Benchmarking Working Group Call #6
Friday 3rd May 4pm BST (GMT+1), 3pm GMT,

Attending: Kate Willett (KW), Victor Venema (VV), Ian Jolliffe (IJ), Claude Williams (CW), Robert Lund (RL), Steve Easterbrook (SE), Lucie Vincent (LV), Rachel Warren (RW)

Not attending: Peter Thorne (PT)

ACTIONS FROM PREVIOUS MEETING:
i) ACTION KW: Kate to contact Robert/Ian/Victor/Lucie/Mike for help over the next week. - PARTIAL
ii) ACTION KW: send code/data to Robert and Mike and Claude (within 1 week) - DONE
iii) ACTIONKW : Circulate revised draft - all to sign off (or comment) by Monday please - DONE
iv) ACTION KW: Reply to all in response to Ian's email - NOT DONE
v) ACTION KW: Send round a doodle poll. - VERY LATE
vi) ACTION KW: Redraft time lines and circulate with Prog Report - DONE
vii) ACTION KW: Kate put date in diary to discuss WG meet up later this year. - THANKS MATT
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. ?

ACTIONS FROM THIS MEETING:
ACTION: KW to test different methods:
- Improve missing data handling – DONE – INTERPOLATE SMALL MD AND TEST MISSING DATA ALLOWANCE
- play with and without loess – VARIED LOESS FITTING
- Different GCMs – atmosphere only – NOT DONE
- compare variance and autocorrelation of difference series – DONE – SEE COMPARISON FIGURES EMAILED

ACTION VV: Lets try and get big Qs sorted by Friday next week. - STARTED

ACTION VV/IJ: VV pass on HOME documentation to Ian. IJ have a look. - DONE

AGENDA:
5 mins: Hello from Rachel/Mike (if he's around) - to tell us what you're up to if you like?
Daily temperature data from GHCND - starting with Wyoming. Simulating realistic daily temperature data - clean,
RL - Wyoming is a tricky state to look out because of topography. Little relationship with ENSO.

10 mins: Progress on Team Creation - with GCM method (Kate - see end of document and emailed figures) and without GCM method (Robert)
KW: Send docs in cross-platform formats in future
RL: need St Dev and removal of trend and clim simultaneously
KW: Compare variance in the difference series
KW: Try with a linear trend
KW: Try with and without the model
VV Cross-correlations from data, rest from model?
RL: Atmosphere only model - could use obs-model difference series.

KW: to test different methods
KW: could start with USHCN or COOP data - get from CW - get some data to Robert
VV - missing data - can you cope with it in a better way?
Can later add missing data masks later to either match reality or be worse.
ACTION: KW to test different methods:
- Improve missing data handling
- play with and without loess
- Different GCMS – atmosphere only
- compare variance and autocorrelation of difference series

20 mins: Victors suggestions for Team Corruptions error worlds which look like an excellent starting point. VV: The document I send around was not intended as a starting point for the ISTI. It may give some ideas for points we need to discuss, but many further points valuable for the ISTI are missing in the errors worlds for the idealised study, I send around.
VV: In searching for interesting research questions we should take the strengths of the ISTI benchmark over previous studies into account: global (a wide and realistic range of different network densities + various climates) and GCM based (thus many covariates available that allow for the use of physical relationships for the inhomogeneities).
VV: Suggestion, split benchmark into two realms: 1) realistic and 2) idealised

1) The realistic scenarios would be blind and are intended to estimate errors after homogenization in the real ISTI dataset.
2) The idealised scenarios are open (solution also immediately published) and are intended to improve our understanding of the algorithms.

1). Realistic scenarios
A. One realistic scenario could mimick HOME, but then make it global. Having the same statistical properties in all stations world-wide makes understanding the results easier. But we do should add some inhomogeneities that cause trend biases, that was missing in HOME.
Ai: abrupt shifts only
Aii: gradual shifts only
Aii: both types of shifts
B. Another realistic scenario could be that the mean size of the biased inhomogeneities in Tx depent on insolation and in Tn depend on column integrated water vapout. That would be helpful to study the influence of the historical improvements in radiation protection. Stefan Brönnimann has offered help with such a scenario. This would also nicely use the additional benefit of using a GCM to generate the benchmark.
C. Something with realistic spatial dependence in which you can see the boundaries of the countries in the biases (for a certain period).
D
E
To 2. Idealised scenarios
A. Many small breaks.
B. No breaks. Study false alarm rate.
C. Shall we also have one HOME-type scenario in the open for testing the algorithms?
D. Strong spatial dependence. One biased break at the same date per country. Too extreme?
E. A lot of missing data.
ACTION VV: Lets try and get big Qs sorted by Friday next week.

10 mins: VV: Also discuss general plans for team validation?
VV Climatologist validation - what are the errors/uncertainties? trend, decadal variability
VV Algorithm validation - how good are the methods/weaknesses? break detection

VV: Can IJ have a look through the validation done for HOME to see if any of it is useful?
ACTION VV/IJ: VV pass on HOME documentation to Ian. IJ have a look.

15 mins: Plan of action for potential workshop in June - what can we achieve in 3 days and how?
3rd - 6th June -

AOB

Minutes sign off:

Team Creation Stats:
Example gridbox in the USA containing 6 stations: -69.375W, 41.25N New England

REMOVED station 6 too short 000000074492

STATION ANOMALIES autocorrelation:
1 = 0.2637 2 = 0.2284 3 = 0.2659 4 = 0.1743 5 = 0.2443

STATION RESIDUAL STANDARDISED ANOMALIES autocorrelation:
1 = 0.2182 2 = 0.1883 3 = 0.2253 4 = 0.1281 5 = 0.1696

NEW STATION RESIDUAL STANDARDISED ANOMALIES autocorrelation:
1 = 0.2290 2 = 0.1576 3 = 0.2272 4 = 0.1591 5 = 0.1884

NEW STATION ANOMALIES (GCM trend added) autocorrelation:
1 = 0.3830 2 = 0.3134 3 = 0.3784 4 = 0.3105 5 = 0.3389
A little too high - but for other gridboxes it is a little too low.

Station correlations for RSAs old/new:

STATION	1	2	3	4	5
1		0.95/0.96	0.94/0.96	0.89/0.89	0.89/0.90
2			0.92/0.93	0.93/0.92	0.88/0.88
3				0.88/0.91	0.90/0.93
4					0.91/0.93
5					

Station correlations for Anomalies old/new:

STATION	1	2	3	4	5
1		0.95/0.97	0.94/0.96	0.89/0.91	0.89/0.91
2			0.91/0.94	0.92/0.93	0.89/0.90
3				0.87/0.92	0.90/0.94
4					0.91/0.94
5					

See emailed figures:
1) MCDW_-69.375_41.25_stationseries_loess04_MAY2013.eps - shows the original stations (1987-2011) and their standardised anomalies with a smoothed filter (grey) that is later removed. Note the poor temporal coverage in Station 6 - this station cannot be simulated. This smoothed filter is removed to reduce the influence of any inhomogeneity within the data and also ensure its stationarity.

2) MCDW_-69.375_41.25_smoothanomresids_loess04_MAY2013.eps - shows the Residual standardised anomalies (RSAs) and their distributions. These should be stationary and deseasonalised but still have some autocorrelation and spatial covariance.

3) MCDW_-69.375_41.25_NEWsmoothanomresids_loess04_MAY2013.eps - shows the simulated RSAs using an AR(1) model in matrix form so that it retains the spatial covariance structure between stations.
Note that I have simulated a much longer record than was available in the real data (1970-2011) - missing data in various degrees of horribleness can be added by Team Corruption.

4) MCDW_-69.375_41.25_modelseries_loess04_MAY2013.eps - shows the monthly mean time series from the HadCM3 GCM for the relevant gridbox for the 1970-2011 period. An A1B scenario is used which contains realistic historical forcing of greenhouse gases until 1999 and then a continued high emissions (increasing) scenario from then onwards. A smoothed filter has been fitted to the standardised anomalies (grey). We use this as the low frequency component of our simulated series.

5) MCDW_-69.375_41.25_NEWA1Bstationseries_loess04_MAY2013.eps - shows the simulated absolute and standardised anomaly time series with the GCM smoothed filter added for the entire length of record 1970-2011.

Issues:
- Are these simulated data realistic enough - how do we test that?
- Do these simulated data contain artefacts that may be equivalent to inhomogeneities? e.g., around 1976
- Gridboxes are treated discretely - can this method/R cope with simulation of a few hundred stations simultaneously? Where do we draw boundaries over contiguous land masses and how do we deal with the poor spatial covariance across these boundaries (assuming its probably not possible to simulate 40000 stations simultaneously?)