Evaluating and benchmarking land surface models

Graham P. Weedon[1@], Martin J. Best[2] and Heather Ashton[2]

1 -Met Office, Maclean Building, Benson Lane, Crowmarsh Gifford, Oxfordshire, OX10 8BB, UK

@ - graham.weedon@metoffice.gov.uk

2 – Met Office, Fitzroy Road, Exeter, Devon, EX1 3PB, UK

Benchmarking of land surface models (LSMs) involves adopting widely agreed standards for judging performance. Unlike evaluation or validation, benchmarking requires comparison of outputs with pre-defined targets or thresholds; allowing meaningful intercomparisons of independent models.

Observations relevant to the assessment of modelled latent and sensible heat are available at a wide range of spatial scales (from the footprint area of flux towers to the large pixels of GRACE satellite products) and time steps (e.g. sub-daily to monthly). Thus, an evaluation system, as part of benchmarking system, should be flexible enough to process model output and observations at whatever time steps and spatial resolutions they are available. NASA's LVT (Land Verification Toolkit, Kumar et al., 2012 *Geosci. Model Dev.*), developed as part of the LIS suite, has this flexibility. It also provides a wide range of analytical metrics with in-built assessment of the uncertainties of the metrics (e.g. 95% CIs).

In this presentation we will describe the development of a new land surface benchmarking system based upon LVT. Initially the system will consider fluxes of momentum, heat, moisture and carbon from point observations, but will be further developed to include satellite data for state variables such as land surface temperature and soil moisture. The system will build upon early work on benchmarking identified in the PLUMBER experiment (Best et al., 2015 *J. Hydrometeorol.*) to identify suitable metrics that can be equally applied for both stand alone and coupled simulations. It is anticipated that the basis of this system could be expanded to cover all aspects of LSMs with the eventual aim of establishing an international standard LSM benchmarking suite.