Introduction to Land surface Verification Toolkit (LVT)

Sujay V. Kumar

Science Applications International Corporation/NASA Hydrological Sciences Laboratory
The LIS modeling suite

Land Information System (LIS)

Land surface Data Toolkit (LDT)
- Land surface parameter processing
- DA/OPTUE preprocessing
- Downscaling support
- Forcing adjustments (bias correction)
- Restart/ensemble generation

GREY BOX
- DA (EnKF, EnKS)
- Optimization/ Uncertainty Estimation (LM, GA, MCMC)
- Models (Noah, VIC, CLSM, JULES, SAC/ SNOW17, FLake, HyMAP)
- Meteorological data (NLDAS, MERRA, GPM, ECMWF...)
- RTMs (CRTM, CMEM)
- Remote sensing data (SMOS, SMAP, MODIS, VIIRS, GOES, IMS...)
- High performance computing support

Land surface Verification Toolkit (LVT)
- Model evaluation and benchmarking
- Hydrological products (drought indices, flood indicators)

SNOTEL
SCAN
MODIS
Motivation

Quantitative measures of fidelity of model simulations are essential for improving the usage and acceptability of model forecasts for real-world applications.

Characterization of accuracy and uncertainty in model predictions - to be used as a benchmark for future model enhancements.

Need formal evaluation procedures to improve the "observability" of LSM processes.

Need a general benchmarking framework capable of capturing useful modes of variability of LSMs through a range of performance metrics is necessary for further advancing the performance and predictability of models.
The objective of this paper is to provide guidance to the LSM community on how to make better use of eddy covariance data, particularly via MDF. We first outline the philosophical principles behind model-data fusion for model improvement. We then discuss the structure of typical land surface models and how they are parameterised. Next we discuss new techniques for model and data evaluation, focussing on new techniques enabling the MDF concept. MDF is used to estimate model parameters to yield model predictions that match the calibration data, particularly via MDF. We conclude with consideration of the constraints in model evaluation and improvement, and effectiveness of model-data fusion, MDF (Raupach et al., 2005). MDF encompasses a range of procedures for combining a set or sets of observations and a model, while quantitatively incorporating information from models and data. Rather, it is a multi-stage process (Fig. 1). At each of these stages, there is interplay between data, model formulation, characterization and evaluation data. Multiple data constraints are used to help formulating, characterizing and evaluating models in a structured manner. A comprehensive evaluation and benchmarking framework is essential for enabling the MDF concept.
LVT functions both as a **verification** and **benchmarking** environment.

**Evaluation** - model outputs are compared to observations to derive an error measure.

**Comparison** - model is not just compared to observations, but also to other models.

**Benchmarking** - performance expectation defined a priori.

---

**The Plumbing of Land Surface Models: Benchmarking Model Performance**


*ARC Centre of Excellence for Climate Science, University of New South Wales, Sydney, New South Wales, Australia
*ECMWF, Reading, United Kingdom
*CNRM-GAME, Météo-France, Toulouse, France
*Helmholtz Centre for Environmental Research-UFZ, Leipzig, Germany
*Center for Ocean–Land–Atmosphere Studies, George Mason University, Fairfax, Virginia
*NOAA/NCEP/EMC, College Park, Maryland
*Oceans and Atmosphere Flagship, CSIRO, Canberra, Australian Capital Territory, Australia
*KNMI, De Bilt, Netherlands
*Hydrological Sciences Laboratory, NASA GSFC, Greenbelt, Maryland
*Oceans and Atmosphere Flagship, CSIRO, Aspendale, Victoria, Australia
*Laboratoire des Sciences du Ciel et de l’Environnement, UMR 8212, IPSL–LASC, CEA-CNRS-UVSQ, Gif-sur-Yvette, France

(Manuscript received 27 August 2014, in final form 19 December 2014)

**ABSTRACT**

The Protocol for the Analysis of Land Surface Models (PALS) Land Surface Model Benchmarking Evaluation Project (PLUMBER) was designed to be a land surface model (LSM) benchmarking intercomparison. Unlike the traditional methods of LSM evaluation or comparison, benchmarking uses a fundamentally different approach in that it sets expectations of performance in a range of metrics a priori—before model simulations are performed. This can lead to very different conclusions about LSM performance. For this study, both simple
Comparisons (MIPs ..)

- Identifies metrics for which one model performs better than another, or where errors in multiple models are systematic.
- Indicates where performance improvements are possible/not possible relative to other models.
- Too much reliance on model comparisons - models may end up being developed too similar to each other.
Simply comparing models and observations – canonical “evaluation” – can’t tell us whether any of the models are doing a good job.

Benchmarking involves defining expectations of performance in any metric of interest a priori – before running model. Options include:

- previous model version (weak – both models could be poor)
- fit for a particular application (stronger / useful – can tell us if a model is “good enough”)
- effectively utilizes available information (strong – can give us an objective definition of whether a model is “good”) - defines a priori expectations based on the complexity of the model and the amount of information given to it.

We would typically accept this as a good simulation (good correlation visually)

Benchmarking will reveal that this is in fact a poor simulation

source: Gab Abramowitz
How well should we expect a LSM to predict latent heat (Qle) flux at Amplero site?

1. Take several (19) flux tower sites other than Amplero
2. Train a linear regression between downward shortwave radiation and Qle
3. Use these regression parameters to predict Qle at Amplero using site meteorology

This will tell us:

- The extent to which Qle is predictable from SWdown alone.
- How predictable Qle is at Amplero site - is it unusually difficult?

Even the 1-variable regression beats the model!

(source: Gab Abramowitz)
LIS was/is being used in many different configurations (557, NCEP, NOHRSC, CRREL, NRL, NLDAS, GLDAS, FLDAS, MSFC, NU-WRF, ICBA, …)

LIS outputs being produced in many different formats (grib, NetCDF, binary), different resolutions, map projections, modes (tile, grid, ensembles)

The typical next step is to compare the model outputs to reference datasets for evaluation

LVT was originally designed to bridge this gap - by having a framework that allows the comparison of LIS output against other datasets

- Includes support for a range of in-situ, remote sensing and model/renalisation products
- Supports the analysis of outputs from various LIS subsystems (LIS-DA, LIS-OPT/UE)
- Includes the capability to generate end-user oriented hydrological products (drought/flood percentiles, indicators)
- Very LIS-reliant, non-LIS datasets require pre-processing to make them “LIS-like”
Redesigned to handle **any two land relevant datasets** (need not be a LIS output)

In addition to all other existing capabilities, some initial benchmarking capabilities have been developed.

The supported datasets in LVT can be used to develop benchmarks using simple (regression) to more complex (ANN-ish) methods.
General capabilities

🌟 Reconciles the differences in spatial and temporal resolutions between the two datastreams being compared, by bringing them to a common (user specified space and time domain)
General capabilities

- Emphasis on supporting datasets natively, as much as possible - Users can download the data by themselves and employ them in LVT
  - E.g. ARM-CART (NetCDF), AGRMET (Grib), SCAN in-situ (ASCII) …
  - A reader/processor needs to be built for each dataset

- Many options for masking/stratification of metrics
  - Data count based mask
  - External static mask
  - External time varying mask
  - Variable-based stratification (e.g. day-night stratification using SWdown)
  - External data based stratification (e.g. landcover, soils, elevation)

- RMSE for Evergreen Needleleaf Forest
- RMSE for Evergreen Broadleaf Forest
- RMSE for Deciduous Needleleaf Forest
- RMSE for Deciduous Broadleaf Forest
General capabilities

- Analysis outputs provided in both gridded (NetCDF/binary) and ASCII formats

- Time-lagged computations

- Supports water-year (flexible year specification)
  - User specifies the starting month of the year specification

- Smoothing support (limited)
  - Uses a moving window average for the computation of analysis metrics

- Computes confidence intervals (currently CIs in the spatial domain is supported; It will be extended to include temporal CIs)
General capabilities

- Spatial averaging modes for analysis metric can be computed on a pixel-by-pixel basis or at basin averaged basis
  - Pixel-by-pixel - each pixel in datastream 1 is compared to a pixel in datastream 2
  - Basin-averaged - datastream 1 and 2 values are averaged to the basin scale and then compared using the analysis metric

- Computes derived variables
  - e.g. Bowen ratio can be computed through LVT (and used for analysis) if both Qle and Qh are present, A column averaged, weighted root zone soil moisture if individual soil moisture layer values are present
  - Energy/Water/Evaporation balance values

- Analysis metric computations are performed
  - Across the entire analysis period
  - At specified temporal intervals
  - Average seasonal/diurnal cycles (if specified by the user)

- Supports outputs from all LIS computational subsystems
  - Data assimilation diagnostics from the LIS-DA output
Software architecture

- 3-layer architecture
- Specified as an object oriented framework with plugins defined for
  - Analysis metrics
  - Datastreams
  - Training algorithms
- Analysis instances are enabled by a config file (no external scripting required)
Supported data streams

**In-situ**
- Ameriflux fluxes
- ARM fluxes, soil moisture, soil temperature
- ARS soil moisture
- CEOP fluxes, soil moisture, soil temperature
- CPC precipitation
- FLUXNET fluxes
- FMI SWE
- GHCN snow depth
- GLERL lake fluxes, temperature
- ISMN soil moisture
- NASMD soil moisture
- PBOH2O soil moisture, snow depth
- SCAN soil moisture
- SMOSREX soil moisture
- SNODEP snow depth metobs
- SNOTEL SWE
- SURFRAD radiation
- USGS streamflow
- USGS groundwater
- ...

**Satellite/Remote Sensing**
- ALEXI
- AMSR-E SWE/snowdepth
- LPRM AMSR-E soil moisture
- ESA CCI soil moisture
- GloMMS NDVI
- GlobSnow SWE
- GRACE TWS
- ISCCP LST
- MOD10A1 snow cover
- MOD16A2 ET
- MODIS LST
- SMOOPS soil moisture
- SMOS L1 Tb
- SMOS L2 soil moisture
- UW ET
- ...

**Model/Reanalysis**
- AGRMET
- GLDAS2
- NLDAS2
- LIS outputs
- MERRA2
- SNODAS
- CMC
- GL6 JULES
- ERA interim Land
- MERRA Land
- COAMPS
- …
# Supported analysis metrics

<table>
<thead>
<tr>
<th>Metric class</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Diagnostics</strong></td>
<td>Mean, Standard deviation, Anomaly, Tendency, Min, Max, Sum, Maxtime, Mintime</td>
</tr>
<tr>
<td><strong>Accuracy</strong></td>
<td>ACC, Bias, CSI, ETS, FAR, FBIAS, MAE, NSE, PODY, PODN, POFD, Correlation, Anomaly Correlation, Tendency Correlation, unbiased RMSE</td>
</tr>
<tr>
<td><strong>Indicators</strong></td>
<td>SPI, SRI, SSWI, SSGI, percentiles, probabilistic percentiles</td>
</tr>
<tr>
<td><strong>Ensemble</strong></td>
<td>Mean, Likelihood, Spread, Cross correlation, ME</td>
</tr>
<tr>
<td><strong>Information theory</strong></td>
<td>Metric entropy, Information gain, Effective complexity, Fluctuation complexity</td>
</tr>
<tr>
<td><strong>Scale decomposition</strong></td>
<td>Discrete wavelet transforms</td>
</tr>
<tr>
<td><strong>Spatial similarity</strong></td>
<td>Hausdorff norm</td>
</tr>
</tbody>
</table>
A suite of common, normalized indicators has been developed:

- SPI, SRI, SSWI, SSGI, percentiles

These indicators are computed as deviations from long term (fitted/computed) distributions.
Examples of ensemble analysis

![Graph A](image)

- Soil Moisture (m3/m3)
- Ensemble Mean
- Obs
- 2010/05 to 2010/09

![Graph B](image)

- Uncertainty Importance
- 2010/05 to 2010/09

Uncertainty importance: An assessment of the relative contribution of each parameter to the ensemble spread (cross correlation between the simulated variable and the parameter, across the ensemble)

Can be used to guide parameter optimization/uncertainty estimation studies
How much of this improvement will be obtained at coarser spatial resolutions where the topography is not well resolved?
Information theory metrics

Time series analysis designed to detect patterns

- Periodical
- High-structured
- Random

Intuitive relationship between information and complexity from Pachepsky et al. (2006)

Change in metric entropy through the assimilation of AMSR-E soil moisture retrievals

NASA AMSR-E

LPRM AMSR-E
LVT provides two capabilities related to benchmarking:

- Develop a benchmark dataset by training any two of the supported datasets
- Compare the model runs to the benchmark dataset

Training algorithms available:

- One-variable regression
- Two-variable regression
- ANN (coming soon..)
Currently LVT works only in a serial mode. Multi-processor capabilities are being added.

Spectral/cross-spectral analysis (along Weedon et al., JHM 2015)

Expand the suite of indicators (e.g. multi-variable based)