget, instruments needed to evaluate them etc... in collaboration with CCT TG EV (Consultative Committee for Thermometry Task Group for Environment) Members?
VV: That kind of work is very important to understand the measurement process. There is clearly a wide span between "statistical" and "physical", I thought we were working physically compared to statistical relative homogenization.
VV: In general we will not have the information one would need for a more physical model, albedo of instrument and ground, insolation, specific humidity profile, clouds, wind, sensor size, soil moisture, cleaning and replacing of the screen, deterioration of the electronics, heat flux between instrument and sensor, air resistance. This sounds to be more something for understanding a single instrument very well, which would require a dedicated experiment. The data of Theo Brandsma at the KNMI may be very good for that. One main aim of POST is to get a handle on biases in the global record, thus we need information on a large number of stations/networks/climates/instruments.

Suggestions for available parallel datasets are welcome. A large part of the datasets know to us is listed here (but should be updated):
https://ourproject.org/moin/projects/parallel

Next deadline: Finish basic data processing of the parallel database; Quality control, homogenization [homogeneity assessment] and computation of indices has to be coded. Published for code review
Victor Venema, Enric Aguilar, Renate Auchmann: July 2015

2. Update on REZATEC activities - Richard

Richard and Peter met with Rezatec representatives after the previous steering committee call. At that stage the plan was to have a preliminary working web-based interface to the ISTI databank, with skeleton capability, in place by the end of March. Following this meeting, Richard spent a day at Rezatec to iron out some of the technical details; it was agreed there that Rezatec would focus on

developing the infrastructure for querying the database and delivering the results online, and that Richard would write some R scripts for simple visualisation and exploratory analysis. An agreed set of analyses / capabilities in the first instance was as follows:

- Single-station analyses of either Tmin, Tmax or Tmean: plot time series of raw data or anomalies; offer choice of anomaly calculation methods; possibly show monthly boxplots to explore seasonality
- Multi-station analyses (again, either Tmax, Tmin or Tmean): bubble or colour maps of mean temperatures, maps by month, spatially organised plots of time series
- Data quality analyses: show percentage of missing observations in each station / year

Subsequently, Rezatec encountered some difficulties with the format of the data - they were advised to contact Jared to resolve these [I don't know whether this led to anything???].

From Jared by email:

- I received word from a Susmit Nayak back in March about them adding the ISTI data into an SQL database. They initially had to turn the ASCII files to CSV before being added, which caused some headaches on their end. They pointed out a couple of formatting issues (some similar to what David Jones provided some time ago). These included special non-ascii characters and extra underscores that weren't needed. However to my knowledge they were able to successfully get the stations in there after doing some manual fixes. I had told them I would incorporate fixes to these issues when I apply to the suite of Stage 2 data (which should be around the time before we initiate a new version of the merge)

Partly as a result of this, and also for other technical reasons, progress has been slower than anticipated. The current state of play is that (a) all of the ISTI data have been loaded into a single table, indexed for fast searching by users and cross-referenced to supporting tables with station names, quality flags etc. (b) a server version of RStudio has been deployed with an experimental RShiny front end that will ultimately be used to deliver online analytical capability. So far, the work has not reached a stage where it is possible to start writing visualisation scripts - we have been promised a timeline within the next few days. Despite the delay, Rezatec have provided assurance that "Rezatec's end goal remains ... to contribute to the ISTI initiative, by providing a public domain global surface temperature showcase". The leader of the Bioclim project (which funded Rezatec to do the work) regards ISTI as a flagship case study for the work, and wants an "official" launch of the Bioclim web site, with the ISTI data on it, to coincide with the Paris climate change meeting in July.

PT: Note that Xiaolan and I will be there.

3. Irish possible grant - Peter

The general proposal is to spend 5 million euros on ISTI.

1. Crowdsourced digitisation
2. Crowdsourced HOMER
3. Serious gaming using automated algorithms and the benchmarks
4. Portal
5. Assess outcomes

I'd be very busy if it is successful ...

Looking for some international partners who may help supervise students or make use of data rescued (NCEI, MO, possibly KNMI and others). Money must be spent in Ireland. But can factor in travel support for int'l partners to enable cooperation.

4. General update on activities - all

Please note events that have taken place and outcomes here

BT: The first meeting of the WMO Task Team on Homogenization took place in Geneva 20-22 April. Clearly this group has some overlap with ISTI and will be looking to work collaboratively where possible. Victor is co-chair together with Matthew Menne of this TT and I am a member. XW is also a member.
VV: We will write a guidance on monthly homogenization. An update of Aguilar et al. (2003). We will probably also have a section on daily data, but the state of the science is not that far yet that would give real recommendations.
VV: TT-HOM supports POST, a databank with all climatic elements (would also help in homogenization, especially of daily data) and a global climate reference network.
VV: A comprehensive format for metadata is being formulated by WMO. For historical known breaks a simpler compatible system would be useful to share this information internationally. We will try to formulate a "standard".
KW: How may benchmarking help?
VV: May be too late to inform this activity. At least this cycle.
BT: May be a next cycle activity.
XW: I think there is a fundamental difference between TT-HOM and ISTI, which is automated homogenization should not be the goal to recommend to WMO states to produce their national homogenized datasets. High quality homogenized climate data have many users other than climate change researchers.
VV: They are complementary activities.

BT: The inclusion of daily data in monthly CLIMAT messages is to be trialled soon (I believe in the US - Produced at NCEI Asheville. Colin Morice has offered to help test the Daily CLIMAT messages), but even if the trial goes well it sounds like it won't be rolled out globally until after the next Commission for Climatology meeting (i.e. not until after 2018).
JL: Is making progress. Producing at NCEI in Asheville monthly. UKMO will help validate.

KW and BT attended the WCRP Grand Challenge on data for extremes workshop (Sydney, February 2015). A talk was given on ISTI and issues surrounding data availability for extremes. This included noting poor support for raw data archiving (rescue, storage, maintenance, presentation) and vulnerability due to single point host institutes. Issues of inhomogeneity and the need for benchmarking were also noted.
PT: Will the outcomes be reported anywhere?
KW: Good question - I think there is meant to be some write up and there is certainly a list of plans with names against that list. (BT- there's an article in the latest edition of the AMOS Bulletin at http://issuu.com/amosadmin/docs/bamos_april_2015_web?e=6873938/12769877 - pretty sure this is accessible to non-members).
KW: Found them - just need to see if they're online somewhere. Not yet online. Relevant ones are:

- Roadmap to existing data sources/datasets/data portals – short guidance document (link to reanalyses.org) standard documentation/wiki? (BT and KW involved)
- Guidance document on how datasets should/should not be used for extremes (linked to above) BAMS article? Think about scope (AKT, BT involved)
- Best practice guidance paper for gridding daily/extremes datasets and applications e.g. model evaluation (part 1, part 2?) (KW involved)
- Develop new datasets using updated data gathering activities and intercompare all existing ETCCDI products (AKT involved)
- Liaise with data rescue community – ensure digitised data are added to database. Roadmap for data rescue activities and raw databases. Identify regions of most value for least effort e.g. South America (RA, PT involved - although they may not know it yet?)

KW: Poor state of archives was a 'surprise' to many of the participants.
VV: Also lack of awareness of inhomogeneities at daily timescale.

KW also visited the Australian Bureau of Meteorology and CSIRO (Hobart) and gave talks that included ISTI/benchmarking.

MdeP joined WMO CIMO Expert teams
A2. Expert Team on New In-situ Technologies
C1. Expert Team on Regional Instrument Centres, Calibration and Traceability
Attending *Metrology for Climate* meeting (21/22 May) on ECVs at NPL as ISTI representative PT: So am I (also gruan, gaia-clim :->)
AM joined WMO CIMO Expert team on Observation In Site Technologies (ET-OIST)
AM joined WMO CCl group of Rapporteurs on Climate Observational Issues on reference stations
PT I seem to be on hook to chair that CCl group :-s
Next MMC 2016 (Metrology for Meteorology and Climate) will be held in conjunction with WMO CIMO TECO, Madrid 26-30 September 2016.

Please note events that will take place and opportunities here

Extension of WMO Resolution 40 to cover certain types of historical climate data will be considered by WMO Congress later this month/early June. There is a separate agenda item (fairly loose at this stage) which makes reference to potential for crowdsourced digitisation and is generally supportive of open data. WMO Congress documents are available at http://cg-17.wmo.int/documents-english - the extension of Resolution 40 is at item 8.1(2) and the item on open data policies is 9.1.

C3S workshop (http://www.ecmwf.int/en/copernicus-workshop-climate-observation-requirements) - Copernicus Workshop on Climate Observation Requirements, early July, Reading, UK - anyone attending this? I (KW) have registered but not heard anything yet. It looks like it will cover rescue/archiving/homogenisation to some extent.
PT: Am double booked with two events in Paris at same time.
XW: I will attend both the ETCCDI meeting and the WCRP conference "Our Common Future under Climate Change", 6-10 July 2015, UNESCO, Paris PT: I will be at that one too :-)

Any plans from anyone to attend COP20 in Paris in December? I'm trying to put together a side event and on the look out for potential collaboration on this. Ideally it would present the story of how observations are used to build robust adaptation strategy with the key point being that there is much more that could/should be done in terms of data availability and high quality assessment. It would be nice to get across data rescue and sharing issues, archiving vulnerabilities, data quality issues and the potential for how things could be much better and show how all of this is currently pulled through to modelling and impact assessment. Any ideas/thoughts welcome.

EUMETNET Data Management Workshop (DMW) 28-30 October in St. Gallen (Switzerland) on "High quality climate data - the foundation of climate services" (www.meteoswiss.ch/dmworkshop2015). BT is likely to attend as it coincides with his next period at WMO. I imagine other ISTI people will be there too? Abstract deadline is 31 May so would need to move quickly.
VV: I will be there as well.
XW: I would send Lucie Vincent there had we known this workshop earlier. It is too late to put this in my travel plan now.
VV: I think a large part of EU homogenization community will be there. The DMW workshop is mainly about national datasets, but otherwise exactly what we do as ISTI.

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

AOB

The evolving WMO I-DARE WWW Data portal that will be 'aired' at a side event at WMO Congress coming up is at http://www.idare-portal.org/ An ISTI profile within it would be very useful. Contact Peter Siegmund at KNMI if that is desired.

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

We will aim to have an all participants call in early to mid-July