Publication and Collation of Data Products and Presentation of Audit Trails

Song Lianchun, Zhou Zijiang, Peter Thorne, Kate Willett, Guo Jianxia and Karl Taylor

Contents

- Data product platform and format and enabling intercomparison

- Acceptable publication criteria

- Correct Acknowledgment and Intellectual Property Rights Issues
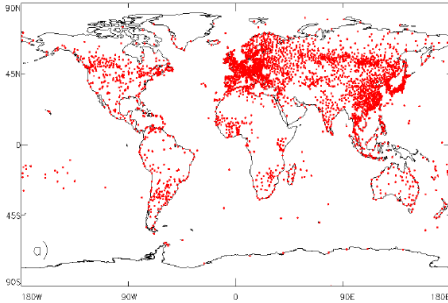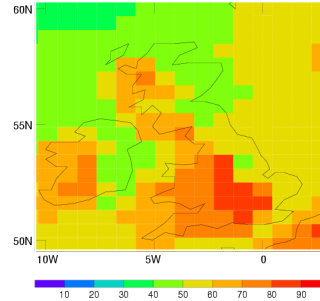
- Audit trails

Data Product Format, Platform and Enabling Intercomparison

Data Product Format, Platform and Enabling Intercomparison

eg. ASCII/netCDF

eg. gif/JPEG

METADATA

Something machine understandable

PLATFORM – like CMIP5 portal
Tables of all data products – level, version, spatial/temporal coverage, station/gridded, quality controlled, homogenised, created by etc.
Search engine
Files names according to a common format

INTERCOMPARISON
Graphics tools to create and compare maps, timeseries and basic statistics

Acceptable Publication Criteria of the Data Product to the Data Bank

Acceptable Publication Criteria for the Data Product to the Data Bank

List of source data (version, stations, period) and selection rationale

Documentation of the quality control method and flags

Homogeneous version of the data – where homogenisation methodology is documented in a previous peer-reviewed publication

Quality assessment report produced by running against at least a minimum set of the common test cases (WP 9 and 10)

Published paper documenting data construction method and related products in an ISI recognised journal

Audit trail document describing all processing steps – strong preference to include source code used

Correct Acknowledgments and Intellectual Property Rights Issues

Correct Acknowledgments and Intellectual Property Rights Issues

All users of the data bank and data products should clearly indicate source (level/version, region)

Appropriate acknowledgment statements for use of the portal/data bank should be stored on the data bank web portal

Data product users should also reference the associated journal publication

Audit Trails

Audit Trails

Sufficiently detailed step by step documentation of choices and methodology such that the product is reproducible:

list station used and periods considered

processing steps – particularly statistical, quality control and homogenisation methodology
THIS SHOULD INCLUDE CODE/SOFTWARE – IN CASES WHERE COPYRIGHT PROHIBITS THIS A STEP BY STEP LIST OF WHAT THE CODE DOES MAY SUFFICE

quality control flag metadata

breakpoint locations and adjustment factors

any ancillary datasets used

Recommendation Summary

Recommendation Summary

1) Publication Criteria for the data product in the data bank:
> - source data list and rationale
> - description of quality control method and metadata flags
> - homogeneous and/or gridded version of the data
> - quality assessment report
> - Peer reviewed publication
> - audit trail with publication of code used

2) Common data format from a common data portal akin to CMIP

3) Manipulation tools provided to allow endusers to access the data, produce plots and make intercomparisons

Questions?