Integrated Metrics and Benchmarking for the North American Land Data Assimilation System (NLDAS)

David M. Mocko[1,2], Christa D. Peters-Lidard[1], Sujay V. Kumar[1,2], Shugong Wang[1,2], Kristi Arsenault[1,2], Grey Nearing[1,2], Youlong Xia[3,4], Michael B. Ek[3], and Jiarui Dong[3,4]

1 – Hydrological Sciences Laboratory, NASA/GSFC, Greenbelt, MD
2 – SAIC at NASA/GSFC, Greenbelt, MD, USA
3 – NOAA/NCEP/EMC, College Park, MD, USA
4 – IMSG at NCEP/EMC, College Park, MD, USA

This work is funded in part by the NOAA Climate Program Office's Modeling, Analysis, Predictions, and Projections (MAPP) Program

Acknowledgements:
David Lawrence, Augusto Getirana, Yuqiong Liu, Yudong Tian, Gabrielle De Lannoy, Rolf Reichle, Qing Liu, Sarith Mahanama, Ben Livneh, Benjamin Zaitchik, Hiroko Kato Beaudoing, Mike Cosh

Presentation Outline

- Introduction to the North American Land Data Assimilation System (NLDAS) project and the new/upgraded LSMs for the next phase on NLDAS
  - **Noah-3.3, Catchment/Fortuna-2.5 (shown here); SAC-HTET-3.5.6/SNOW-17, VIC-4.1.2.l (in LIS-7)**
- Simulations with the Land Information System (LIS)
  - **Using the LIS software framework for data assimilation of remotely-sensed water states**
- Model evaluations and benchmarks using the Land Verification Toolkit (LVT)
  - **Soil moisture, Surface fluxes, Snow, Streamflow**
  - **Regression model development and evaluation**

NLDAS Land Surface Models (LSMs)

- NLDAS Phase 2 is currently running in operations in near-real-time (3.5-day lag) to drive a suite of LSMs (**Noah-2.8**, **Mosaic**, **SAC/SNOW-17**, and **VIC-4.0.3**).
- The NLDAS-2 LSMs have been extensively evaluated against many observations in papers led by Xia et al.
- For the next phase of NLDAS, new and upgraded LSMs are being run using the NASA-developed LIS software framework, including the use of data assimilation.
- 1/8th-deg. centered over CONUS (25-53° N; 125-67° W).
- A 60-year spin-up, then simulations from 1979 – 2012.

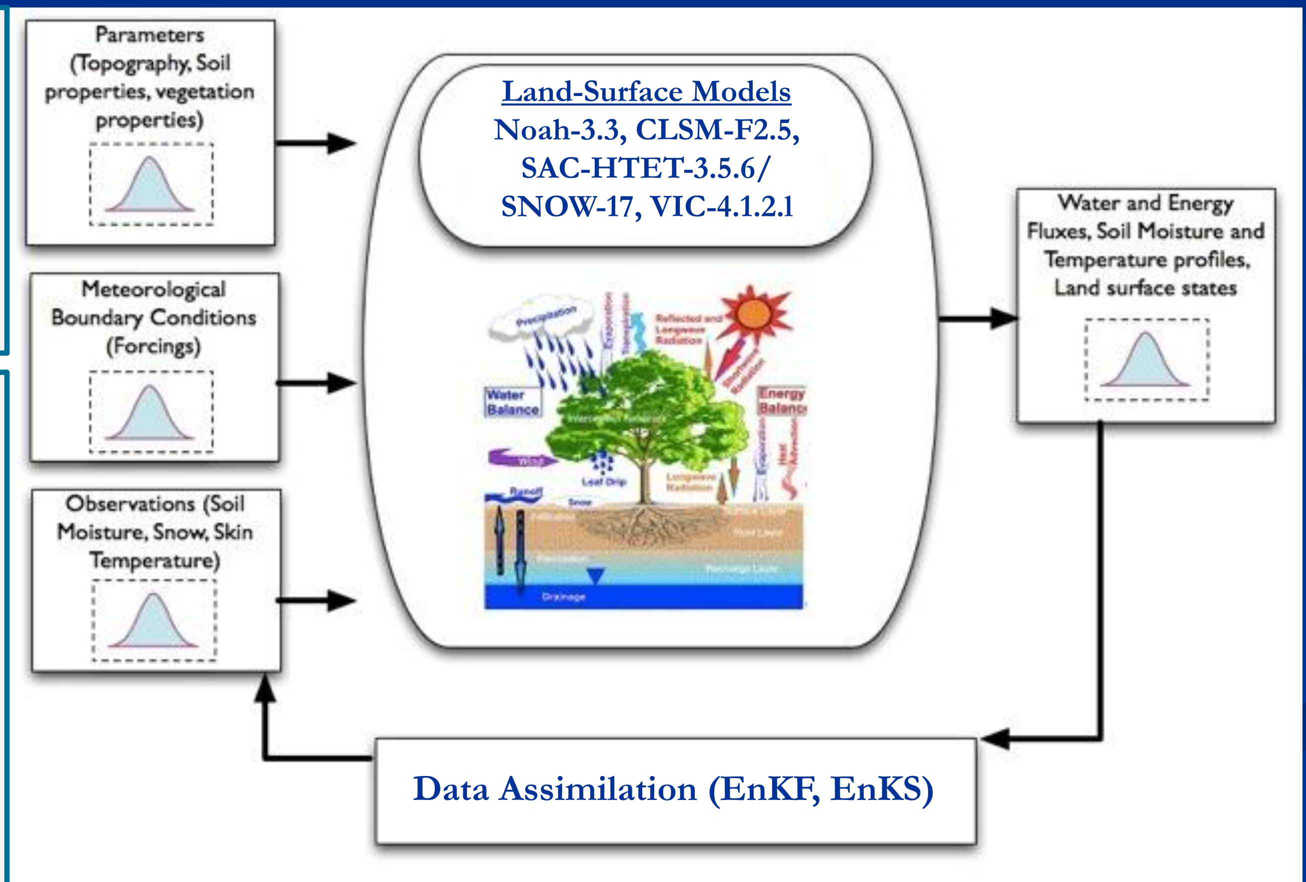
Reference(s): Xia et al. (2012a&b) – NLDAS-2 introduction and streamflow evaluation – *JGRa*; Xia et al. (2013) – soil temperature evaluation – *J. Appl. Meteor. Climatol.*; & numerous others.

The Land Information System (LIS)

- LIS is a flexible land-surface modeling and data assimilation framework developed with the goal of integrating satellite- and ground-based observed data products with land-surface models.

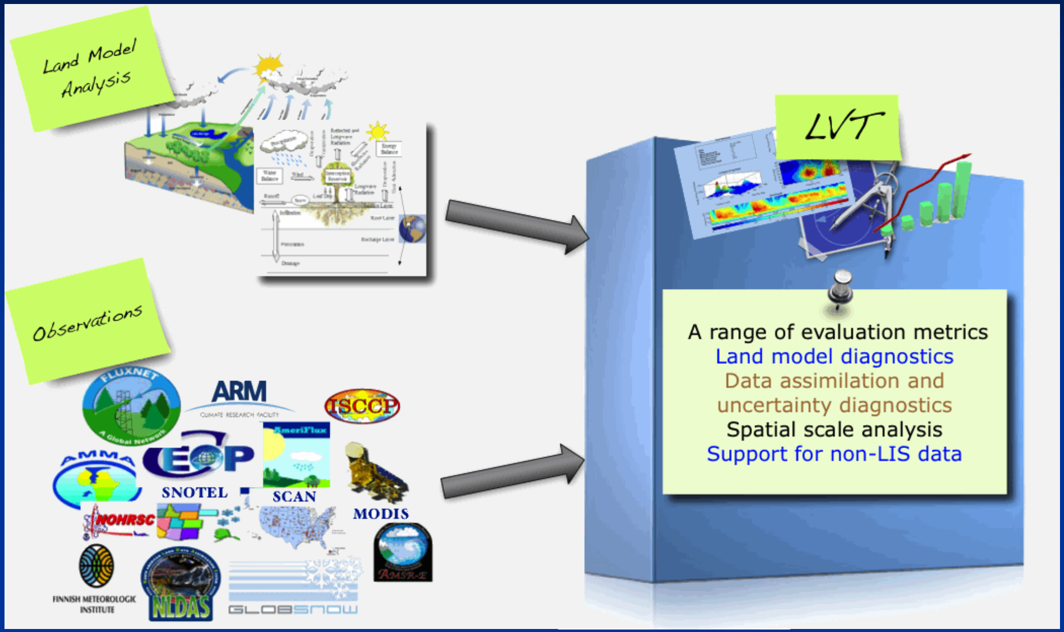
1979-present NLDAS-2 Forcing and Parameters

Data assimilation of: soil moisture, snow depth/area, terrestrial water storage (TWS), and irrigation intensity



Reference(s): Kumar et al. (2006) in *Environmental Modelling & Software*
Peters-Lidard et al. (2007) in *Innovations in Systems and Software Engineering*

The Land Verification Toolkit (LVT)

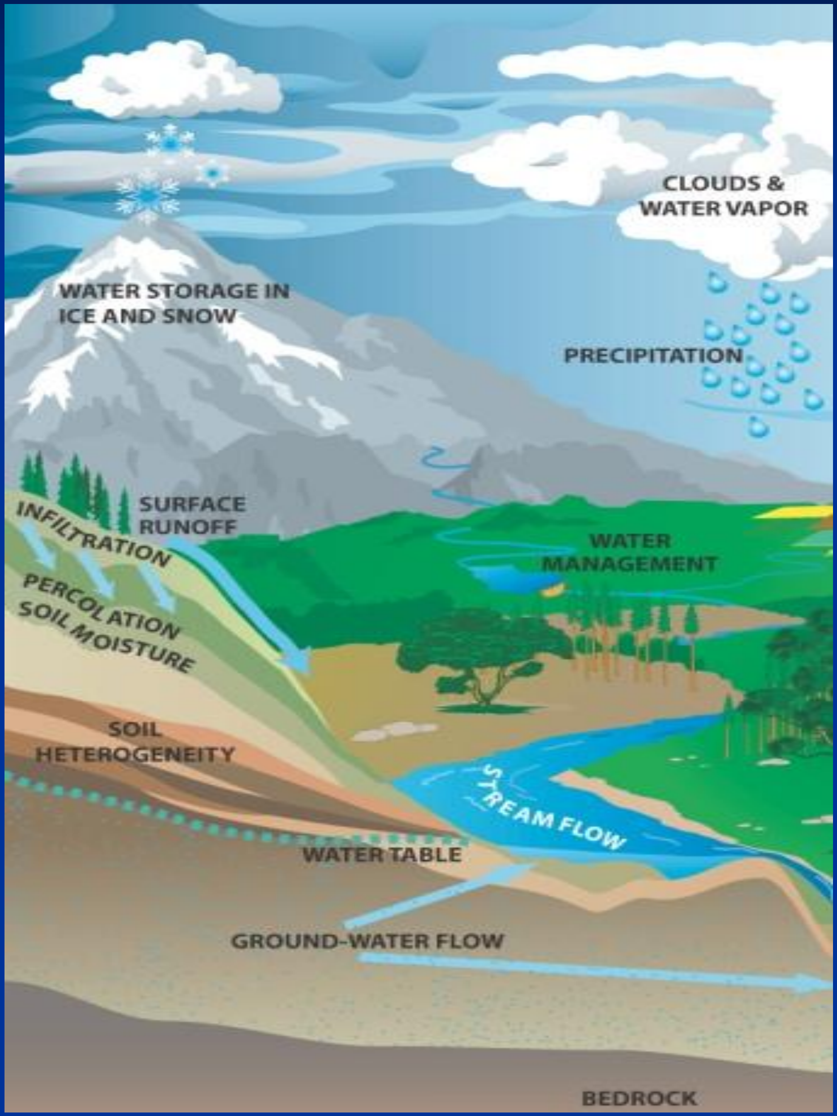


Metric Class	Examples
Accuracy	RMSE, Bias, Correlation
Ensemble	Mean, Standard deviation, Likelihood
Uncertainty	Uncertainty importance
Information theory	Entropy, Complexity
Data assimilation	Mean, variance, lag correlations of innovation distributions
Spatial similarity	Hausdorff distance
Scale decomposition	Discrete wavelet transforms

- LVT is a NASA-developed open-source software framework developed to provide an automated, consolidated environment for systematic land surface model evaluation and benchmarking
- Includes support for a range of in-situ, remote-sensing, and other model and reanalysis products in their native formats

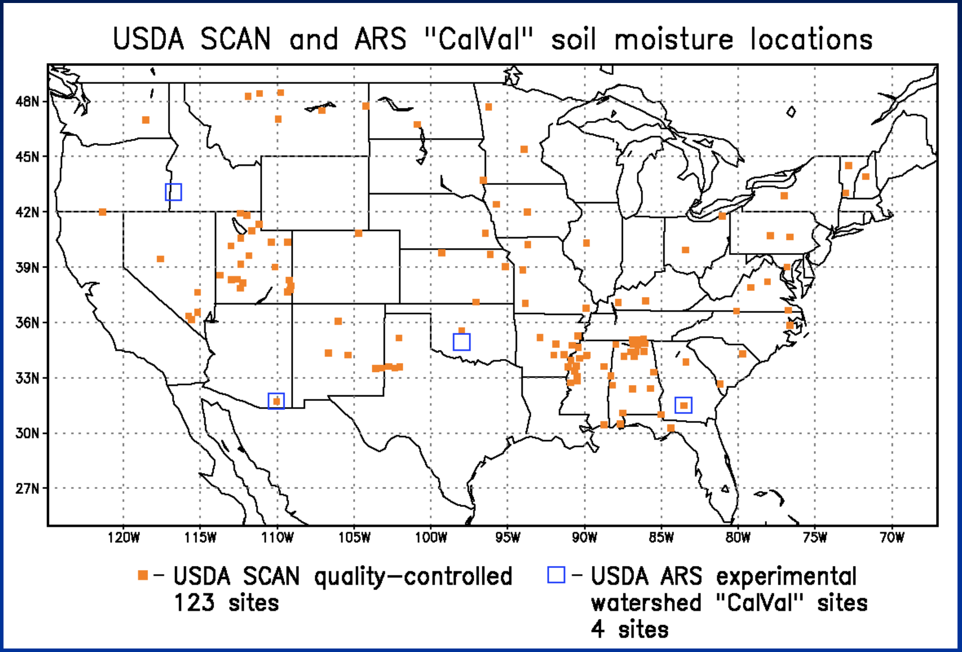
Reference: Kumar et al. (2012): Land surface Verification Toolkit (LVT) – A generalized framework for land surface model evaluation. *Geosci. Model. Dev.*

Benchmarking philosophy



1) Compare against available independent data – evaluating against every component of the water cycle, including budgets, balances, and combination variables (SWE/P, E/P, Q/P).
2) Normalize metrics to allow integration between different observations/components, and to ease comparison of the LSMs.
3) Investigate if the LSMs provide additional information compared to a regression model. Similar to the PALS benchmarking protocol.

Evaluation of soil moisture fields



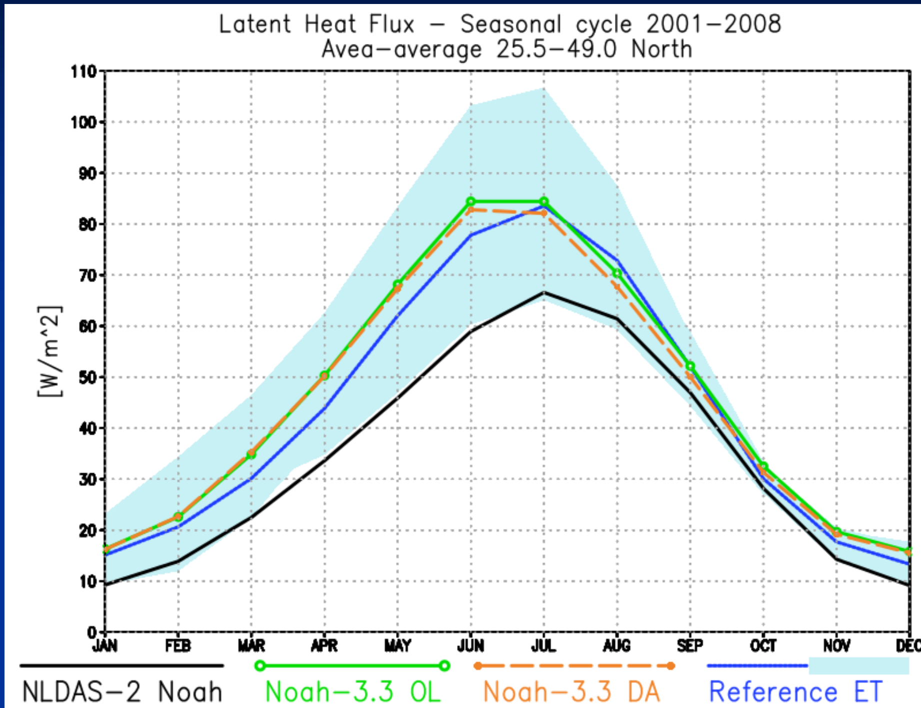
LSM (version)	ARS	SCAN
NLDAS-2 Mosaic	0.758	0.661 +/- 0.025
NLDAS-2 SAC	0.757	0.701 +/- 0.027
NLDAS-2 VIC-4.0.3	0.592	0.495 +/- 0.029
NLDAS-2 Noah-2.8	0.694	0.634 +/- 0.028
Noah-3.3 OL	0.711	0.614 +/- 0.024
CLSM-F2.5 OL	0.581	0.493 +/- 0.030

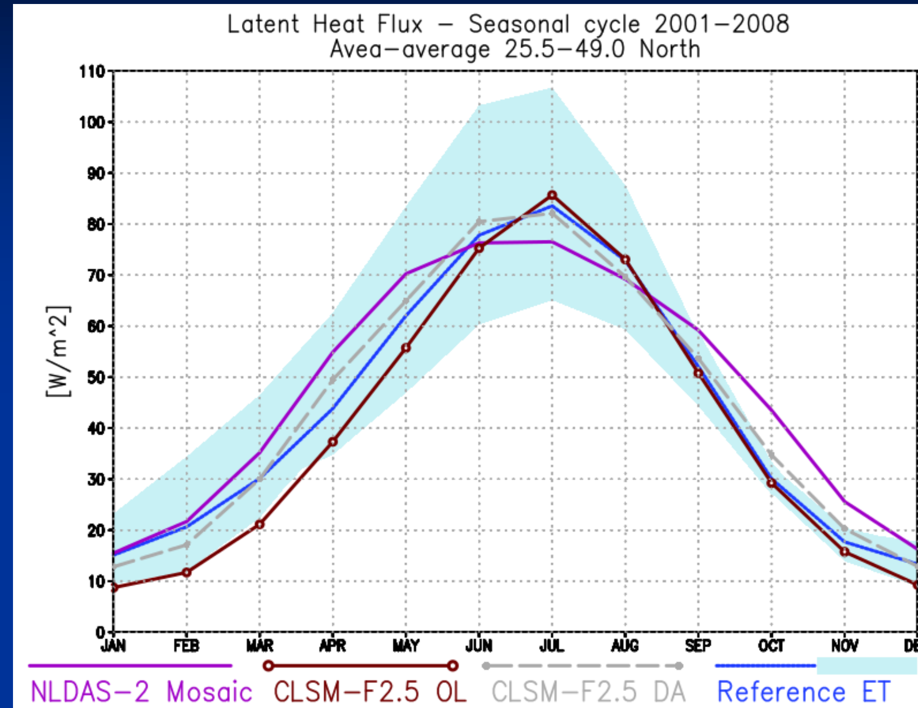
Anomaly R values

The surface soil moisture (5cm) was evaluated against quality-controlled soil moisture observations from the **USDA SCAN (123 sites, 2000-2012)** and **ARS "CalVal" (4 sites, 2001-2011)** networks. Data assimilation results can be found in Kumar et al. (2014, *JHM*) – & his talk Tue 11:00am (Paper 3.1).

Reference(s): Jackson et al. (2012) – ARS – *IEEE TRGS*; Schaefer et al. – SCAN – *J. Atmos. Oceanic Technol.*; Liu et al. (2011) – soil moisture skill with land DA – *JHM*

Evaluation of gridded surface fluxes

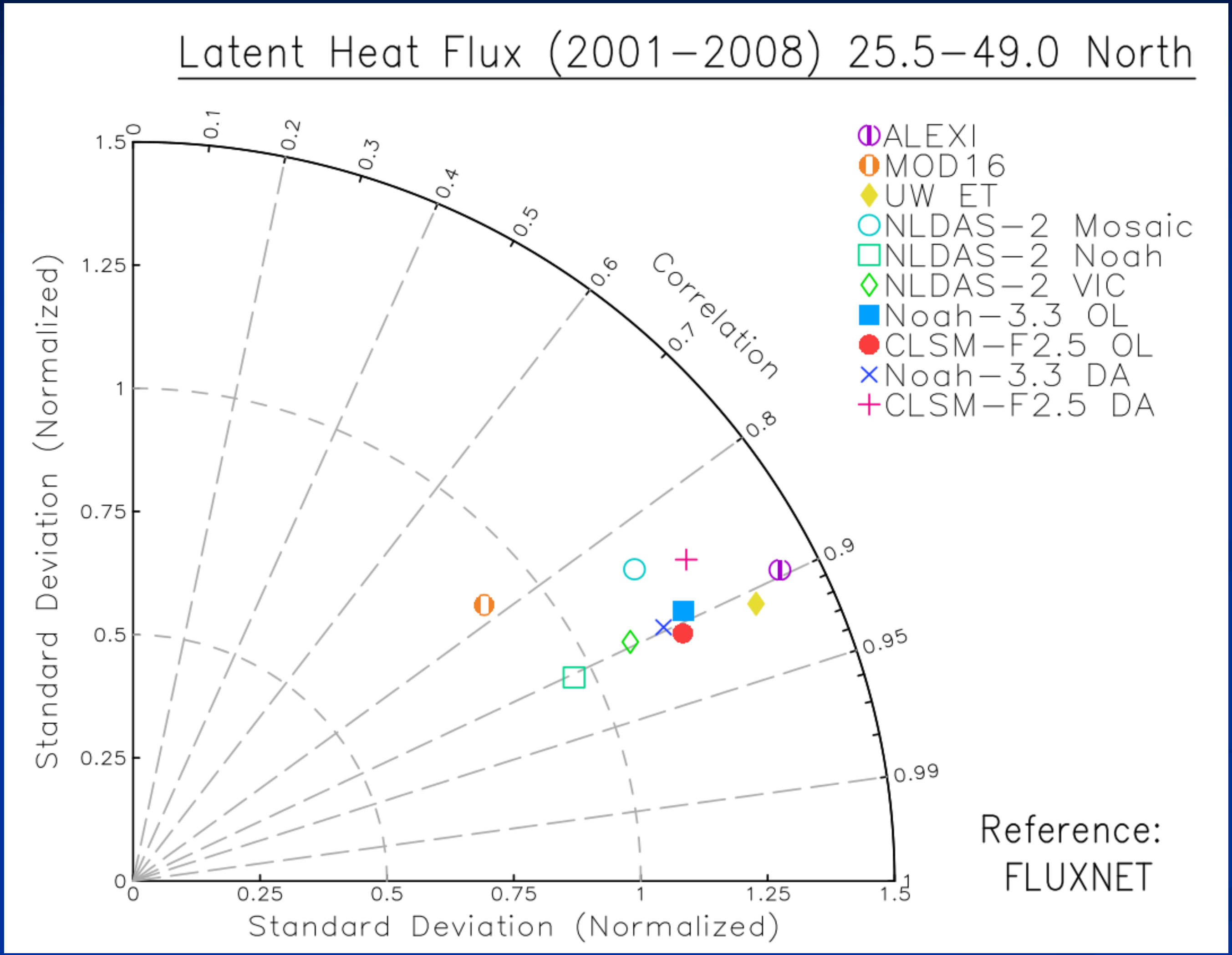




Gridded monthly surface flux reference products (in blue, the **mean**/**range**) based on FLUXNET surface observations, on MODIS retrievals (MOD16 and UW ET), and on thermal remote sensing (ALEXI). (Left) NLDAS-2 Noah and Noah-3.3; and (Right) NLDAS-2 Mosaic and CLSM-F2.5.

Reference(s): Jung et al. (2009) – FLUXNET – *Biogeosci.*; Mu et al. (2011) – MOD16 – *Rem. Sens. Environ.*; Tang et al. (2009) – UW ET – *JGR*; Anderson et al. (2007) – ALEXI – *JGRa*

Evaluation of gridded surface fluxes



Taylor diagram using FLUXNET product as reference dataset.

ALEXI, MOD16, and UW ET compared to FLUXNET are shown.

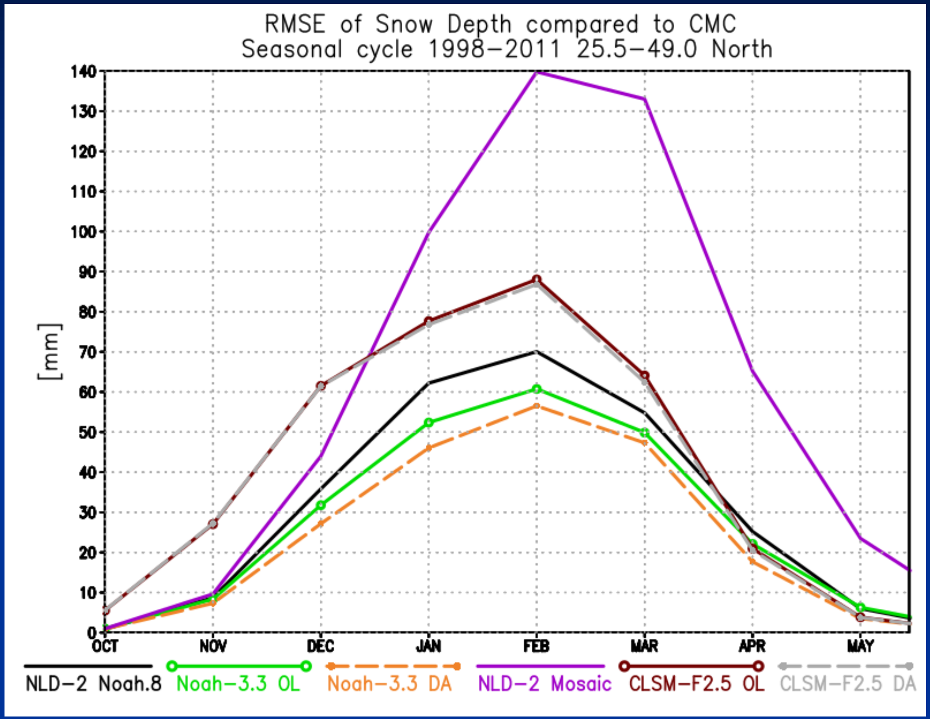
NLDAS-2 LSMs shown with open marks.

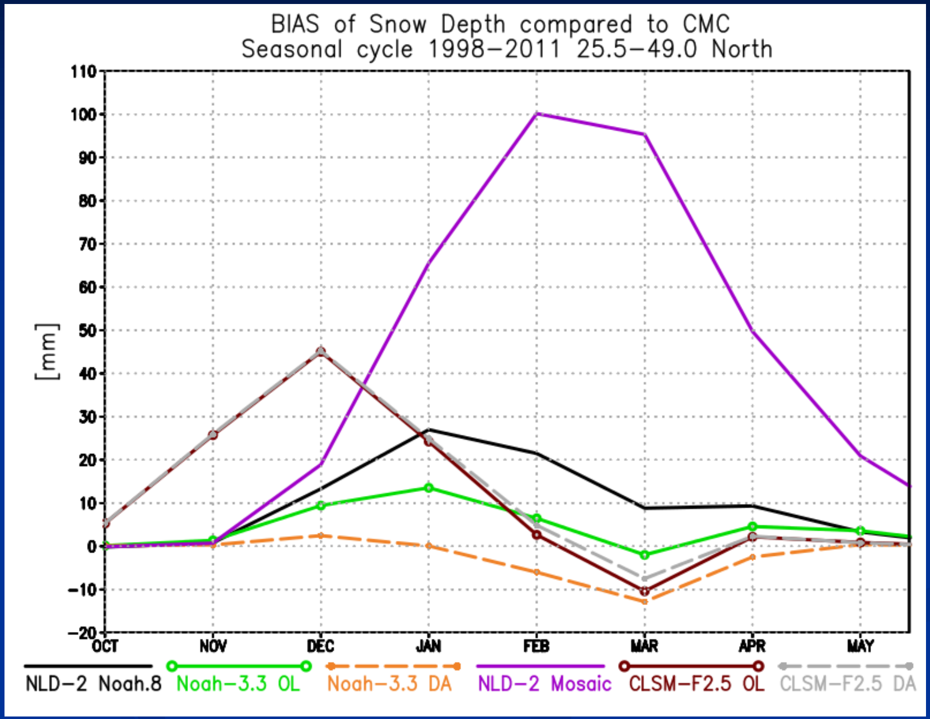
Noah-3.3/CLSM-F2.5 OL runs shown with closed marks.

Noah-3.3/CLSM-F2.5 DA runs shown with "X" and "+".

Reference(s): Jung et al. (2009) – FLUXNET – *Biogeosci.*; Mu et al. (2011) – MOD16 – *Rem. Sens. Environ.*; Tang et al. (2009) – UW ET – *JGR*; Anderson et al. (2007) – ALEXI – *JGRa*

Evaluation of snow depth

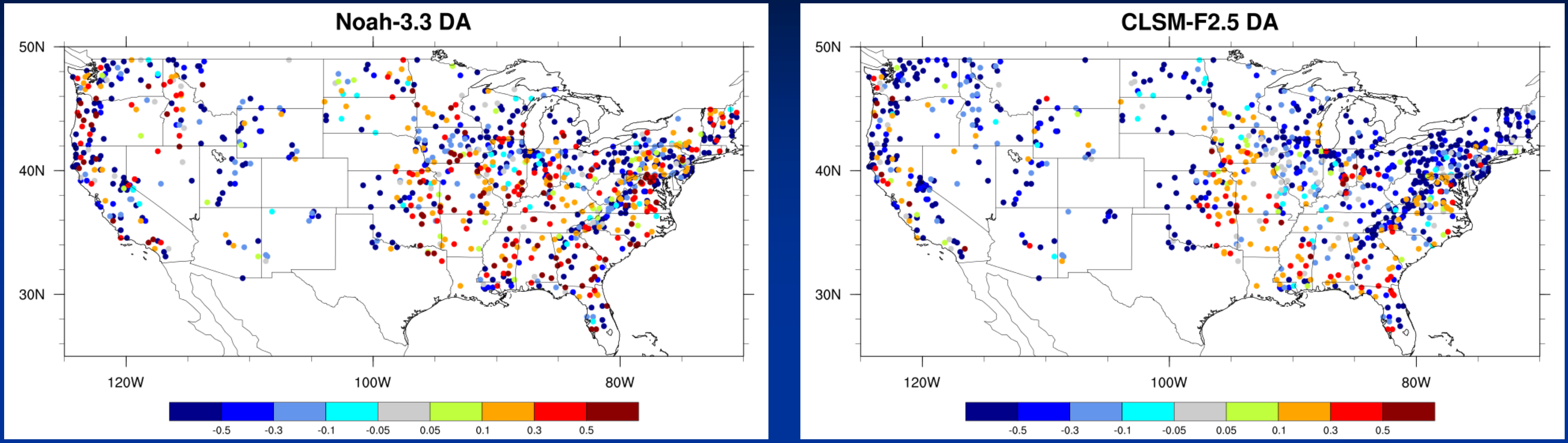




Gridded snow depth analyses (GHCN, CMC, SNODAS) are used for snow depth evaluations. (Left) RMSE & (Right) Bias against the CMC daily snow depth analysis. Additional detail in Kumar et al. (*GRL*, revised version submitted) and Liu et al. (*WRR*, in revision). And Kumar et al.'s presentation on Tue at 11:00am.

Reference(s): Menne et al. (2012) – GHCN – *J. Atmos. Oceanic Technol.*; Brown and Brasnett (2010) – CMC – *Environ. Canada*; Barrett (2009) – SNODAS – *NSIDC, Boulder, CO*

Evaluation of streamflow



Nash-Sutcliffe Efficiency against 939 USGS small unregulated basins, 2000-2012.

(Left) Noah-3.3 DA and (Right) CLSM-F2.5 DA runs compared to USGS streamflow observations. The NLDAS-2 LSMs use the NLDAS router to route surface runoff and baseflow. The new LSMs used the HyMAP router, which also provides streamflow height and includes floodplains. We've also used a large-basin naturalized streamflow dataset (Mahanama et al., *JHM*, 2012).

Reference(s): Lohmann et al. (2004) – NLDAS router and USGS streamflow obs – *JGR*; Getirana et al. (2012) – HyMAP router – *JHM*; Mahanama et al. (2012) – Naturalized streamflow – *JHM*;

Benchmarking system

- David Lawrence (NCAR) et al. have proposed a benchmarking system for land models as part of The International Land Model Benchmarking (ILAMB) project. One aspect of this system is the development of a "score" system using the standard. dev. of the observations to normalize different error metrics. One example:

$$S = 1 - (\text{Global avg. RMSE} / \sigma_{observations})$$

LSM (version) vs. FLUXNET	RMSE (W/m^2)	S (with $\sigma_{obs} = 23.7\ W/m^2$)
NLDAS-2 Mosaic	20.1	0.152
NLDAS-2 VIC-4.0.3	14.4	0.392
NLDAS-2 Noah-2.8	12.3	0.481
Noah-3.3 OL	18.0	0.241
CLSM-F2.5 OL	12.9	0.456

S varies between 0 and 1; higher values show lower overall RMSE

Benchmarking system

- This type of “score” system can be produced for other metrics (unbiased RMSE, Mean Absolute Error, etc.) all available in LVT.
- Lawrence has also proposed adding weights to this score system, subjectively determined to represent confidence/representativeness we have in a particular observational dataset as well as importance of each water budget term.
- As an example, here we use an average weight each of 0.25 for each latent heat flux product to produce a single S “score” for each LSM.

LSM (version)	ALEXI	FLUXNET	MOD16	UW ET	Average
NLDAS-2 Mosaic	0.191	0.152	0.000	0.253	0.149
NLDAS-2 VIC-4.0.3	0.097	0.392	0.090	0.369	0.237
NLDAS-2 Noah-2.8	0.124	0.481	0.161	0.375	0.285
Noah-3.3 OL	0.318	0.241	0.114	0.331	0.251
CLSM-F2.5 OL	0.235	0.456	0.090	0.531	0.328

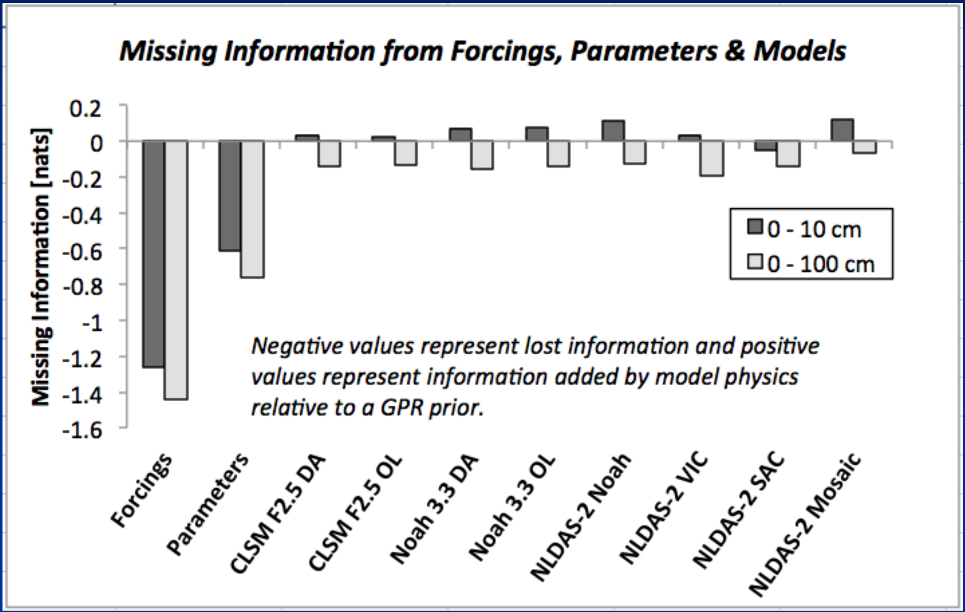
Model inputs vs. Model physics

Information from: **Model Inputs** vs. **Model Physics**

"for the most part, the models *under-utilise* the information available to them" - (Abramowitz, 2005; GRL)

- How to measure information use efficiency?
  - Abramowitz (2005; 2012) showed that regressions on the boundary conditions often out-perform physics-based LSMs. **Implies that models do not use all available information.**
  - Gong et al. (2013) measured information *lost by the model*.
  - Nearing et al. (WRR, "The Amount and Quality of Information in a Model", accepted) measured information *provided by model physics*.
- Use a Leave One Out (LOO) regression model

Benchmarking Results



Standard benchmarking approach (*e.g., Abramowitz, 2005*) using a split data record at each site

- 15-Day lagged forcing as inputs
- Single kernel density function (a Gaussian process regression) trained on ~100 data points at all sites
- 101 SCAN sites with sufficient training data

Explanation of Results

- First columns measure (Shannon) information missing from Forcings
- Second column measures information missing from Parameters
- Model columns measure info gain (or loss) from model physics

Summary/Conclusions

- LSMs from NLDAS Phase 2 and from the next phase of NLDAS are being evaluated against observations using LVT, including soil moisture, surface fluxes, snow, and streamflow.
- Benchmarking system is being developed to evaluate changes from new LSMs and using data assimilation, normalized and weight-averaged between observational datasets and water budget terms.
- Regression model analysis was performed to show the info lost (or gained) from the model physics and parameters.
- Improving model parameters in the new LSMs for NLDAS may provide better simulated results for the Open Loop and for using data assimilation.

Future steps for NLDAS

- Evaluate the VIC-4.1.2.l and SAC-HTET-3.5.6/SNOW-17 Open Loop simulations and add/test the effects of DA.
- Model parameter calibration through the benchmarking system.
- Further evaluations (other variables) with the regression model.
- Drought uncertainty analysis using LIS-OPT/UE subsystem.
- Probabilistic modeling will be performed towards reducing the uncertainty from the LSMs with respect to independent validation data, instead of simply showing improvements in error metrics from a single deterministic realization.
- Add latest versions of Noah-MP and CLM LSMs into LIS (groundwater, etc.) and run and test over the NLDAS domain.
- Transition the latest version of NLDAS using LIS and DA to NOAA/EMC for near-real-time operational data production.

NLDAS & LIS websites

- NLDAS at NASA:
  http://ldas.gsfc.nasa.gov/nldas/
- NLDAS datasets at the NASA GES DISC:
  http://disc.gsfc.nasa.gov/hydrology/
- NLDAS at NOAA/NCEP/EMC:
  http://www.emc.ncep.noaa.gov/mmb/nldas/
- LIS website at NASA:
  http://lis.gsfc.nasa.gov/
- LVT website at NASA:
  http://lis.gsfc.nasa.gov/LVT/

Additional slides

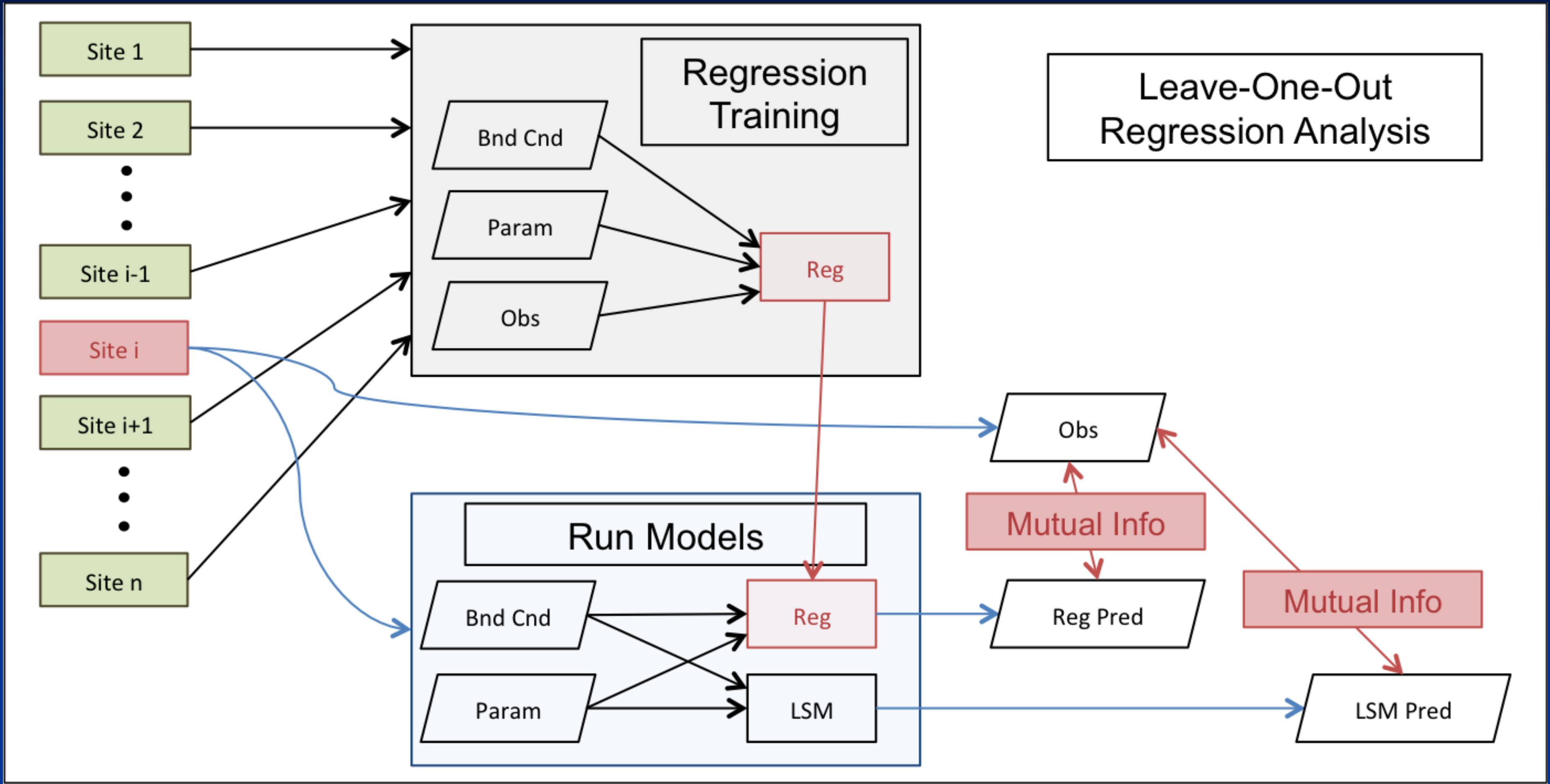
NLDAS Land Surface Models (LSMs)

NLDAS-2	Major LSM changes	Next phase of NLDAS	References
Noah-2.8	• Common code by NCAR/NCEP • Warm season updates • Snow physics upgrade	Noah-3.3	Chen et al. (1996, JGR); Ek et al. (2003, JGR); Wei et al., 2012, HP); Livneh et al., 2010, J. Hydromet.)
Mosaic	• Topographic catchments instead of 1-D soil moisture layers • 3 soil moisture regions: saturated, sub-saturated, and wilting	Catchment/ Fortuna-2.5 (CLSM-F2.5)	Koster et al. (2000, JGR); Reichle et al. (2011, J. Climate); same version of code as for MERRA-Land
VIC-4.0.3	• Canopy energy balance • Snowpack improvements	VIC-4.1.2.1	Liang et al. (1994, JGR); Gao et al. (2010, book chapter)
SAC/ SNOW-17	• Distinct soil layers for soil moisture/temps (HT) • Includes the Noah LSM's evapotranspiration physics (ET)	SAC-HTET-3.5.6/ SNOW-17	Burnash et al., (1973); Anderson (1973); Koren et al. (2007, 2010, NOAA Tech Memos)

Multi-variate Data Assimilation

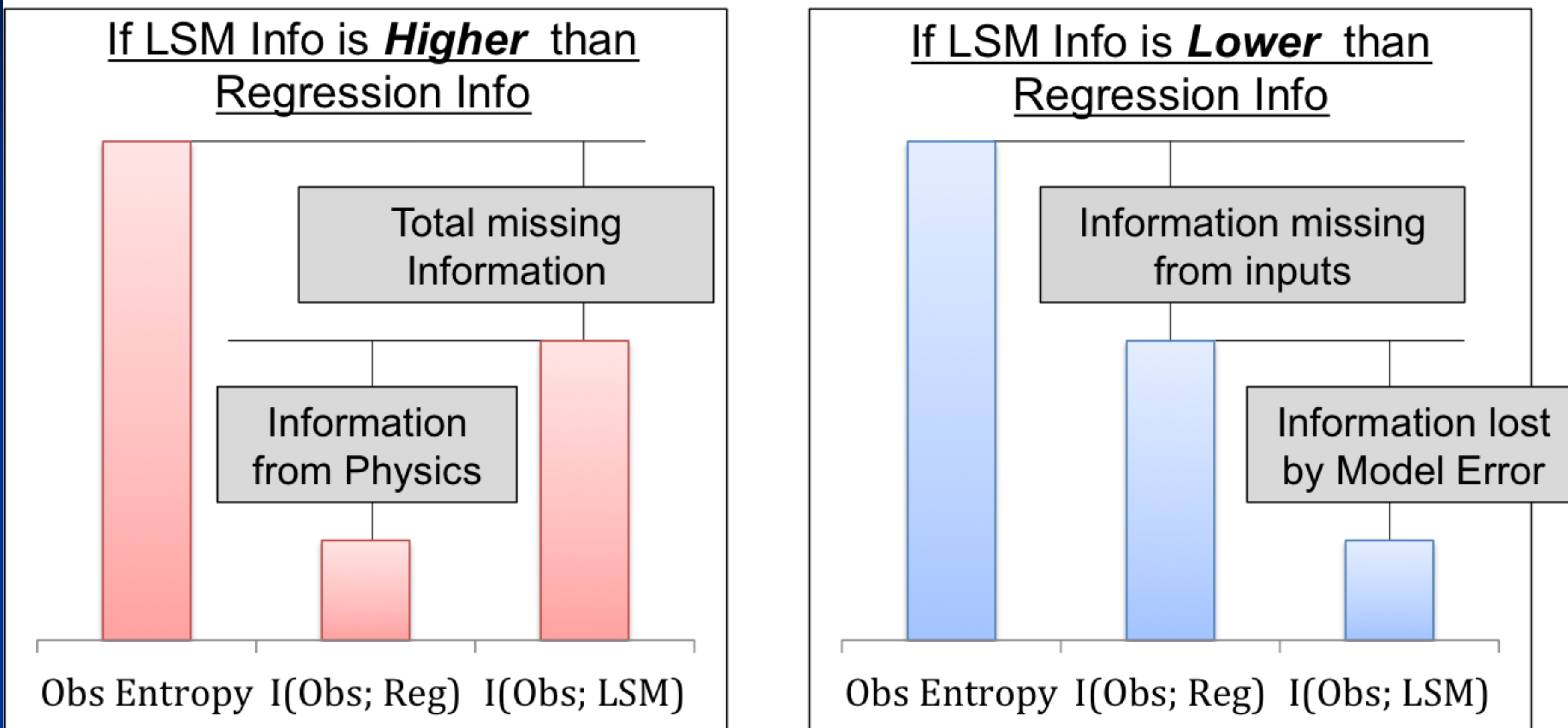
Water state	Platform/Product(s)	Period	Notes
Soil Moisture	• ECV (Essential Climate Variable), merged product from multiple platforms • AMSR-E LPRM product • ASCAT (SMOPS) product	ECV (1979-2002) AMSR-E (2002-2011) ASCAT (2007-)	Microwave soil moisture not assimilated in locations/times with heavy vegetation, precipitation, snow cover, frozen ground, RFI. Observations are scaled to the LSM's climatology using CDF matching.
Snow	• Snow depth from SMMR, SSM/I, AMSE-E • Snow-covered fraction (SCF) from MODIS and IMS	SMMR (1979-1987) SSM/I (1987-2002) AMSR-E (2002-2011) MODIS (2000-) IMS (1997-)	Passive microwave snow depth bias-corrected using the Cressman method using in-situ observations from the Global Historical Climate Network (GHCN). Visible SCF data used as an additional constraint during data assimilation.
TWS	• Terrestrial water storage (TWS) anomalies from GRACE	GRACE (2002-)	Ensemble Kalman Smoother (EnKS) used only with the CLSM-F2.5 LSM.
Irrigation intensity	• Irrigated areas determined from a MODIS climatology	Demand-driven during 1979-2012	Demand-driven "sprinker" scheme based on Ozdogan et al., JHM, 2010. Triggered when root zone soil moisture falls below a specific threshold. Irrigation requirement computed as an amount of water and added to the precipitation forcing.

Regression analysis



With observations at "n" Sites, Site "i" is left out of the training of the regression model, which uses the inputs and observations at each other site. After training, the inputs and observations at Site "i" are used to determine information obtained from the observations, the regression model, and the physics of the land-surface model(s). This is repeated for all "n" Sites.

Regression analysis



The X-axis – I(X;Y) – is Shannon's mutual information between X and Y; it is equal to the amount of information that Y gives about X and vice versa. Mutual information measures **predictability**. If the Info obtained from the LSM is **Higher** than from the regression model, then the physics from the LSM has **provided additional information**. If the Info obtained from the LSM is **Lower** than from the regression model, then the physics from the LSM has **lost information** through model error.

Discussion

- The previous analyses and description of tools are towards the development of a **systematic approach** to LSM benchmarking.
- LSM parameters should be **distributed and optimized** for the validation datasets that we most trust (and ideally, co-located).
- The benchmarking environment should be automated and consider all aspects of the water and energy balances (a particular LSM may improve its simulation of snow but degrade the fluxes).
- The end goal of this work is to identify and correct model and parameter deficiencies towards improved model fluxes/states.