

Benchmarking and Assessment Working Group

2012 Progress Report

January 2013

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 Brönnimann - 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.
Ingvild Antonsen - Justervesenet - The Norwegian Metrology Service, Norway
Change of job

New Members:

Mike Finney - Department of Mathematical Sciences, Clemson University, USA
Rachel Warren - College of Engineering, Mathematics and Physical Sciences, University of Exeter, UK
Giuseppina Lopardo - Istituto Nazionale di Ricerca Metrologica (INRiM), Italy

October 2011 to October 2012 Objectives:

- 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
- 6) 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

Objectives Met:

- 6) 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

December 2011

Ian Jolliffe's poster at the 5th International Verification Methods Workshop, Melbourne, Australia: Benchmarking and Assessment (Verification) of Homogenisation Algorithms for the International Surface Temperature Initiative (ISTI)

March 2012

Peter Thorne International temperature symposium, USA – presentation

May 2012

Kate Willett's NCDC Visit with Claude Williams and Robert Lund - presentation

Kate Willett's Clemson University/Robert Lund visit – presentation

Lucie Vincent's CMOS presentation

June 2012

Peter Thorne's presentation at the Earth Temperature Network workshop, Edinburgh, UK

September 2012

Victor Venema at a side meeting on homogenization at the EMS in Lodz, Poland

November 2012

Kate Willett's presentation at the 5th ACRE Meeting, Toulouse, France

Kate Willett's visit to the National Physical Laboratory (Michael Podesta, Stephanie Bell, Alistair Forbes, Tom Gardiner)

Objectives Not Met:

1) *Benchmark Cycle concepts and plan formalised and submitted to JAOT or similar by April 2012*

This is in first draft form but not yet ready for submission – aim for Summer 2013

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*

Various methods have been tested by Kate Willett with little success – getting the spatial covariance realistic (correlated but not too correlated) is proving difficult. Robert Lund has come up with a method that appears to work in the majority, but not all cases – it depends on the actual station autocorrelation and neighbour covariances. This is being coded up in R which has better (more robust?) statistical routines than IDL and is free/open access. However, this is a slow process and program run time is also far slower. All other tasks are pending on this one being complete so this is holding up everything else. A paper describing the methods is in preparation but depends on getting the code up and running for the example set of stations (MCDW over the USA – from ISTI databank stage 2).

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*

This work has not been started as it is waiting on further progress from Team Creation – see above.

4) *Design methods and create software for assessing the results of tests on the benchmarks ready for the Benchmark cycle*

This work has not been started as it is waiting on further progress from Team Creation – see above.

5) *Create a platform for guiding users of the benchmarks in how to use them and how the assessment works*

This work has not been started as it is waiting on further progress from Team Creation – see above.

Other Efforts and Achievements:

Funding for a PhD student with Met Office CASE sponsorship and supervision from Kate Willett has been found through EPSRC based at the College for Engineering, Mathematics and Physical Sciences, University of Exeter with Trevor Bailey and Ian Jolliffe as supervisors: Development of statistical tools for validation of methods for creating homogeneous observational climate daily datasets. Three candidates were interviewed and the PhD offered to and accepted by Rachel Warren. This PhD will develop a benchmarking process for daily data and work in collaboration with ISTI and the European homogenisation community.

Robert Lund now has Mike Finney working as a PhD student. Mike will study time series techniques to generate clean (homogeneous) data from a single station or a collection of stations. This research will focus on vector autoregressive time series models as a means to handle the spatial and temporal autocorrelations in our generated data. The developed methods will be able to simulate temperature series at daily levels, replete with their non-Gaussian aspects. Comparison to other regression and spatial based downscaling techniques will be investigated.

Victor Venema contacted Marlis Hofer; she may be able to obtain funding for a more advanced automatic downscaling method that would not only take the crosscorrelations between the stations into account, but also the influence of circulation types and consequently generate additional (decadal) variability in the difference time series.

2012 Annual Overview:

Progress during 2012 has been very slow and most of the objectives were not met. This is largely due to development and proof of concept of methods for creating the analog-known-worlds taking far longer than expected. It is also because the databank release has been slower than originally envisaged and the analogs have always been intended to be spun off this release which is still pending. Kate Willett has spent a lot of time working on the creation of analog-known-worlds but on top of other commitments has been unable to bring this to a stage at which it is ready to go forward. This has really stalled the rest of the objectives of the group. The intention is to call a meeting to update members on progress (or lack of) to date and develop a time line to go forward. Ideally, the analog-known-world method will be completed

during spring 2013 and both the overall benchmarking concept paper and analog-known-world methods paper submitted by summer 2013. Realistically, complete benchmark data will not be ready until late 2013.

Objectives for October 2012 to October 2013:

- 1) Benchmark Cycle concepts and plan formalised and submitted to JAOT or similar.
- 2) Create fully functioning open source software for producing the analog-known-worlds and submit a methods paper.
- 3) Create a first suite of analog-known-worlds.
- 4) Design methods and create software for producing the analog-error-worlds.
- 5) Design methods and create software for assessing the results of tests on the benchmarks ready for the Benchmark cycle
- 6) 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

Suggested timeline and plan for achieving objectives:

Objective	Description	Responsible Members	Deadline
Benchmark Concepts Paper	Overview outlining the entire Benchmark concept from analog-known-worlds, added error-models and validation	Kate Willett to circulate first draft, all to edit.	June 2013
Analog-known-worlds proof-of-concept	Create software to produce analog-known-worlds on a proof-of-concept scale and submit methods paper	Team Creation – lead by Robert Lund and Kate Willett	May 2013
Analog-known-worlds global scale production	Produce analog-known-worlds for as many ISTI databank stage 3 stations as possible	Team Creation – code probably run and data hosted by Kate Willett	May 2013
Analog-error-worlds concepts finalised	Decide upon number and type of error-models to create (including how to ensure that these are blind tests for each cycle)	Team Corruption – lead by Claude Williams	May 2013
Analog-error-worlds proof-of-concept	Create software to produce analog-error-worlds on a proof-of-concept scale and submit methods paper (if desired?)	Team Corruption – lead by Claude Williams	October 2013

Group meet up	Attempt to get as many together as possible, possibly a networked code sprint with a USA and Europe (Australia?) hub	All	September/October 2013
Analog-error-worlds global scale production	Produce analog-error-worlds from the analog-known-worlds ready for distribution	Team Corruption – lead by Claude Williams	November 2013
Validation concepts finalised	Decide upon number and type of tests with which to perform validation	Team Validation – lead by Ian Jolliffe	August 2013
Validation proof-of-concept	Create software and score system/intercomparison tables to run the validation on a proof-of-concept scale and submit methods paper (if desired?)	Team Validation – lead by Ian Jolliffe	October 2013
Validation global scale production	Produce software and framework ready for running on the global scale – automated or manual	Team Validation – lead by Ian Jolliffe	End 2013
Benchmark Cycle Release of analog-error-worlds	Release first official benchmarks – publicise widely	All – lead by Kate Willett	November/December 2013
Benchmarking Platform Design	Create a webpage showing step-by-step ‘How to benchmark’ with appropriate links to data, validation and intercomparison tables with registration so that feedback can be provided and contact maintained	All – lead by Kate Willett	December 2014 (ideally earlier but more important to get benchmarks created first)
Benchmark Cycle – release of the ‘answers’	Release the ‘answers’ (analog-known-worlds)	All – lead by Kate Willett	June 2016
Organise benchmark cycle 1 wrap-up workshop	Plan and run a workshop, perhaps in conjunction with full ISTI meeting or other conference?	All – lead by Kate Willett	June 2016 ready for late 2016?