



## **The Community Intercomparison Suite (CIS)**

Duncan Watson-Parris (1), Nick Schutgens (1), Nick Cook (2), Zak Kipling (1,3), Phil Kershaw (4), Ed Gryspeerd (5), Bryan Lawrence (4), and Philip Stier (1)

(1) Atmospheric, Oceanic and Planetary Physics, University of Oxford, Physics, Oxford, United Kingdom (duncan.watson-parris@physics.ox.ac.uk), (2) Tessella Ltd, Abingdon, Oxford, UK, (3) ECMWF, Reading, UK, (4) Centre for Environmental Data Analysis, STFC Rutherford Appleton Laboratory, Didcot, UK, (5) Department of Physics, Imperial College London, London, UK

Earth observations (both remote and in-situ) create vast amounts of data providing invaluable constraints for the climate science community. Efficient exploitation of these complex and highly heterogeneous datasets has been limited however by the lack of suitable software tools, particularly for comparison of gridded and ungridded data, thus reducing scientific productivity. CIS (<http://cistools.net>) is an open-source, command line tool and Python library which allows the straight-forward quantitative analysis, intercomparison and visualisation of remote sensing, in-situ and model data.

The CIS can read gridded and ungridded remote sensing, in-situ and model data - and many other data sources 'out-of-the-box', such as ESA Aerosol and Cloud CCI product, MODIS, Cloud CCI, Cloudsat, AERONET. Perhaps most importantly however CIS also employs a modular plugin architecture to allow for the reading of limitless different data types. Users are able to write their own plugins for reading the data sources which they are familiar with, and share them within the community, allowing all to benefit from their expertise.

To enable the intercomparison of this data the CIS provides a number of operations including: the aggregation of ungridded and gridded datasets to coarser representations using a number of different built in averaging kernels; the subsetting of data to reduce its extent or dimensionality; the co-location of two distinct datasets onto a single set of co-ordinates; the visualisation of the input or output data through a number of different plots and graphs; the evaluation of arbitrary mathematical expressions against any number of datasets; and a number of other supporting functions such as a statistical comparison of two co-located datasets. These operations can be performed efficiently on local machines or large computing clusters - and is already available on the JASMIN computing facility.

A case-study using the GASSP collection of in-situ aerosol observations will demonstrate the power of using CIS to perform model evaluations. The use of an open-source, community developed tool in this way opens up a huge amount of data which would previously have been inaccessible to many users, while also providing replicable, repeatable analysis which scientists and policy-makers alike can trust and understand.