

Creating surface temperature datasets to meet 21st Century challenges

Met Office Hadley Centre, Exeter, UK

Confirmed sponsors: WMO (CLPA and OBS); WCRP; GCOS; RMS; Met Office; University of Exeter 7th-9th September 2010

Agenda version 2 26/07/10

International scientific organizing committee	Local organizing committee
Peter Thorne (Chair, CICS-NC (formerly Met Office)) Peter Stott (Met Office) Simon Gilbert (Met Office) Albert Klein Tank (CCI / KNMI) Liz Kent (WCRP WOAP / NOCS) Blair Trewin (BoM) Lianchun Song (CCI / CMA) Ian Jolliffe (Univ. Of Exeter) John Christy (UAH) Nigel Fox (NPL) Jay Lawrimore (NOAA NCDC) Adrian Simmons (GCOS / ECMWF)	Kate Willett (Chair) Peter Stott Simon Gilbert Julia Slingo John Kennedy Jo Mclellan Chris Gordon Vicky Pope

Workshop aims

- To agree a well defined plan of how as an international community to go about undertaking necessary data rescue, analysis and verification to produce global surface temperature records at monthly, daily and sub-daily resolution that are fit for climate services. The focus is limited to land surface temperature records so as to allow a well defined remit but with the understanding that the concept should be taken forwards for other meteorological parameters if successful. It is also recognised that close collaboration with the SST community is essential to create truly global high quality assessments of surface temperatures.
- To agree how to ascertain the strengths and weaknesses in the resulting estimates of historical changes and quantify the inevitable uncertainties that result from not having made historical observations primarily for climate monitoring and these measures not being fundamentally traceable to standards.
- To engender broad input into the process design from expert communities outside of the traditional climate change community and key stakeholders including likely end-users.
- To ensure that the envisaged outputs will be usable by a broad range of stakeholders and accessible to everyone.

The workshop will start out from the Met Office proposal to the Commission for Climatology (reproduced at www.surface temperatures.org) to produce an agreed plan with broad community sign on. The final meeting report which will be led by the international organising committee will appear either in BAMS or EOS as a public accessible summary and in more complete detail as a WMO Technical Document.

For those sessions that have an associated breakout scheduled selected individuals / groups will be solicited by the international organising committee to provide white papers (preferably 2, maximum 3 sides of text plus any necessary figures) to facilitate discussions and provide a starting point. Participants will be able to modify these positions in both breakout groups and plenary. These white papers will be posted in advance of the meeting and comments on these will be solicited through a moderated blog. *The white papers will provide the main input for the final meeting report.*

Primer and nomenclature

The final workplan will be developed and agreed by the meeting participants but here we outline some of the essential ingredients of that workplan. A single **primary databank** will be created to hold observations as originally recorded. The **databank** would contain observations at all available temporal resolutions: monthly, daily and sub-daily. The **databank** will have free and open access, contain all available metadata describing the observations and will have strong version control in place. The identification of **fundamental station records** or other primary observation sources is in some cases not a trivial task and research, data archaeology, digitisation, rescue, all with an agreed audit trail will be necessary.

This **databank** will be used to produce **derived datasets** by a broad range of traditional and non-traditional participants to meet particular requirements, but all derived **datasets** should be traceable to the **primary databank**. Such **datasets** might be for regional applications, incorporate data adjustments or homogenisation or be presented on regular grids. The derived **datasets** must be documented to agreed standards and therefore essentially reproduceable whilst recognising that exact reproducibility may not be possible for some approaches. Mechanisms should be in place to allow researchers to make the results of enhanced analysis such as improved quality assurance, data adjustments and homogenisation available to the wider community, perhaps as an adjunct to the **primary databank**.

Methods to ensure the accuracy of **derived datasets**, including the effectiveness of data homogenisation and the construction of gridded products, should be in place. For example, **dataset** development methodologies can be benchmarked for performance against a common set of realistic test-cases where synthetically produced data can provide a true solution against which results can be assessed.

Another important aspect is to ensure openness, transparency and outreach. To be successful the process must engage widely beyond the community of research scientists, including with funding bodies, the general public, policy makers and international organisations. Developing nations must be a priority for engagement, ensuring both that datasets are as globally complete as possible and that all nations contributing **station records** can benefit from the entire climate record. Finally, how this effort is to be governed and how it interacts with pre-existing efforts and related issues needs to be defined.

Timing will be strictly adhered to – there will be a time keeper who will police talks.

Tuesday 7th September

Chair: Prof. Julia Slingo OBE, President RMS, Met Office Chief Scientist

Registration 08.30-09.00

1. Welcome 09.00-10.00

09.00-09.10 local logistics – Kate Willett, Chair, Local Organising Committee

09.10-09.40 Introductory remarks – Julia Slingo, Met Office Chief Scientist

09.40-10.00 Overview of Workshop Objectives – Peter Thorne, Chair, International Organising Committee

2. Overview talks on data recovery, digitisation, management and data policy

Overview talks (this session and other overview sessions- 7 and 12) are to be a maximum of 25 minutes with 5 minutes for questions. The idea is that these talks are scene setters so it is imperative that they are focussed. Speakers are specifically requested to:

- 1. focus upon the lessons learned that may be applicable to the task in hand*
- 2. avoid generalist overviews of their project*
- 3. reflect upon what worked, what didn't and what they would do differently if they were to start out again.*

Please provide talks at least a week in advance of the workshop so that they can be reviewed by the IOC.

- **ICOADS: A multinational data rescue, digitisation, archiving, and access success for the oceans– Steve Worley**

10:30-11.00 Coffee break

- **NCDC: A World Data Centre perspective on the data management issue – Jay Lawrimore**
- **CCI/CLIVAR/JCOMM regional initiatives on high resolution data and extremes indices -Albert Klein Tank**

12.00-13.00 White paper presentations on data rescue efforts, updates and policy

3. Retrieval of historical data

White paper lead: Peter Thorne

White paper authors are requested to consider in preparation of their document inter-alia (a finite list): what other known data sources exist that are not part of the current databases; what other efforts are currently under way (avoidance of duplication); other potential sources; a practical model for the data rescue effort (e.g. whether one or more workshops are required, the best mechanism to solicit data release); data version reconciliation between data banks; digitised records that are not made available internationally; and the potential of crowd sourcing digitisation.

4. Near real-time updates

White paper lead: Jay Lawrimore

White paper authors are requested to outline existing real-time exchange policies and discuss: potential for expansion; how existing and planned data exchange networks can be utilised; discrepancy between theory and practice (how to decrease the number of missing data); update timeliness; and real-time data checking.

5. Data policy

White paper lead: Albert Klein Tank

White paper authors are requested to consider IPR issues pertaining to: creation of a single databank holding; possibilities and obstacles for collation into a distributed databank; development and dissemination of datasets from the databank; WMO RES 40 and historical data; whether because we want to be open and transparent do we limit to only data without restriction? If so, what impact does this have?; and whether there are novel approaches to overcome the data IPR issues that could be pursued.

6. Data provenance, version control, configuration management

White paper lead: John Christy

White paper authors are requested to consider: The metadata requirements to ensure tracability back to the original data record; the version controlling requirements (what represents BAU and what constitutes a fundamental update to the databank); how to best retain previous versions of the databank; how to configure the archives to maximise usefulness (data format, file indexing, appending the homogenised data estimates etc.); how to version control and archive the datasets produced from the initial databank (see Section 8).

13.00-14.00 Lunch

14.00-16.00 Breakout groups on data rescue, updates and policy

Institutions with multiple participants are politely requested to ensure that they are represented in as many breakout groups as possible in these daily breakout sessions. Organisers may assign individuals to breakout groups prior to meeting commencement based upon participant expertise (TBD).

16.00-16.30 Coffee break

16.30-18.00 Plenary discussion on output of breakout groups

18.30-20.30 Icebreaker Exeter University

Wednesday 8th September

Chair: Prof. John Mitchell OBE FRS, Met Office Principal Fellow

08.30-10.30

7. Background talks on homogenisation and analysis

- Lessons learnt from USHCN and GHCN most recent homogenisation cycle – Matt Menne
- The COST HOME project – homogenisation across timescales and with input from multiple groups – Olivier Mestre

- Hadley Centre radiosonde project: what can be learnt from automation and systematic validation against plausible test-beds? – Peter Thorne
- SST homogenisation and analysis – what lessons can be learnt for land homogenisation? – John Kennedy

10.30-11.00 Coffee

11.00-12.00 White paper presentations on dataset creation and performance benchmarking

8. Creation of quality controlled homogenised datasets from the databank

White paper lead: Blair Trewin

White paper authors are requested to consider: whether different methods will be required for different timescales (monthly, daily, sub-daily); whether some common metrics would be appropriate; whether efforts should be focussed on a particular area or timescale first and if so why; the role of national and regional assessments; how to engender multiple efforts; the potential for novel approaches (e.g. an open source community effort, the use of reanalyses output instead of neighbours as the expectation field and reanalyses feedback files, Bayesian approaches); the importance of a consistent approach to assessing uncertainty.

9. Benchmarking homogenisation algorithm performance against test cases

White paper lead: Kate Willett

White paper authors are requested to consider: how to create realistic synthetic test cases; whether there is a need for full space and time sampling to the observational network characteristics; from what raw data (reanalyses, climate models, others?) these test cases should be constructed; how to design a set of test cases that are orthogonal and can optimally answer multiple questions (e.g. is performance predicated upon there being an underlying trend? Does the break clustering matter? Does breaks having a sign bias impact performance? What about trend like effects?); whether the test case construction should be truly blind to the dataset algorithm creators; how to avoid potential for over-tuning of algorithms based upon these test cases.

10. Dataset algorithm performance assessment based upon all efforts

White paper lead: Peter Stott

White paper authors are requested to consider: How can we objectively discriminate between different datasets that result from this exercise; how to make use of the benchmarking results to quantify the fundamental uncertainties across a range of space and time scales; whether this assessment needs to be independent of the dataset creators and if so how to achieve such an independence; whether such an assessment should look to produce a unified best estimate of the changes and if so how; whether it will be feasible to use the test cases to ascertain fundamental performance strengths and weaknesses of competing algorithms; how to handle / incorporate any efforts, such as reanalyses, that can't be run on the test cases.

11. Spatial and temporal interpolation

White paper lead: Liz Kent

Recognising that many customers will require spatially and temporally complete fields the white paper authors are requested to consider: methodological approaches to

infilling; caveats that need to be applied to interpolated data; whether interpolation techniques will need to be different for different timescales (monthly vs. daily vs. hourly), development of uncertainty and error covariance estimates, and the role of reanalyses.

12.00-16.00

Breakout groups on dataset creation and assessment

[Lunch 13.00-14.00]

16.00-16.30 Coffee

16.30-18.00 Plenary

19.30 Conference banquet (TBC)

Thursday 9th September

Chair: Thomas Peterson, President Commission of Climatology

08.30-09.30

12. Data management, outreach, education overview talks

- The satellite experience: version control, configuration management and serving multiple independently derived datasets – John Christy
- Outreach, visualisation, user engagement, and education: lessons from google.org – Amy Luers

09.30-10.45 White papers on communication, outreach and governance

13. Publication, collation of results, presentation of audit trails

White paper lead: Lianchun Song

White paper authors are requested to consider: what an acceptable audit trail for dataset acceptance would be (intermediate steps, code, others) and what would be desirable beyond this minimum requirement; whether datasets should be served from a common platform in a common format; how to enable intercomparison of the results between the datasets that will result; what will constitute acceptable publication criteria (peer-reviewed literature; other?); how to ensure that datasets are correctly credited to their creators and related Intellectual Property Rights issues on these value added products; how to handle approaches that may be substantially different e.g. reanalyses.

14. Solicitation of input from the community at large including non-climate fields and discussion of web presence.

White paper lead: Nigel Fox

White paper authors are requested to consider: how to engage non-traditional participants and input from relevant fields outside climate; the role of the internet (the web presence and whether and if so how this should be interactive); whether to have a central dataset repository or distributed dataset serving from participants host sites; the potential for user visualisation of the databank and resulting datasets and manipulation of the data (maps, graphs etc.); how to engage truly global participation and the need for technical help / support and off the shelf software; how to provide tools for end users to manipulate the data to meet their climate service needs.

15. Governance

White paper lead: Simon Gilbert

White paper authors are requested to consider: how this process should be governed; the need for and frequency of meetings and their purpose; funding support; and periodic reporting requirements and assessments as to whether the project is being successful and should be continued; and how to ensure regular updates and truly global participation in the process through e.g. regional workshops.

16. Interactions with other activities

White paper lead: Peter Thorne

White paper authors are requested to consider: how the concepts discussed for this challenge may most optimally be expanded to other climatic variables / observing technologies; liaison and collaboration with the SST community to ensure our ability to create a suite of consistent truly global products; how to avoid redundancy with ongoing efforts; how this effort can optimally interface with ongoing and planned reanalyses efforts (data for input and also useful data back to aid dataset development),

10.45-11.15 Coffee

11.15-13.00 Breakout groups on publication, outreach, education and project governance

13.00-14.00 Lunch

14.00-15.15 Plenary discussion of issues regarding management, outreach, education and project governance

15.15-15.45 Coffee

15.45-17.30

16. Wrap up and agreement of broad brush challenge ToR