# White Paper 10
# Dataset algorithm performance assessment based upon all efforts

- Consideration of the steps that need to be taken to provide suitable performance assessment for algorithms to develop datasets.

- How to make use of the benchmarking results developed in White Paper 9 and other information in order to assess suitability of datasets for a particular purpose.

- About how users can assess how well datasets represent real world behaviour for their particular application.

# White Paper 10

- Possible assessment criteria could include

    - ability to indentify long term trends

    - extent to which algorithm has retained variance of climate variable

    - ability of algorithm to identify and adjust for breaks in dataseries and suitable criteria for false positive identification of non existent break points

    - non-stationarity in the mean

    - realistic temporal correlations

    - realistic spatial correlation

- Assessed quality will depend on risks associated with making wrong inferences based on algorithm outputs

- Testing algorithms

    - Model-based tests

    - Reanalyses

    - Reference very high quality datasets

# Recommendations (1)

- Assessment criteria should be developed independently of dataset developers

- Assessment criteria should be produced that are appropriate to the purpose to which dataset will be put

- Output of an assessment will be to identify whether a dataset is fit for a particular purpose

- Outputs should be documented in such a form as to enable a clear decision tree for users

- Validation of an algorithm should always be carried out on a different dataset from that used to develop and tune the algorithm

- A key issue is to determine how well uncertainty estimates in datasets represent a true measure of the difference between the derived value and the "true" value

- Consider the future needs for climate services by identifying an appropriate set of regions or stations that any assessment should include

- New efforts should be coordinated with on-going regional and national activities to rescue and homogenize data

# Recommendations (2)

- Validation of an algorithm should always be carried out on a different dataset from that used to develop and tune the algorithm

- A key issue is to determine how well uncertainty estimates in datasets represent a true measure of the difference between the derived value and the "true" value

- Consider the future needs for climate services by identifying an appropriate set of regions or stations that any assessment should include

- New efforts should be coordinated with on-going regional and national activities to rescue and homogenize data

# White Paper 10 Feedback on blog

- Suggestion that rather than providing a single best data value at each particular point provide many data representing coherent spatio-temporal realisation from the distribution field of possible outcomes that users can Monte Carlo

- Don't get too hung on on winners and losers

  - Depends on purpose

  - Suggestion to pick "bad" methods that should not be considered further

- But - disagreeing with above point : a clear and detailed evaluation of efficiency characteristics for homogenisation methodologies will allow improvements to be made easily to existing methods or identify need to switch to alternative techniques

- Not just trend point detection, other inhomogeneities need to be detected, eg trend inhomogeneities. Overall goal is to time series with characteristics closest to the real world

- John Graham-Cumming offers to help eg with writing or reviewing code, writing tests or documentation