Geosci. Model Dev. Discuss., 5, C184–C186, 2012
www.geosci-model-dev-discuss.net/5/C184/2012/

Geoscientific
Model Development
Discussions

Interactive comment on "Land surface Verification Toolkit (LVT) – a generalized framework for land surface model evaluation" *by* S. V. Kumar et al.

Anonymous Referee #2

Received and published: 20 April 2012

This paper describes a new model evaluation system, called the Land surface Verification Toolkit (LVT), designed to operate in conjunction with NASA's Land Information System (LIS). LVT uses a large collection of *in situ*, remote sensing, reanalysis, and data assimilation-based model results to judge the fidelity of hydrological model results from multiple land surface models operating within LIS or results translated into LIS-style outputs. LVT offers a variety of metrics, including those for point comparisons, spatial correspondence, ensemble and uncertainty measures, information theoretic metrics, and spatial scale impacts on model performance.

§1 is well-written and appropriately motivates discussion of LVT.

In §2, the authors describe several community-wide efforts—citing GSWP, ALMIP, and C-LAMP—as providing "benchmarks for the simulation of continental scale water and

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

energy budgets." However, the authors fail to mention that C-LAMP also included biogeochemical model evaluations of global forest phenology, global primary productivity, CO_2 seasonality, and regional carbon stocks and dynamics. Overall, §2 offers a good background of previous and current model evaluation efforts and provides the context for development of LVT.

While the system is conceptually well described in the paper, architectural details are lacking, resource requirements are not described, and computational performance is not discussed. In particular, §3 describes the three-layer structure of LVT, but offers no details on how the *Analysis Metrics* and *Observations* in the Abstractions layer are implemented in the Fortran 90 and C languages. Are *Metrics* templated for ease of extensibility? ESMF would seem to be a heavy-weight solution to provide only clock/time, configuration, and logging infrastructure in the Core Structure and Features layer. Does it also provide the geospatial transformations functionality, or are other packages employed for that purpose? What are the typical memory and CPU requirements of the LVT package, and what are its computational and I/O performance characteristics? Does LVT employ shared- or distributed-memory parallelism?

In §4, the authors describe a philosophy of preserving datasets in their native formats, but it is not clear if LVT reprocesses those datasets every time an evaluation is executed to make them consistent, spatially and temporally, with model results. For high resolution remote sensing datasets, generating products for spatial comparison with low resolution model grids and low frequency output could take considerable compute time. How are these issues handled? §4.2 offers useful motivation for careful consideration of appropriate analysis metrics.

The examples provided in §5 are extremely valuable, but no details are provided about how graphical diagnostics are produced. Are one or more graphical packages employed by LVT, or is the user left to generate graphics from numerical output from LVT? For the MDF paradigm example described in §5.1, it is slightly misleading that the authors suggest "calibration of model parameters helps in improving the model perfor-

GMDD
5, C184–C186, 2012

mance, by correcting both these systematic biases", when the sensible and latent heat biases are really the result of a single bias in energy partitioning.

§6 does a good job of summarizing the key points of the paper.

This paper does not describe the relative importance of different metrics or if LVT provides a mechanism for weighting metrics or scoring models based on a collection of these metrics, incorporating model and observational uncertainties. Does LVT offer such a mechanism for judging the overall performance of models for one or more specific application areas?

A few potential typographical errors were noted. On page 244, line 20, the article "the" may be extraneous before "near-optimal performance." On page 246, line 13, "of" is missing before "such sensitivities." On page 248, lines 13 and 14, the "in" and "to" that span those lines might more appropriately be one word. On page 248, line 28, an article ("a" or "the") may improve the readability before "sum of orthogonal components". On page 240, line 11, "representation" should probably be plural.

Interactive comment on Geosci. Model Dev. Discuss., 5, 229, 2012.

GMDD

5, C184–C186, 2012