

Benchmarking and Assessment Working Group

2011 Progress Report

October 2011

Current Members:

Kate Willett (Chair) - UKMO Hadley Centre, UK
Claude Williams - NCDC, USA
Ian Jolliffe - Exeter Climate Systems, University of Exeter, UK
Robert Lund - Department of Mathematical Sciences, Clemson University, USA
Lisa Alexander - Climate Change Research Centre, University of New South Wales, Australia
Olivier Mestre - Meteo France, France
Stefan Brönniman - University of Bern, Switzerland
Lucie A. Vincent - Climate Research Division, Environment Canada, Canada
Aiguo Dai - Climate and Global Dynamics Division, NCAR, USA
Steve Easterbrook - Department of Computer Science, University of Toronto, Canada
Victor Venema - Meteorologisches Institut, University of Bonn, Germany
David Berry - National Oceanography Centre, Southampton, UK

Peter Thorne (ex-officio) - CICS-NCDC, USA

Ex-Members:

Chris Wikle - Department of Statistics, University of Missouri, USA
Chris had too many other commitments and had to step down.

New Members:

David Berry - National Oceanography Centre, Southampton, UK

October 2010 to October 2011 Objectives:

- 1) Invite members and set up the group including email lists, website and blogsite.
- 2) Devise a structure for creation of the Benchmarking cycle and set out a timeline for achievements and submit to the Implementation Plan
- 3) Construct a Benchmarking and Assessment working group Terms of Reference
- 4) Publicise the aims and objectives of both the ISTI and the work of the Benchmarking and Assessment working group widely and engage with as many similar efforts as possible
- 5) Design the concepts behind the benchmarks and begin to construct them

Objectives Met:

1) Invite members and set up the group including email lists, website and blogsite.
A selection of international scientists with a variety of skills (climate, mathematics, spatial statistics, verification statistics, homogenisation, benchmark creation, computer science) were contacted and a group formulated. This includes members of the COST HOME action which has been a pioneer in benchmarking homogenisation algorithms and forms a valuable scientific basis for the work of ISTI. Due to Chris

Wikle's departure from the group we are lacking a spatial statistician but others have experience in this area (Olivier Mestre, Victor Venema, Robert Lund) and can provide expertise. The email list benchmarking@surfacetemperatures.org has been set up to communicate with all members. Teleconferences are hosted every 2-3 months.

The ISTI website now has a Benchmarking and Assessment Working Group page (www.surfacetemperatures.org/benchmarking-and-assessment-working-group) outlining who we are and what we're aiming to do. It also hosts the minutes of all teleconferences, important documents, listings of conferences attended and a timeline of deliverables.

We have a blog site (<http://surftempbenchmarking.blogspot.com>) that is open to all members to post threads and anyone to comment on threads. All members are invited to post threads at any time.

2) Devise a structure for creation of the Benchmarking cycle and set out a timeline for achievements and submit to the Implementation Plan

Four teleconferences have been held with at least 7 members in attendance for each. We have discussed: aims and running logistics; creation of the analog-known-worlds; overall concepts and homogenisation questions to address; and a plan of action. The concepts behind benchmarking on such a large scale are complex and have taken a long time to formulate. A first draft of the concepts paper has been drafted outlining initial ideas behind analog-known-world creation, error structure implementation to create the analog-error-worlds and assessment methods. These may change as the appropriateness of these concepts becomes clear.

Three task teams have been established to govern the three components of the benchmarks. Team Creation is lead by Robert Lund and will coordinate the design and creation of the homogeneous synthetic analog-known-worlds. Team Corruption is lead by Claude Williams and will coordinate the design of error structures and software to add these to the analog-known-worlds creating the analog-error-worlds. Team Validation is lead by Ian Jolliffe and will coordinate the design and creation of suitable assessment tools. All members have joined one or more groups. Kate has promised to be a very active member of all groups recognising that unlike other members she has some (10%) of her official work time allocated to this and ISTI related work. Email lists have been set up for each of these teams: teamcreation@surfacetemperatures.org; teamcorruption@surfacetemperatures.org; and teamvalidation@surfacetemperatures.org.

A timeline for deliverables has been laid out in the Implementation Plan and is hosted on the BAWG website.

3) Construct a Benchmarking and Assessment working group Terms of Reference

A BAWG Terms of Reference has been written and agreed on by all members. It is now hosted on the BAWG website. The working group will report to the Steering Committee, giving a verbal progress report at every quarterly phone call and a written annual progress report.

4) *Publicise the aims and objectives of both the ISTI and the work of the Benchmarking and Assessment working group widely and engage with as many similar efforts as possible*

The work of the Benchmarking and Assessment working group has been publicised at a number of conferences – all presentations are hosted on the website:

February 2011

Kate Willett's informal presentation at the National Climate Data Center (NC, USA): *Devising a Benchmarking System for Homogenisation Methods of Climate Data-Products*

April 2011

Kate Willett's Poster for EGU 2011: *Robust Benchmarking of Homogenisation Algorithms for the Surface Temperature Initiative*

May 2011

Kate Willett's presentation for MARCDATIII, Frascati, Italy, 2011: *Is it Good Enough? Benchmarking Homogenisation Algorithms and Cross-cutting with Efforts for Land Observations*

October 2011

Steve Easterbrook's poster presentation at the WCRP Open Science Conference, Denver, CO, USA: *Benchmarking and Assessment of Homogenisation Algorithms for the International Surface Temperature Initiative (ISTI)*

Kate Willett's presentation at the COST HOME 7th Seminar for Homogenisation and Quality Control of Climate Databases, Budapest, Hungary: *Creating a Global Benchmark Cycle for the International Surface Temperature Initiative*

Further mentions have been made within ISTI general presentations.

5) *Design the concepts behind the benchmarks and begin to construct them*

A first draft of the concepts paper has been drafted outlining initial ideas behind analog-known-world creation, error structure implementation to create the analog-error-worlds and assessment methods. These may change as the appropriateness of these concepts becomes clear. Three task teams have been established to coordinate these efforts as described above. We recognise that all three are strongly linked and dependent on one another but that there is value in expertise being channelled to specific areas.

The basic concepts are in place. We now need to fine tune our decisions of exactly what to include, how, and to set up software to allow multiple productions of benchmarks with easily tweaked parameters.

Objectives Not Met:

None

Other Efforts and Achievements:

Provisional acceptance of a PhD proposal to work on geospatial statistical methods in building daily benchmarks with Kate Willett (UK Met Office) and Prof. Trevor Bailey and Prof. Ian Jolliffe from Exeter University and with CASE NERC quota funding from the UK Met Office.

NSF Research proposal incorporating benchmarking submitted through Clemson University, University of North Carolina and NCDC but rejected.

Potential for collaboration on a European daily benchmarking .

Standard talk and poster now prepared and presented – this will be updated as appropriate and made available on the website.

2011 Annual Overview:

2011 has seen the beginnings of the Benchmarking and Assessment Working Group. This has involved bringing together expert individuals and setting up a distributed working environment (teleconferences, website, email lists and blogsite). This is a new concept and so time has been taken to explore the issues, develop ideas and ensure that all members are in agreement of the objectives. It has been difficult to make quick progress in a distributed, sporadic working framework that crosses multiple timezones and that is largely conducted in people's own time. It is recognised that at any one time some members cannot contribute as much as others due to other commitments but the input of all members is highly valued and very much appreciated. However, despite these obstacles, there has been progress made in developing the framework of benchmarking. It should still be feasible to begin the benchmark cycle in November 2012.

Objectives for October 2011 to October 2012:

- 1) Benchmark Cycle concepts and plan formalised and submitted to JAOT or similar by April 2012
- 2) Design methods and create software for producing the **analog-known-worlds** ready for November release of pilot benchmarks and creation of the official benchmarks for the Benchmark cycle
- 3) Design methods and create software for producing the **analog-error-worlds** ready for November release of pilot benchmarks and creation of the official benchmarks for the Benchmark cycle
- 4) Design methods and create software for assessing the results of tests on the benchmarks ready for the Benchmark cycle
- 5) Create a platform for guiding users of the benchmarks in how to use them and how the assessment works
- 7) Publicise the aims and objectives of both the ISTI and the work of the Benchmarking and Assessment working group widely and engage with as many similar efforts as possible