Land Data Assimilation of Satellite-Based Soil Moisture Products Using the Land Information System Over the NLDAS Domain

David M. Mocko[1,2], Sujay V. Kumar[1], Christa D. Peters-Lidard, and Yudong Tian[3]
NASA Goddard Space Flight Center, Hydrological Sciences Laboratory

Poster Number: **24**
Advances in Methods and Applications of Land Data Assimilation Systems
26th Conference on Hydrology

Also: 1 – Science Applications International Corporation (SAIC)
2 – NASA Goddard Space Flight Center Global Modeling and Assimilation Office
3 – Earth System Science Interdisciplinary Center (ESSIC), Univ. of Maryland

Data assimilation with the Land Information System (LIS)

The Land Information System

The Land Information System (LIS) is a software framework for high-performance computer modeling and data assimilation, developed at NASA Goddard's Hydrological Sciences Laboratory. LIS (Kumar et al., 2006; Peters-Lidard et al., 2007) can use many different surface forcing and land parameter datasets to drive a land-surface model (LSM) to produce output of water/energy fluxes, soil moisture/temperature profiles, and other land-surface states. Several different LSMs are also available in the LIS architecture. LIS can also include surface observations via data assimilation techniques to improve the depiction of hydrologic model states. LIS can also be run coupled to the WRF (Weather Research and Forecasting) Model. LIS's domain/resolution are configurable, and supports high-performance computing through advanced parallelization. In addition to data assimilation, LIS also supports various optimization and uncertainty techniques. Figure 1 is a schematic showing data flow using LIS in an uncoupled mode. This figure demonstrates the typical configuration for a Land Data Assimilation System (LDAS) using data assimilation of surface land state information to improve simulated output from an LSM.

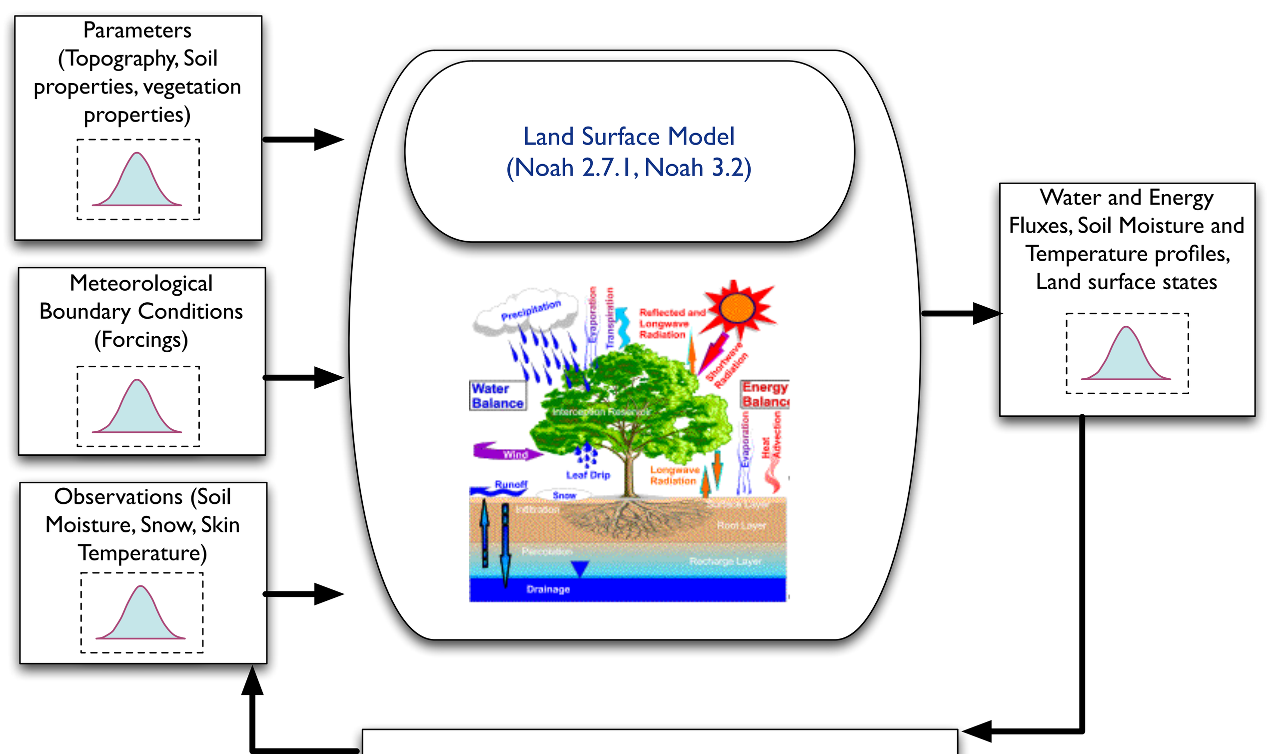

Figure 1: Schematic depicting the data flow in the Land Information System for a uncoupled land data assimilation system. Two versions of the Noah LSM were used in this study, but many additional LSMs are available within the LIS framework.

NLDAS Phase 2 Description and Forcing Data

The North American Land Data Assimilation System (NLDAS) is a collaborative project among several groups: National Centers for Environmental Prediction's (NCEP) Environmental Modeling Center (EMC), NASA's Goddard Space Flight Center (GSFC), Princeton University, the National Weather Service's (NWS) Office of Hydrologic Development (OHD), the University of Washington, and NCEP's Climate Prediction Center (CPC). NLDAS Phase 2 (Xia et al., 2011a; Ek et al., 2011) has produced an hourly 33-year+ dataset (Jan 1979 to present, in near real-time with a ~4-6 day lag) 1/8th-degree surface meteorology and hydrology dataset over the contiguous United States, southern Canada, and northern Mexico. The non-precipitation land-surface forcing fields for NLDAS-2 are derived from the analysis fields of the NCEP North American Regional Reanalysis (NARR). These fields are 32-km spatial resolution and 3-hourly temporal frequency, and are spatially interpolated to the NLDAS grid and then temporally disaggregated to hourly temporal resolution. NCEP/CPC daily gauge precipitation is interpolated with the PRISM (Parameter-elevation Regressions on Independent Slopes Model) algorithm to the NLDAS grid. Daily gauge precipitation is then temporally disaggregated into hourly using Stage II radar precipitation estimates. If the radar is not available, the CMORPH (CPC MORPHing technique) precipitation analysis, CPC HPD (Hourly Precipitation Data), or NARR precipitation are used (depending on data availability/extent). NARR downward shortwave (SW) radiation also is bias-corrected using retrieved GOES (Geostationary Operational Environmental Satellite) data. The NLDAS-2 forcing is used to drive four land-surface models (LSMs) – NCEP/Noah, NASA/Mosaic, Princeton/VIC, and OHD/SAC – to output water/energy fluxes and model state variables. NLDAS datasets are being evaluated against numerous observations, and are used in a wide variety of applications. The forcing and all four models' outputs are staged on NCEP/EMC public ftp servers via their NLDAS website. At the same time, the NASA GES DISC distributes NLDAS forcing and Mosaic model output, with additional search, sub-setting, format, and visualization options available. NCEP/EMC also hosts the NLDAS Drought Monitor with products to support the National Integrated Drought Information System (NIDIS). Table 1 lists the various datasets used for NLDAS Phase 2 forcing, their coverage period, temporal/spatial resolutions, and other information.

Forcing	Coverage	Temporal Resolution	Spatial Resolution	Notes
NARR Model	1979-2003	3 Hourly	32km	
R-CDAS Model	2003-Present	3 Hourly	32km	Realtime version of NARR
GOES Radiation	1996-2000	Hourly	1/8th degree	Used to bias correct NARR
CPC PRISM Gauge	1979-Present	Daily	1/8th degree	Used over CONUS, Mexico
CPC Gauge	1979-Present	Hourly	2 X 2.5 degree	Used over CONUS
CMORPH Precip	2002-Present	1/2 Hourly	8km	Used over CONUS, Mexico
Stage II Precip	1996-Present	Hourly	4km	Used over CONUS

Table 1: Overview of datasets used in NLDAS Phase 2 surface forcing.

Remotely-Sensed Soil Moisture, Data Assimilation, and Experimental Design

Peters-Lidard et al. (2011) configured LIS on the same domain as the 1/8th-degree NLDAS grid to examine two different versions of the Noah LSM (2.7.1 and 3.2) as well as the effects of data assimilation of two different soil moisture datasets from AMSR-E retrievals. The primary focus of this paper was to examine the simulated evapotranspiration (ET) of the different models and to explore possible improvements in ET from the assimilation of soil moisture states. NLDAS Phase 2 forcing is used to drive the two LSMs (in separate simulations) from 1979 to 2010 (after a 15-year model spin-up). The simulations with data assimilation included the AMSR-E surface soil moisture products from 2002-2008 in two separate simulations using the Noah3.2 LSM. The two AMSR-E products used are: 1) the NASA Level 3 "AE_Land" product; and 2) the Land Parameter Retrieval Model (LPRM) product developed at GSFC and VU Amsterdam. Various quality-control measures are taken on these products before data assimilation, including flags for dense vegetation, precipitation, snow cover, frozen soil, and radio frequency interference.

The data assimilation simulations use a 1-D Ensemble Kalman Filter (EnKF) algorithm with an ensemble size of 12. Perturbation parameters (based on earlier data assimilation studies, Kumar et al., 2009) are applied to both the surface forcing (precipitation and surface shortwave & longwave) and to the Noah LSM soil moisture states. The surface (10cm) soil moisture is updated, as well as layers 2 through 4 in Noah using smaller perturbations away from the surface. These perturbations are used to simulate uncertainty in the soil moisture fields. Because an algorithm such as EnKF corrects random zero-mean errors and assume unbiased observations relative to the model background, the observations are scaled (prior to data assimilation) to the model's climatology using a cumulative distribution function (CDF) approach (Reichle and Koster, 2004). The model CDF and the observations CDF are computed over the 2002-2008 period separately for each grid point, and the observations (roughly the top 2cm of soil) are re-scaled to the model's 10cm surface soil moisture climatology.

References: **Ek et al., 2011**: North American Land Data Assimilation System Phase 2 (NLDAS-2): Development and Applications. *GEWEX* News, May 2011; **Jung et al., 2009**: Towards global empirical upscaling of FLUXNET eddy covariance observations: validation of a model tree ensemble approach using a biosphere model. *Biogeosci.*, **6**, 2001–2013; **Kumar et al., 2006**: LIS - An Interoperable Framework for High Resolution Land Surface Modeling. *Env. Model. & Software*, 21, 1402–1415; **Kumar et al., 2009**: Role of subsurface physics in the assimilation of surface soil moisture observations. *J. Hydrometeor.*, doi: 10.1175/2009JHM1134.1; **Kumar et al., 2012**: Land surface Verification Toolkit (LVT) – A generalized framework for land surface model evaluation, submitted to *Geosci. Model. Dev.*; **Lohmann et al., 2004**: Streamflow and water balance intercomparison of four land-surface models in the North American Land Data Assimilation System (NLDAS). *JGR*, 109(D07S91), doi:10.1029/2003JD003517; **Mu et al., 2011**: Improvements to a MODIS Global Terrestrial Evapotranspiration Algorithm. *Rem. Sens. Environ.*, **115**(8): 1781–1800, doi: 10.1016/j.rse.2011.02.019; **Peters-Lidard et al., 2007**: High-performance earth system modeling with NASA/GSFC's Land Information System. *Innov. Sys. & Software Engin.*, 3(3), 157–165; **Peters-Lidard et al., 2011**: Estimating evapotranspiration with land data assimilation systems. *Hydro. Proc.*, doi: 10.1002/hyp.8387; **Reichle and Koster, 2004**: Bias reduction in short records of satellite soil moisture. *GRL*, **31**(L19501): 4, doi: 10.1029/ 2004GL020938; **Reichle et al., 2007**: Comparison and assimilation of global soil moisture retrievals from the Advanced Microwave Scanning Radiometer for the Earth Observing System (AMSR-E) and the Scanning Multichannel Microwave Radiometer (SMMR). *JGR*, **112**(D09108), doi: 10.1029/2006JD008033; **Rodell et al., 2004**: The Global Land Data Assimilations System. *BAMS*, **85**(3), 381-394; **Xia et al., 2011a&b**: Continental-Scale Water and Energy Flux Analysis and Validation for the North-American Land Data Assimilation System Project Phase 2 (NLDAS-2), Part 1: Intercomparison and Application of Model Products. *JGR*, doi:10.1029/2011JD016048; Part 2: Validation of Model-Simulated Streamflow. *JGR*, doi:10.1029/2011JD016051.

Comparison to evapotranspiration, soil moisture, and streamflow data

Evaluation against two independent monthly-averaged evapotranspiration datasets

The simulated fluxes from the simulations were compared against the Global LDAS (GLDAS; Rodell et al., 2004) as well as against two reference ET datasets. The first is a gridded global monthly ½-degree product (Jung et al. 2009) synthesized from FLUXNET tower data. The other is a global 1km dataset (MOD16; Mu et al., 2011) based on MODIS satellite data. The LDAS and other datasets were averaged up to ½-degree for the comparisons. Figure 2 depicts the seasonally-averaged latent heat flux from the LDAS simulations compared to FLUXNET and MOD16 during the period of the available soil moisture products. The NLDAS Noah2.7.1 and Noah3.2 results here do not include the data assimilation. Table 2 provides the RMSE and Bias values (also for 2002-2008) for the various LDAS results as compared to both the FLUXNET and MOD16 datasets. The assimilation of the NASA "AE_Land" product actually increases the RMSE and Bias, while the assimilation of the LRPM product reduces both RMSE and bias. Further details of these results can be found in Peters-Lidard et al. (2011) as well as Kumar's AMS talk Tue 24 Jan 2012 at 2:15pm.

	FLUXNET RMSE	FLUXNET Bias	MOD16 RMSE	MOD16 Bias
Latent Heat Flux				
GLDAS	24.7 ± 0.3	5.5 ± 0.4	28.0 ± 0.2	4.4 ± 0.3
NLDAS-Noah2.7.1	19.3 ± 0.3	11.9 ± 0.4	21.5 ± 0.2	10.3 ± 0.3
NLDAS-Noah3.2	27.6 ± 0.3	12.9 ± 0.4	22.7 ± 0.2	11.2 ± 0.3
NLDAS-Noah3.2 + NASA DA	29.4 ± 0.3	15.9 ± 0.4	24.5 ± 0.2	14.2 ± 0.3
NLDAS-Noah3.2 + LPRM DA	25.6 ± 0.3	10.9 ± 0.3	21.9 ± 0.2	9.2 ± 0.3
Sensible Heat Flux				
GLDAS	23.4 ± 0.2	−5.6 ± 0.4	N/A	N/A
NLDAS-Noah2.7.1	21.1 ± 0.3	−7.0 ± 0.4	N/A	N/A
NLDAS-Noah3.2	32.5 ± 0.3	−9.2 ± 0.4	N/A	N/A
NLDAS-Noah3.2 + NASA DA	34.5 ± 0.3	−12.2 ± 0.4	N/A	N/A
NLDAS-Noah3.2 + LPRM DA	30.4 ± 0.3	−7.3 ± 0.4	N/A	N/A

Table 2: Domain-averaged RMS and Bias values in W/m^2 (with 95% confidence intervals) of latent and sensible heat fluxes from five simulations as compared to reference datasets of FLUXNET and MOD16. Sensible heat not available in MOD16.

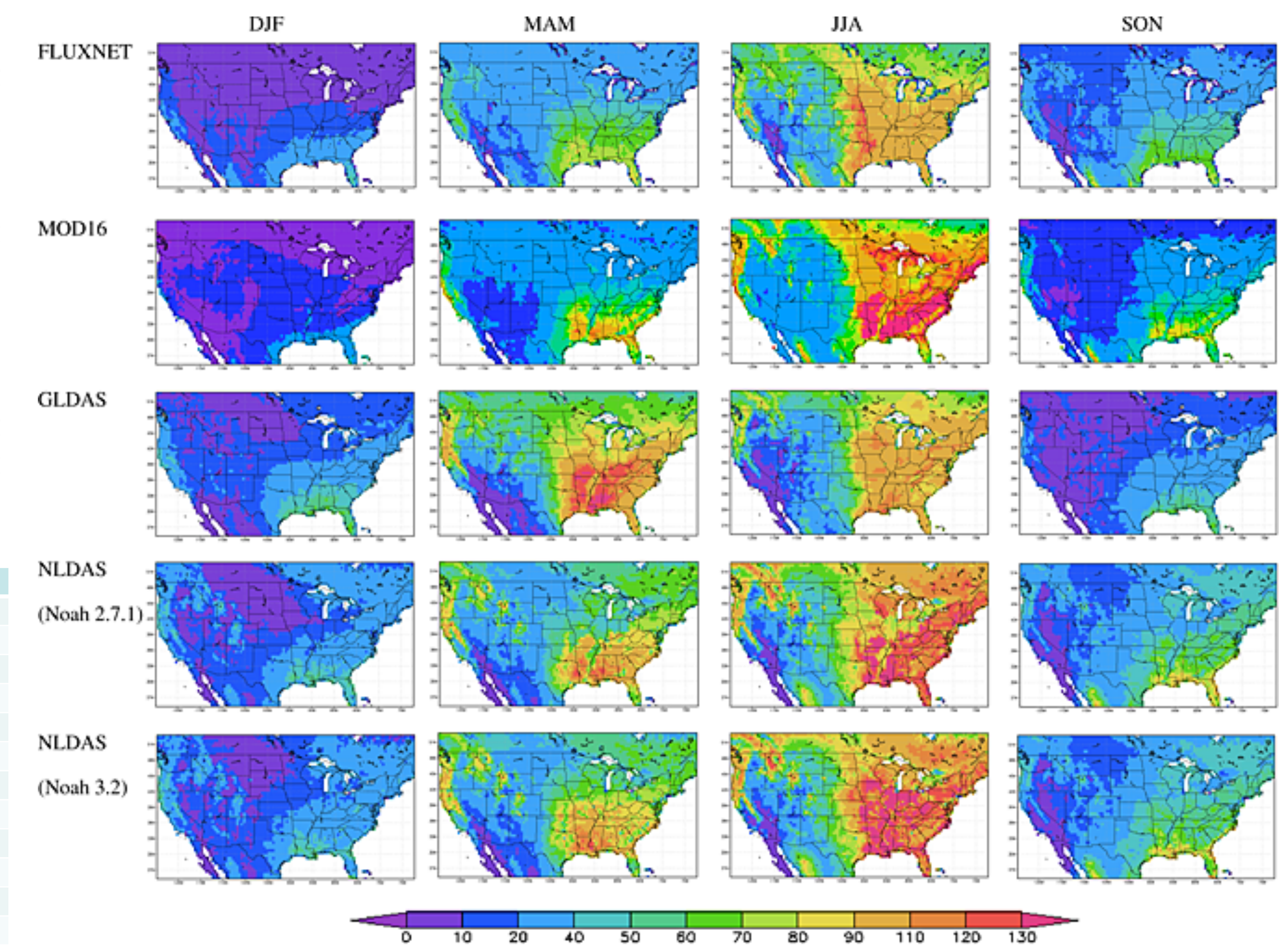

Figure 2: Seasonally-averaged latent heat flux for Jan 2002 – Dec 2008 for FLUXNET, MOD16, GLDAS Noah, and NLDAS-forced simulations with Noah2.7.1 and Noah3.2, without soil moisture data assimilation. Units are in W/m^2.

Evaluation against in situ soil moisture

The NLDAS-forced Noah3.2 simulations in LIS were also compared against in situ soil moisture observations from 2002-2009. The observations used were from the USDA's Soil Climate Analysis Network (SCAN); the distribution of the sites used is shown in Figure 3. A preliminary analysis (Table 3) of the Open Loop simulation (with no soil moisture data assimilation) and of the NASA-DA simulation and LPRM-DA simulation again showed that the NASA "AE_Land" product degraded in comparison to the reference dataset relative to the Open Loop. The assimilation of the LPRM product slightly improved the soil moisture at these sites for both the surface and root zone. The comparison was repeated for a subset of the total number of sites that were quality-controlled after Reichle et al. (2007). Again, the NASA-DA simulation resulted in a lower anomaly correlation of soil moisture as compared to SCAN, with a slight improvement in the LPRM-DA simulation, especially in the root zone. Further analysis will continue to compare against other soil moisture and observational datasets using the LIS Verification Toolkit (Kumar et al., 2012).

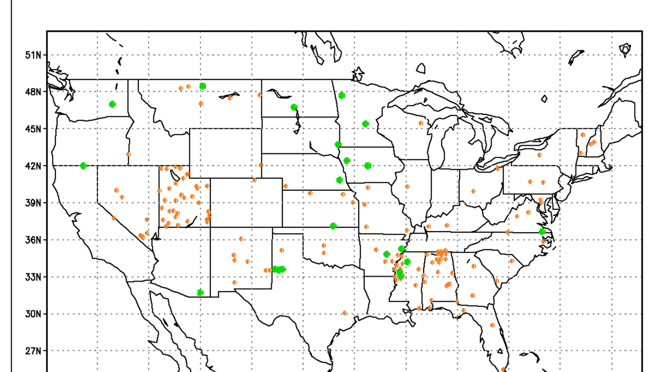

Figure 3: SCAN soil moisture locations. 21 quality-controlled sites shown in green with 158 additional non-quality-controlled sites shown in orange.

Anomaly correlation	OL	NASA-DA	LPRM-DA
Surface (10cm)	0.55 +/-0.01	0.49 +/-0.01	0.56 +/-0.01
Root zone (1m)	0.17 +/-0.01	0.13 +/-0.01	0.19 +/-0.01

Anomaly correlation	OL	NASA-DA	LPRM-DA
Surface (10cm)	0.62 +/-0.05	0.53 +/-0.05	0.62 +/-0.05
Root zone (1m)	0.16 +/-0.05	0.13 +/-0.05	0.19 +/-0.05

Table 3: (Top) Anomaly correlation for all 179 SCAN sites from the Open Loop (OL) and NASA-DA and LPRM-DA; (Bottom) Same, but for only 21 quality-controlled sites after Reichle et al. (2007).

Evaluation against observed streamflow

The runoff from the Noah3.2 simulations were routed into streamflow following Lohmann et al. (2004). USGS data for major water resource basins (Figure 4) were used for comparison using stations detailed in Xia et al. (2011b). The Open Loop and LPRM-DA RMSE and Bias values are shown in Figure 5, and the seasonal cycle for a select 4 basins are shown in Figure 6. For most basins, data assimilation of the LPRM product improved the simulation of streamflow.

Figure 4: Major water resource basins in the United States.

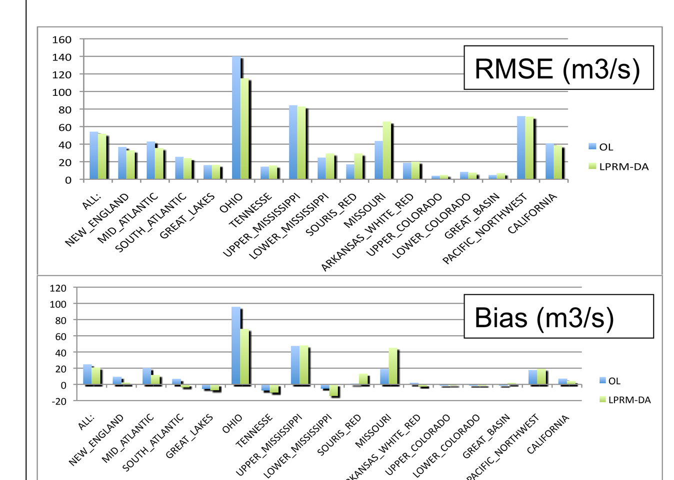

Figure 5: (Top) RMSE & (bottom) Bias for Open Loop and LPRM-DA for major basins in (m3/sec) after Xia et al. (2011b).

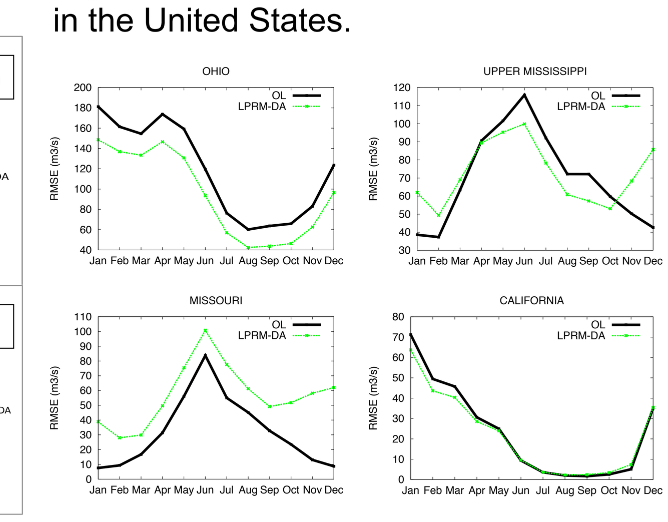

Figure 6: The RMSE (m3/s) in 4 water resource basins comparing the Open Loop against the LPRM-DA simulation.

This research is supported by funding in part by the NOAA Climate Program Office MAPP program. NLDAS documentation and datasets can be found at http://ldas.gsfc.nasa.gov/ and http://www.emc.ncep.noaa.gov/mmb/nldas/. The efforts of all NLDAS/GLDAS participants are acknowledged.