

Dissecting PLUMBER: Why are land surface models performing so poorly?

N. Haughton^a, G. Abramowitz^a, A. Pitman^a

^a *Climate Change Research Centre, UNSW*
Email: ned@nedhaughton.com

Abstract: The PALS Land Surface Model Benchmarking Evaluation Project (PLUMBER) (Best et al. submitted) showed a serious across-the-board problem with the performance of 13 international land surface models. PLUMBER found that in key predicted atmospheric fluxes, land surface models are consistently outperformed by relatively simple, out-of-sample empirical models. The problem is especially apparent in sensible heat flux, where *all* land surface models are outperformed by a simple out-of-sample linear regression on incoming shortwave radiation.

Here we briefly describe the PLUMBER results, and then attempt to rule out potential causes of this poor performance. In particular, we investigate: the aggregation methods used in PLUMBER, to rule out that these results are not due to particular sites or metrics used in the experiment; the potential for time-resolution problems, which are potentially important, because land surface models are designed to predict over longer timescales; and the fact that empirical models are not bound by the laws of energy conservation, potentially giving them an advantage in a benchmarking situation.

We find evidence that the performance problems shown in PLUMBER are not due to any of these causes. These results indicate that the problem lies deeper, and we propose a number of further methods of investigation that may provide more insight.