Geophysical Research Abstracts
Vol. 20, EGU2018-5317-1, 2018
EGU General Assembly 2018
© Author(s) 2018. CC Attribution 4.0 license.

Revisiting the PLUMBER Experiments from a Process-Diagnostics Perspective

Grey Nearing (1), Benjamin Ruddell (2), Martyn Clark (3), Bart Nissan (4), and Christa Peters-Lidard (5)

(1) University of Alabama; Geological Sciences; Tuscaloosa, AL, United States (gsnearing@ua.edu), (2) Northern Arizona University; School of Informatics, Computing, and Cyber Systems; Flagstaff, AZ, United States (, (3) NCAR – Research Applications Laboratory; Boulder, CO, (4) University of Washington – Department of Civil & Environmental Engineering; Seattle, WA, (5) NASA Goddard Space Flight Center – Hydrological Sciences Laboratory; Greenbelt, MD

The PLUMBER benchmarking experiments [1] showed that some of the most sophisticated land models (CABLE, CH-TESSEL, COLA-SSiB, ISBA-SURFEX, JULES, Mosaic, Noah, ORCHIDEE) were outperformed – in simulations of half-hourly surface energy fluxes – by instantaneous, out-of-sample, and globally-stationary regressions with no state memory.

One criticism of PLUMBER is that the benchmarking methodology was not derived formally, so that applying a similar methodology with different performance metrics can result in qualitatively different results. Another common criticism of model intercomparison projects in general is that they offer little insight into process-level deficiencies in the models, and therefore are of marginal value for helping to improve the models.

We address both of these issues by proposing a formal benchmarking methodology that also yields a formal and quantitative method for process-level diagnostics. We apply this to the PLUMBER experiments to show that (1) the PLUMBER conclusions were generally correct – the models use only a fraction of the information available to them from met forcing data (<50% by our analysis), and (2) all of the land models investigated by PLUMBER have similar process-level error structures, and therefore together do not represent a meaningful sample of strucural or epistemic uncertainty.

We conclude by suggesting two ways to improve the experimental design of model intercomparison and/or model benchmarking studies like PLUMBER. First, PLUMBER did not report model parameter values, and it is necessary to know these values to separate parameter uncertainty from structural uncertainty. This is a first order requirement if we want to use intercomparison studies to provide feedback to model development. Second, technical documentation of land models is inadequate. Future model intercomparison projects should begin with collaborative effort by model developers to document specific differences between model structures. This could be done in a reproducible way using a unified, process-flexible system like SUMMA [2].

[1] Best, M.J. et al. (2015) 'The plumbing of land surface models: benchmarking model performance', J. Hydrometeor.
[2] Clark, M.P. et al. (2015) 'A unified approach for process-based hydrologic modeling: 1. Modeling concept', Water Resour. Res.