An overview of benchmarking data homogenisation procedures for the ISTI
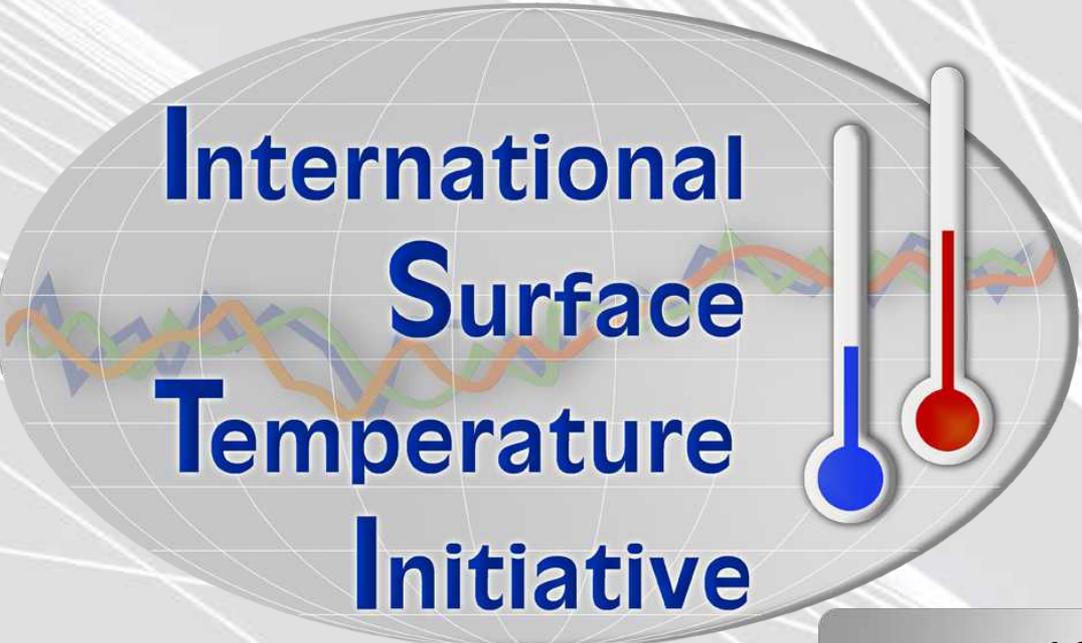
International Surface Temperature Initiative

12th IMSC, June 2013

**Dr Kate Willett
(Met Office Hadley Centre, UK)
and the ISTI benchmark working group**

Partnerships are Essential

- Bring together existing efforts, augment and ensure pull through. e.g. ACRE project (http://www.met-acre.org/), IEDRO (www.iedro.org) and other national / international programs.
- Pursue innovative approaches (crowdsourcing building upon success of oldweather.org, www.data-rescue-at-home.org etc.)
- Build on ICOADS model for sea surface temperatures (http://icoads.noaa.gov/) – easy submission and access to data
- Recognize key partners and contributions
- And you can also help …

www.surfacetemperatures.org/databank

Data.submission@surfacetemperatures.org

http://www.surfacetemperatures.org/databank/ DataSubmission-Stage1-Guidance.pdf?attredirects=0

http://www.youtube.com/watch?feature= player_embedded&v=CEaTQjzc0zo

International Surface Temperature Initiative
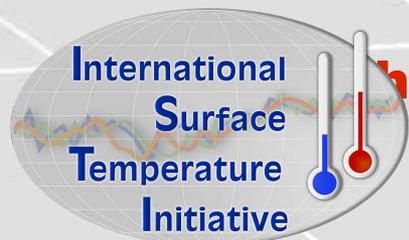
The Basics of Benchmarking for the ISTI

The Plan

HOMOGENEOUS WORLDS
Create a range of homogeneous (clean) worlds of station data analog to the real ISTI databank

INHOMOGENEOUS WORLDS
Design blind inhomogeneous (error) models and a few open inhomogeneous models exploring a range of characteristics of the inhomogeneity and climate that we believe to be problematic

Apply the blind inhomogeneous models to 2-3 different homogeneous worlds and apply the open inhomogeneous models to different homogeneous worlds

ASSESSMENT
Specify an explicit subset of stations to be returned or regional averages for a range of core climate statistics (mean, trend etc.) and a list of found break locations and sizes.

Make assessment of how well each product performed in terms of retrieving regional core statistics and also in terms of hit rates/false alarm rates for correct location and characterisation of breaks.
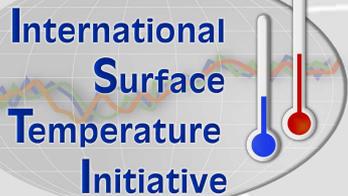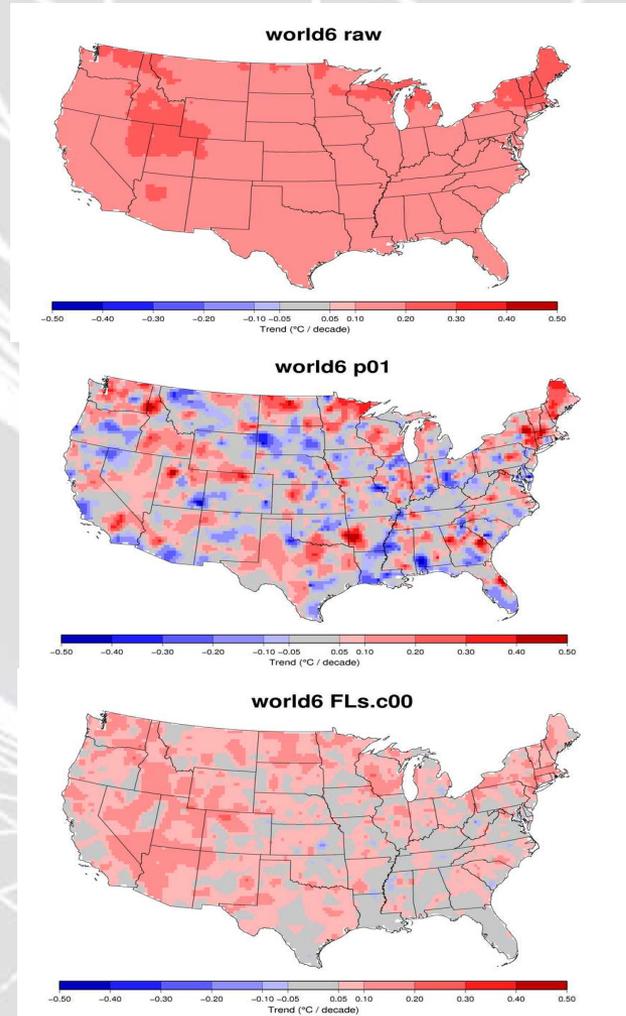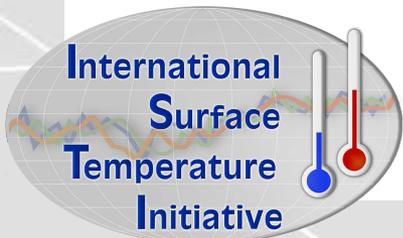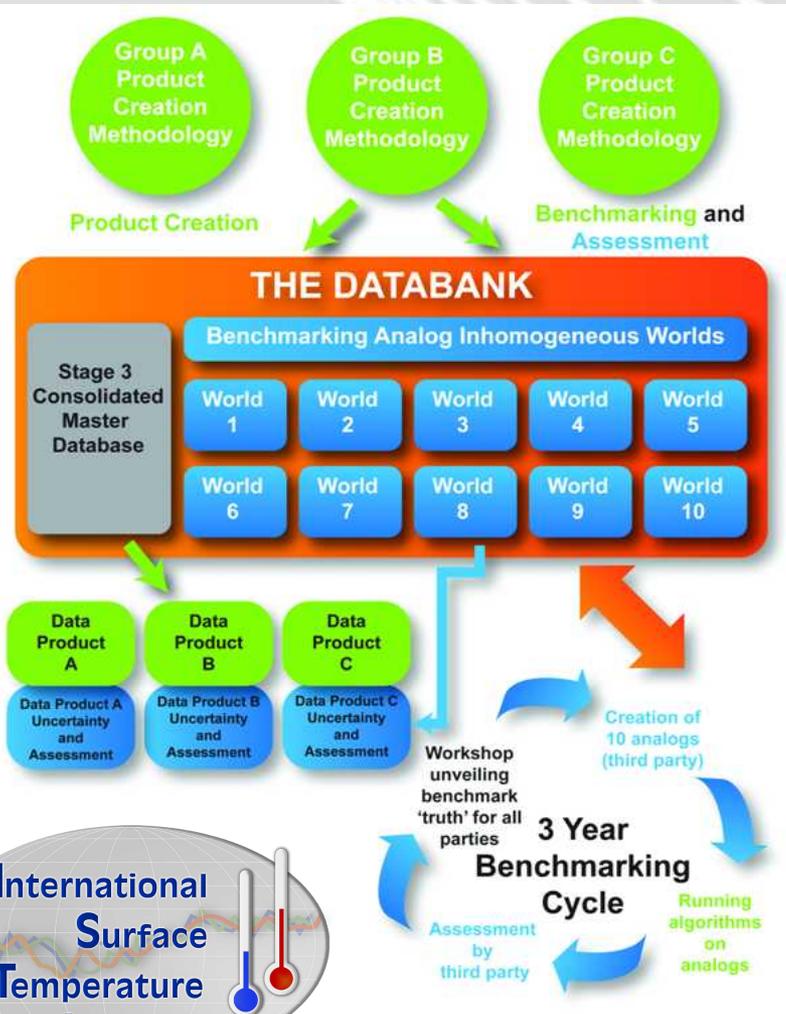
International
Surface
Temperature
Initiative

Provide a concise assessment report for each product

Review the first benchmark cycle and being cycle 2...

Benchmarking Cycle


Example use of benchmark data for USHCN

Team Creation: Creating a 'Clean' Synthetic World

So far...

A VAR(1) model that works on the gridbox scale providing globally realistic low-frequency variability

- remove low frequency filter from standardised anomalies
- simulate the station residuals
- add back in a low frequency filter from a GCM gridbox
- add back real station climatology and variance (apply missing data?)
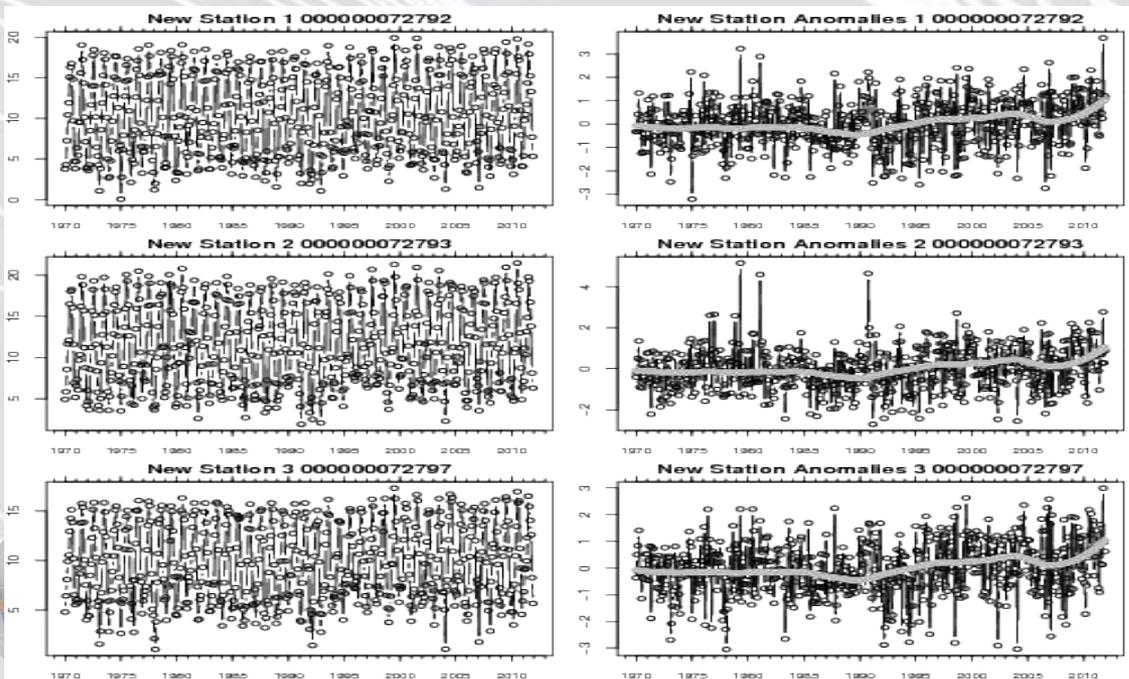
Station autocorrelation
High Freq. only

	OLD	NEW
st1	0.09	0.10
st2	0.15	0.18
St3	0.13	0.10

With low freq. Added

	OLD	NEW
st1	0.17	0.12
st2	0.24	0.19
St3	0.23	0.15

Station cross-correlations

	OLD	NEW
st1:st2	0.88	0.87
st1:st3	0.88	0.87
st2:st3	0.82	0.80

Initiative

Big questions...

Is our method going to work over data sparse regions and data dense regions?

How good is good enough?

How do we ensure smooth cross-correlations globally across matrix boundaries?
    Gibb sampling?

How do we compute this efficiently?
    R package invoking fortran – speedy but complex and less transparent/tweakable
    R code with parallelisation and bigmatrix – transparent/tweakable but complex, computer intensive
    R code without parallelisation and bigmatrix – transparent/tweakable but very computer intensive

Who is going to set up the webpage framework/databank storage framework?

            Who is going to be the handle-turner?
                Probably Kate?

International
Surface
Temperature
Initiative
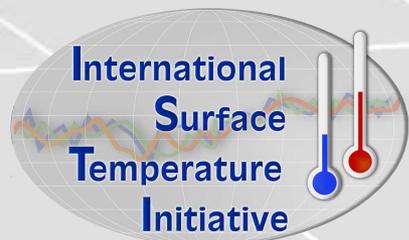
Team Corruption: Creating a set of Dirty/Error-filled Worlds

So far...

BLIND WORLDS

10 – exploring biased records, seasonal cycles in inhomogeneities, data density, gradual trend characteristics, different frequencies and magnitudes of breaks, different background climate variability
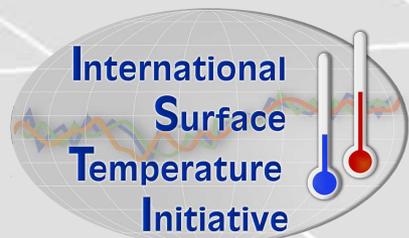
OPEN WORLDS

Up to 5

Best guess, squeaky clean, isolated issues exploration (unrealistic)

STATISTICS:

Set up basic global/regional average values for frequency, size, character, spatially correlated, seasonal cycle

Distribution may be a mixed model (e.g. hybrid pareto?) - specify mean and standard deviation or other?

International Surface Temperature Initiative
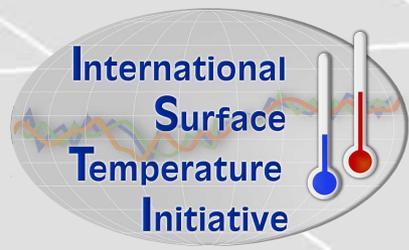
Big questions...

How blind is blind enough?

How many different background 'clean worlds'?

How on earth do we build these things?

How many different things to assess in one world?

What do we ask people to homogenise as a minimum?

International Surface Temperature Initiative

Team Validation: Assessing Skill Against the Benchmarks

So far...

LEVEL 0
How real are the benchmarks? Compare performance against benchmarks and real data to inform assessment and future benchmarks.

LEVEL 1
How well are climate features restored? Look at the difference in the mean, variance, trends and time series for regional averages in pre-specified regions.

Users should return station data, at least for specified regions. We can deal with grids but only partial assessment can be performed.

LEVELS 2 (break location) & 3 (break character)
How well are individual stations restored and what are the specific weaknesses in algorithms?

Users should return a list of break locations, size, duration, seasonal cycle/slope. These will be assessed using contingency tables and cumulative pdf plots. These can be stratified by 'perfection' (e.g., precise location, located +/- 3months, located +/- 6months) or break size or break character.

Surface
Temperature
Initiative

Assessment in form of short (1 page) summary, lengthy paper and hoping to spur other research projects using the data.

Big questions...

How do we assess different style products: grids vs stations?

What if stations are missed out?
 - maybe because they are not of interest
 - maybe because they are 'too hard' to homogenise

What do we want back from users?

Coordinate with VALUE downscaling group in terms of assessment terminology, tests and software?

International Surface Temperature Initiative
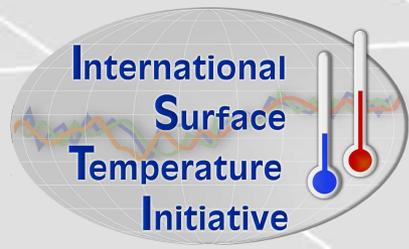
# Other Stuff

Other Progress

Concepts paper 3/4 written

Several talks/posters: SAMSI, EMS, EGU, IMSC, ACRE, MARCDAT, EarthTemp Network, WCRP Open Conference …

Agreement to take part: GHCNM, NIST?

Daily benchmark PhD: Rachel Warren, University of Exeter with Met Office sponsorship and co-supervision (Kate Willett)

Offer of data or product storage from Doug Nychka (NCAR)

Interest in helping with statistical methods in the future: Jaxk Reeves (Georgia Tech, USA), Finn Lindgren (Bath Uni., UK)

International Surface Temperature Initiative
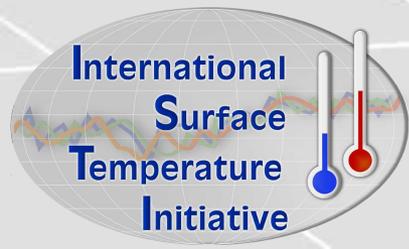
What can we achieve this week?

TEAM CREATION
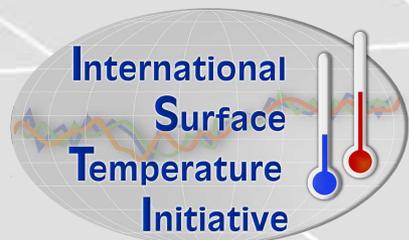Agreement on basic VAR method

TEAM CORRUPTION
Some idea of how to build the error models

TEAM ASSESSMENT
What to ask users to submit? Make it easy as possible

OVERALL
Agreement of plan from here onwards

International Surface Temperature Initiative

Questions and Answers

www.surfacetemperatures.org

General.enquiries@surfacetemperatures.org

Data.submission@surfacetemperatures.org