JOURNAL OF GEOPHYSICAL RESEARCH: ATMOSPHERES, VOL. 119, 23–38, doi:10.1002/2013JD020792, 2014

Hydrological evaluation of the Noah-MP land surface model for the Mississippi River Basin

Xitian Cai,[1] Zong-Liang Yang,[1] Cédric H. David,[1,2] Guo-Yue Niu,[3] and Matthew Rodell[4]

Received 26 August 2013; revised 3 December 2013; accepted 6 December 2013; published 9 January 2014.

[1] This study evaluates regional-scale hydrological simulations of the newly developed community Noah land surface model (LSM) with multiparameterization options (Noah-MP). The model is configured for the Mississippi River Basin and driven by the North American Land Data Assimilation System Phase 2 atmospheric forcing at 1/8° resolution. The simulations are compared with various observational data sets, including the U.S. Geological Survey streamflow and groundwater data, the AmeriFlux tower micrometeorological evapotranspiration (ET) measurements, the Soil Climate Analysis Network (SCAN)-observed soil moisture data, and the Gravity Recovery and Climate Experiment satellite-derived terrestrial water storage (TWS) anomaly data. Compared with these observations and to the baseline Noah LSM simulations, Noah-MP shows significant improvement in hydrological modeling for major hydrological variables (runoff, groundwater, ET, soil moisture, and TWS), which is very likely due to the incorporation of some major improvements into Noah-MP, particularly an unconfined aquifer storage layer for groundwater dynamics and an interactive vegetation canopy for dynamic leaf phenology. Noah-MP produces soil moisture values consistent with the SCAN observations for the top two soil layers (0–10 cm and 10–40 cm), indicating its great potential to be used in studying land-atmosphere coupling. In addition, the simulated groundwater spatial patterns are comparable to observations; however, the inclusion of groundwater in Noah-MP requires a longer spin-up time (34 years for the entire study domain). Runoff simulation is highly sensitive to three parameters: the surface dryness factor (α), the saturated hydraulic conductivity (k), and the saturated soil moisture (θ_{max}).

Citation: Cai, X., Z.-L. Yang, C. H. David, G.-Y. Niu, and M. Rodell (2014), Hydrological evaluation of the Noah-MP land surface model for the Mississippi River Basin, *J. Geophys. Res. Atmos.*, *119*, 23–38, doi:10.1002/2013JD020792.

1. Introduction

[2] Land surface models (LSMs) have evolved rapidly in recent decades due to the advances in high-performance computing, ground-based measurements (e.g., FLUXNET [*Baldocchi et al.*, 2001]), remote sensing [*Murray et al.*, 2013], and emerging concepts such as hyperresolution [*Wood et al.*, 2011] and multiparameterization (or multiple hypotheses) [*Clark et al.*, 2011]. One such LSM is the community Noah LSM with multiparameterization options (hereafter Noah-MP) [*Niu et al.*, 2011; *Yang et al.*, 2011]. Based on the Noah LSM [*Ek et al.*, 2003], Noah-MP has added biophysical processes such as an unconfined aquifer for groundwater storage and a dynamic water table [*Niu et al.*, 2007], an interactive vegetation canopy [*Dickinson et al.*, 1998], a multilayer snowpack [*Yang and Niu*, 2003], and a simple TOPMODEL (TOPography based hydrological MODEL)-based runoff production [*Niu et al.*, 2005].

[3] Model evaluation plays a very important role in LSM development, as LSM benchmarking or better model evaluation is one of the three core activities in the current Global Energy and Water Cycle Exchanges Project (GEWEX) Global Land/Atmosphere System Study (GLASS) [*van den Hurk et al.*, 2011]. Noah-MP has been tested at local scales [*Niu et al.*, 2011] and in global river basins [*Yang et al.*, 2011]. Its runoff simulation was evaluated using the University of New Hampshire-Global Runoff Data Center (UNH-GRDC) gridded runoff data set [*Fekete et al.*, 2002], but it has not yet been evaluated with the U.S. Geological Survey (USGS) streamflow data. The groundwater module was evaluated against the Gravity Recovery and Climate Experiment (GRACE) terrestrial water storage (TWS) data when it was coupled with the Community Land Model [*Niu et al.*, 2007], but it has not been evaluated since its coupling with Noah-MP. In addition, the evapotranspiration (ET) simulation has not yet been evaluated using observational

[1]Department of Geological Sciences, John A. and Katherine G. Jackson School of Geosciences, University of Texas at Austin, Austin, Texas, USA.
[2]Now at University of California Center for Hydrologic Modeling, Irvine, California, USA.
[3]Department of Hydrology and Water Resources, University of Arizona, Tucson, Arizona, USA.
[4]Hydrological Sciences Branch, NASA Goddard Space Flight Center, Greenbelt, Maryland, USA.

Corresponding author: Z.-L. Yang, Department of Geological Sciences, John A. and Katherine G. Jackson School of Geosciences, University of Texas at Austin, 1 University Station C1100, Austin, TX 78712, USA. (liang@jsg.utexas.edu)

2169-897X/14/10.1002/2013JD020792

data. Soil moisture is another important variable for land-atmosphere coupling and drought monitoring and thus needs to be evaluated using observational data. Lastly, since the launch of GRACE in 2002, modeled TWS can now be evaluated at a regional scale with the GRACE-derived TWS.

[4] Therefore, this study comprehensively evaluates the performance of Noah-MP in hydrological simulations of major hydrological variables (runoff, groundwater, ET, soil moisture, and TWS). It features a detailed multivariable evaluation using the best available ground-based and satellite measurements. This type of evaluation is consistent with the current call for benchmarking of LSMs by GEWEX GLASS [*Abramowitz*, 2012; *Kumar et al*., 2012; *van den Hurk et al*., 2011]. Specifically, the following measurements are used: USGS streamflow and groundwater data, AmeriFlux tower micrometeorological ET data, Soil Climate Analysis Network (SCAN) soil moisture data, and GRACE satellite-derived TWS anomaly data.

[5] Before a clean version of the model is obtained for evaluation, hydrological modeling generally requires spin-up, parameter sensitivity tests, and model calibration for specific study areas. Without groundwater dynamics, hydrological models typically only require several years of model spin-up time [*Cosgrove et al*., 2003]; with groundwater dynamics, however, they require much longer time for water table depth (WTD) to reach an equilibrium state, particularly in arid regions [*Niu et al*., 2007]. This study will first investigate how the integration of groundwater dynamics into Noah-MP affects model spin-up. Unlike some parameters (e.g., slope and leaf area index) that can be derived from representative field site measurements or remote-sensing data, many parameters, such as the hydraulic conductivity and the Clapp-Hornberger "b" parameter, cannot be directly derived from measurements and must be estimated by calibration for the specific study areas [*Hogue et al*., 2006]. This study will identify the most sensitive parameters through sensitivity tests and then obtain the optimal combination of these parameters through calibration.

[6] The evaluation of the study is conducted for the period of 2000–2009 for the Mississippi River Basin (MRB) at the North American Land Data Assimilation System Phase 2 (NLDAS-2)'s 1/8° resolution [*Ek et al*., 2011]. Section 2 describes the Noah-MP model, study area, and data sets used in this study. Section 3 introduces model spin-up, parameter sensitivity tests, and model calibration. Section 4 shows the specific evaluations of runoff, groundwater, ET, soil moisture, and TWS. Section 5 summarizes the study.

2. Methodology

2.1. The Noah-MP Model

[7] Noah-MP was enhanced from the original Noah LSM through the addition of improved physics and multiparameterization options [*Niu et al*., 2011; *Yang et al*., 2011]. The improved physics includes a dynamic groundwater component, an interactive vegetation canopy, and a multilayer snowpack. The multiparameterization options provide users with multiple choices of parameterizations in leaf dynamics, canopy stomatal resistance, soil moisture factor for stomatal resistance, and runoff and groundwater. For example, there are four options for runoff and groundwater: (1) TOPMODEL with simple groundwater model (SIMGM) [*Niu et al*., 2007], (2) TOPMODEL with an equilibrium water table [*Niu et al*., 2005], (3) original surface and subsurface runoff (free drainage) [*Schaake et al*., 1996], and (4) Biosphere-Atmosphere Transfer Scheme surface and subsurface runoff (free drainage) [*Yang and Dickinson*, 1996]. The parameterizations used in this study are the default options recommended by *Yang et al*. [2011]: TOPMODEL runoff with SIMGM groundwater, leaf dynamics, Ball-Berry canopy stomatal resistance, a Noah-type (using soil moisture) soil moisture factor controlling stomatal resistance, and the Monin-Obukhov surface exchange coefficient for heat.

[8] Both surface and subsurface runoff are computed by a simple TOPMODEL-based runoff model [*Niu et al*., 2005]. Surface runoff (R_{sf}) is parameterized as

$$R_{sf} = F_{sat}\,p + (1 - F_{sat})\max(0, (p - I)) \tag{1}$$

where p, the effective precipitation intensity (kg m^{-2} s^{-1}), is the rainfall and dewfall reaching the ground plus snowmelt, I is maximum soil infiltration capacity (kg m^{-2} s^{-1}), which is dependent on soil properties and moisture, and F_{sat} is the fraction of saturated area and is parameterized as

$$F_{sat} = (1 - F_{frz})\,F_{max}\,e^{-0.5f\left(z_\nabla - z'_{bot}\right)} + F_{frz} \tag{2}$$

where F_{frz} is a fractional impermeable area as a function of the soil ice content of the surface soil layer, z_∇ is the WTD (m), z'_{bot} is the depth of the model bottom, which is 2 m, and F_{max} is the potential or maximum saturated fraction for a grid cell, which can be derived from high-resolution subgrid topography (e.g., 30 m) of a model grid cell (e.g., 1° resolution) using the TOPMODEL concepts (see *Niu et al*. [2005] or *Niu et al*. [2011] for details). In this study, a global mean $F_{max} = 0.38$ is used, which is derived from the HYDRO1K 1 km topographic index (or wetness index, WI) data [*Verdin and Jenson*, 1996].

[9] Subsurface runoff (R_{sb}) is parameterized as

$$R_{sb} = R_{sb,max}\,e^{-\Lambda - f\left(z_\nabla - z'_{bot}\right)} \tag{3}$$

where $R_{sb,max}$ is the maximum subsurface runoff when the grid cell mean WTD is zero—here globally $R_{sb,max} = 5.0 \times 10^{-4}$ m s^{-1}, derived from calibration against global runoff data through sensitivity tests [*Niu et al*., 2007]—and Λ is the grid cell mean WI—here $\Lambda = 10.46$, which is its global mean value derived from HYDRO1K 1 km WI data.

[10] With an unconfined aquifer added to account for the exchange of water between the soil and the aquifer, the temporal variation of the water stored in the unconfined aquifer, W_a (mm), is parameterized as

$$\frac{dW_a}{dt} = Q - R_{sb} \tag{4}$$

where Q is the recharge rate (mm s^{-1}), which is positive when water enters the aquifer. It is parameterized as

$$Q = -K_{bot}\frac{-z_\nabla - (f_{mic}\psi_{bot} - z_{bot})}{z_\nabla - z_{bot}} \tag{5}$$

where K_{bot} is hydraulic conductivity of the bottom soil layer (mm s^{-1}). The f_{mic} is the fraction of micropore content in the bottom layer soil, which is introduced to limit the upward flow (depending on the level of structural soil) and ranges from 0.0 to 1.0 (see *Niu et al.* [2011] for details), and ψ_{bot} is the matric potential (mm).

[11] Latent heat flux (λE, or potential evapotranspiration E) is calculated using the Penman-Monteith equation following *Bonan* [2008]:

$$\lambda E = \frac{s(R_n - G) + \rho C_p (e_*[T_a] - e_a)/r_H}{s + \gamma (r_W/r_H)} \tag{6}$$

where λ is the latent heat of vaporization (J kg^{-1}), $e_*[T_a]$ is the saturation vapor pressure evaluated at the air temperature (T_a), $s = de_*[T_a]/dT$ is the saturation vapor pressure versus temperature evaluated at T_a, R_n is net radiation (W m^{-2}), G is soil heat flux (W m^{-2}), ($R_n - G$) is net available radiation (W m^{-2}), ρ is dry air density (kg m^{-3}), C_p is specific heat capacity of air (J kg^{-1} K^{-1}), e_a is the vapor pressure of air (Pa), and r_H and r_W are resistance to sensible heat and water vapor, respectively (s m^{-1}). The surface exchange coefficient for heat, C_H, which is used to calculate aerodynamic resistances, can be estimated using either the Monin-Obukhov similarity theory (this study) or the method of *Chen et al.* [1997].

[12] In addition to hydraulic conductivity, runoff is also found to be very sensitive to the surface dryness factor (α). It determines the soil surface resistance to ground evapotranspiration [*Sellers et al.*, 1992], as shown in the following equation:

$$r_{surf} = f_{snow} \times 1.0 + (1 - f_{snow}) e^{(8.206 - \alpha S_1)} \tag{7}$$

where r_{surf} is the soil surface resistance (s m^{-1}), f_{snow} is the snow fraction covering a ground surface, and S_1 is the soil wetness in the top soil layer, varying from 0.0 to 1.0. Thus, α controls the effect of soil moisture on r_{surf}.

[13] In the current Noah-MP, vegetation plays a significant role in the model: the stomatal conductance determines the photosynthesis and the carbon cycle, the dynamic leaf model predicts the leaf area index (LAI) and the green vegetation fraction (GVF), the "semitile" subgrid scheme calculates the surface energy balance for vegetated and bare ground separately, and the canopy water scheme simulates the canopy water interception and evaporation.

2.2. Study Area

[14] The Mississippi River Basin (MRB) is the largest river basin in North America, covering many distinct climate zones. It is also a well-studied river basin, and thus, a variety of meteorological, hydrological, and ecological data are available. For example, it is the study area of the first Continental-Scale Experiment of the World Climate Research Program GEWEX Continental-Scale International Project [*Kumar and Merwade*, 2011; *Roads et al.*, 2003]. The MRB area is 3.28 million km^2, which is approximately 41% of the conterminous U.S. (Figure 1a). It covers six of the 21 major geographic regions defined by the USGS two-digit hydrologic unit code (HUC2, http://water.usgs.gov/GIS/huc.html). Calculated from the NLDAS-2 meteorological forcing data (1998–2009), the basin average annual temperature and precipitation are 11.9°C and 821.0 mm, respectively. Across the various climate zones, there is a large temperature gradient between the south and north, with a minimum of −3°C in the Rocky Mountains and a maximum of 22.9°C in the southern most area of the basin (Figure 1b), and a large precipitation gradient between the southeast and northwest, with a minimum of 126.3 mm in Wyoming and a maximum of 1973.6 mm in the Gulf of Mexico region (Figure 1c). In this study, the Ohio-Tennessee Region is considered a typical wet region and Missouri Region a typical dry region, with the Upper Mississippi Region considered the transitional zone between the two.

2.3. Model Input Data

[15] The NLDAS-2 [*Mitchell et al.*, 2004] meteorological forcings at 0.125° spatial resolution and hourly temporal resolution are used to drive the Noah-MP model. The seven nonprecipitation meteorological forcing fields are derived from the NCEP (National Centers for Environmental Prediction) North American Regional Reanalysis, including air temperature, the U and V components of wind speed, specific humidity, surface pressure, surface downward shortwave radiation, and surface downward longwave radiation. Precipitation field data are derived from the temporal disaggregation of the gaged daily precipitation data from NCEP/Climate Prediction Center with an orographic adjustment based on the monthly climatological precipitation of the Parameter-elevation Regressions on Independent Slopes Model [*Daly et al.*, 1994]. The Noah LSM outputs forced by the same NLDAS-2 meteorological forcings are also downloaded from the NLDAS website, which serves as the baseline model for comparison with Noah-MP. More details regarding the setup and performance of the Noah LSM model can be found in *Xia et al.* [2012a, 2012b].

[16] The static input data for Noah-MP are from various sources. The land-water mask that masks out the water component from simulation (land = 1 and water = 0) and the latitude and longitude coordinate information, which are primarily used for computing the solar zenith angle, are the same as those of NLDAS-2, which uses the latitude and longitude in the center of each 0.125° grid box. The vegetation type and soil texture types (top 30 cm and 30–100 cm depth) are aggregated from the 30 arc-second data of the USGS 24-category vegetation (land use) and the hybrid State Soil Geographic Database (STATSGO) Food and Agriculture Organization soil texture data sets, respectively, both of which are maintained by the NCAR/RAL (Research Application Laboratory, National Center for Atmospheric Research) (http://www.ral.ucar.edu/research/land/technology/lsm.php). Soil color data are used to determine ground surface albedo over visible and infrared bands and include eight categories, with one as the lightest and eight as the darkest. The annual mean 2 m air temperature data (from NCAR/RAL) are also used as the lower boundary layer condition for soil temperature. The monthly climatological GVF data are converted from the 0.144° five year (1985–1990) GVF data derived from National Oceanic and Atmospheric Administration (NOAA)/advanced very high resolution diameter (AVHRR) by *Gutman and Ignatov* [1998].

2.4. Observational Data

[17] USGS streamflow data are used for runoff calibration and validation. As shown in Figure 1a, four USGS gaging stations are selected: the Ohio River at Metropolis, IL collecting runoff from region 5 and region 6; the Mississippi River at Keokuk, IA collecting runoff from region 7; the Missouri River at Hermann, MO collecting runoff from region

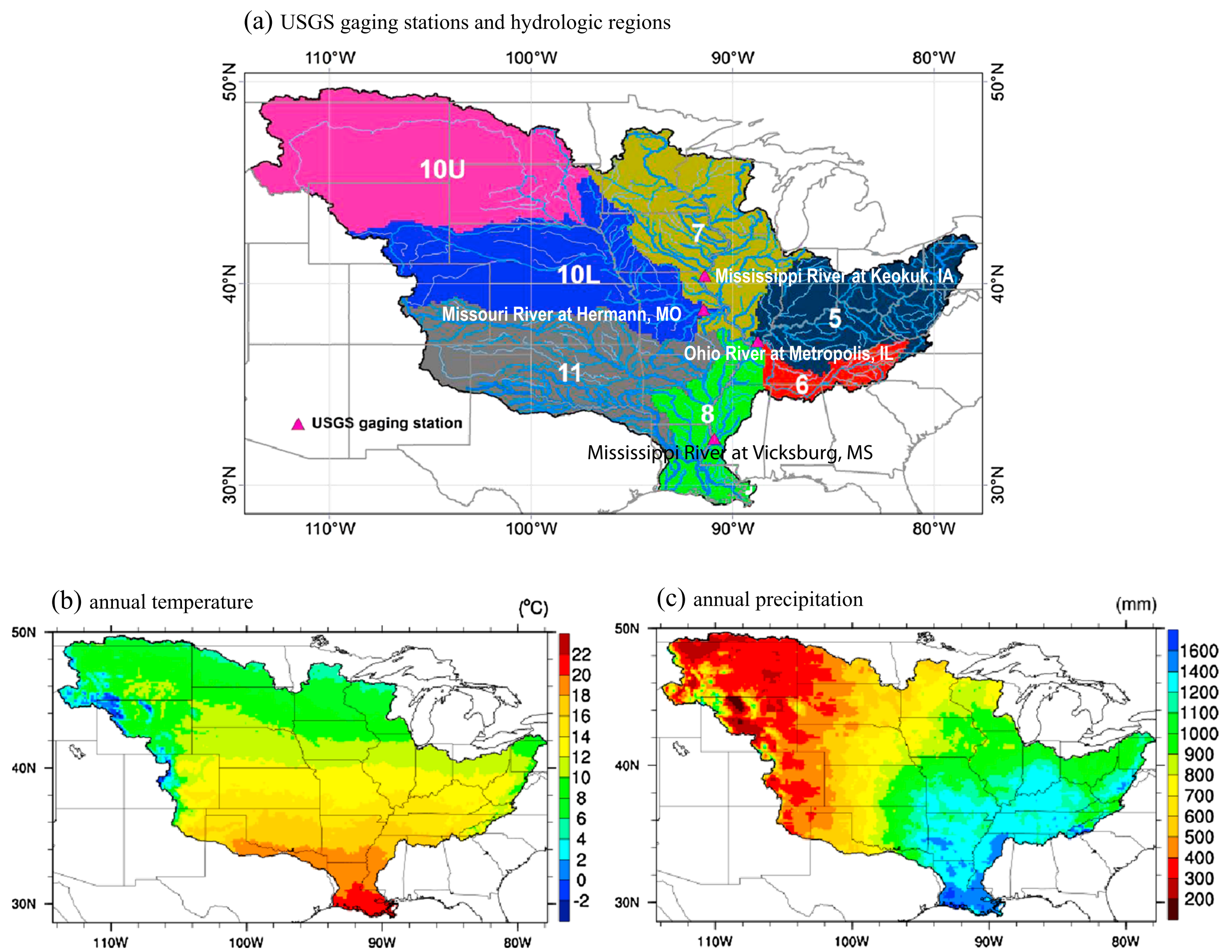

Figure 1. Map of the Mississippi River Basin showing (a) USGS gaging stations and hydrologic regions (Numbers in the shaded area are the two-digit hydrologic unit code (HUC): 5 – Ohio; 6 – Tennessee; 7 – Upper Mississippi; 8 – Lower Mississippi; 10U – Upper Missouri; 10L – Lower Missouri; 11 – Arkansas–White–Red.), (b) average annual temperature and (c) average annual precipitation.

10; and the Mississippi River at Vicksburg, MS collecting runoff from the entire MRB. To compare the spatial distribution of runoff between the model and the observations, we use two types of data sets: the USGS hydrologic unit runoff (http://waterwatch.usgs.gov/) [*Brakebill et al.*, 2011] and the monthly gridded climatological runoff composite fields at 30 min spatial resolution provided by the University of New Hampshire-Global Runoff Data Center (UNH-GRDC). The USGS hydrologic unit runoff data set was developed in 2008 and has been updated annually (D. Wolock, personal communication, 2012). It was calculated from all the available records for 1901–2009 at the eight-digit hydrologic unit code (HUC8) level, which consists of 2110 hydrologic cataloging units for the continental U.S. and 1128 units for the MRB. The UNH-GRDC runoff preserves the accuracy of the observed discharge data and obtains consistent spatial and temporal resolutions from a water balance model [*Fekete et al.*, 2002]; hence, it is considered the best available gridded data set for LSM evaluation, although the values are occasionally lower than the gaged discharge data [*Leung et al.*, 2011].

[18] A climatological WTD map for the U.S. created by *Fan and Miguez-Macho* [2011], which contains 567,946 USGS groundwater observational sites (254,464 sites for the MRB) and is dense enough to show the groundwater spatial pattern for most of the U.S., is used for spatial comparison with the Noah-MP simulated WTD. Daily groundwater storage anomalies for the MRB and its major subbasins were derived from 58 sites with good representation of the subbasin averages by *Rodell et al.* [2007] and are used in this study for temporal comparison with the model results.

[19] Over recent decades, the global network of micrometeorological tower sites with coordinating eddy covariance measurements of CO_2, water vapor, and energy, the FLUXNET (http://fluxnet.ornl.gov/) [*Baldocchi et al.*, 2001; *Running et al.*, 1999], has provided the most reliable ET measurements and has been considered a valuable data source for LSM development [*Stöckli et al.*, 2008] and evaluation [*Blyth et al.*, 2010; *Kumar and Merwade*, 2011; *Li et al.*, 2011]. As part of the FLUXNET, AmeriFlux features much denser flux tower sites in the U.S. than other regional networks. Although there are 21 AmeriFlux tower sites in the MRB, only 15 of them have overlapping observation times with the Noah-MP simulation period (2000–2009), and those sites are used to evaluate the model-simulated latent heat flux (ET). A list of Ameriflux tower sites and their locations, land cover, and available measurement periods are shown in Table 1. The data included in this study are the monthly Level 4 latent heat flux

Table 1. Basic Information Regarding the AmeriFlux Sites

Site No.	Site Name	State	Latitude (°N)	Longitude (°E)	Land Cover[a]	Available Period
1	ARM Southern Great Plains burn site - Lamont	OK	35.55	−98.04	GRA	2005–2006
2	ARM Southern Great Plains control site - Lamont	OK	35.54	−98.04	GRA	2005–2006
3	ARM Southern Great Plains site - Lamont	OK	36.61	−97.49	CRO (DCP)	2003–2006
4	Brookings	SD	44.35	−96.84	GRA	2004–2006
5	Bondville	IL	40.01	−88.29	CRO (DCP)	1996–2007
6	Fort Peck	MT	48.31	−105.10	GRA	2000–2006
7	Goodwin Creek	MS	34.25	−89.97	GRA	2002–2006
8	Lost Creek	WI	46.08	−89.98	CSH (SHR)	2001–2005
9	Morgan Monroe State Forest	IN	39.32	−86.41	DBF	1999–2006
10	Missouri Ozark Site	MO	38.74	−92.20	DBF	2004–2007
11	Mead – irrigated continuous maize site	NE	41.17	−96.48	CRO (DCP)	2001–2006
12	Mead – irrigated maize-soybean rotation site	NE	41.16	−96.47	CRO (DCP)	2001–2006
13	Mead – rain-fed maize-soybean rotation site	NE	41.18	−96.44	CRO (DCP)	2001–2006
14	Niwot Ridge Forest (LTER NWT1)	CO	40.03	−105.55	ENF	1998–2007
15	Willow Creek	WI	45.81	−90.08	DBF	1999–2006

[a]AmeriFlux uses IGBP land cover classification, while Noah-MP uses USGS global 24-category classification. GRA stands for Grassland; CRO: Cropland, DCP: Mixed Dryland/Irrigated Cropland and Pasture; CSH: Closed Shrublands; SHR: Shrubland; DBF: Deciduous Broadleaf Forest; ENF: Evergreen Needleleaf Forest. Unless otherwise indicated by parentheses, Noah-MP uses the same land cover type as AmeriFlux for the corresponding 0.125° × 0.125° model grid. There are six sites where Noah-MP and AmeriFlux use slightly different names. At site 3, for example, AmeriFlux uses CRO, while Noah-MP uses DCP.

data, which are gap filled using Artificial Neural Network and Marginal Distribution Sampling techniques (http://public.ornl.gov/ameriflux).

[20] The SCAN [*Schaefer et al*., 2007] is a nationwide soil moisture and climate information system led by the Natural Resources Conservation Service (NRCS), USDA (http://www.wcc.nrcs.usda.gov/scan/). SCAN soil moisture data are collected by dielectric constant measuring devices at 5 cm, 10 cm, 20 cm, 50 cm, and 100 cm, where possible. The data used in this study have been processed by extensive quality control steps [*Liu et al*., 2011], through which any unrealistic data values (e.g., data outside a reasonable range and inconsistent data affected by sensor calibration or installation) and data measured under frozen conditions were excluded. Figure 2 shows the 60 available SCAN stations in the MRB and their data availability. Due to the low data availability at most stations, the data are aggregated into monthly basin average.

[21] GRACE-derived TWS anomaly data [*Wahr et al*., 2004] can now validate the performance of LSMs in TWS simulation, which is an overall indicator of the model proficiency in simulating the water budget. Recently, *Swenson and Milly* [2006] have used GRACE data for climate model evaluation, *Niu et al*. [2007] and *Lo et al*. [2010] have successfully used GRACE data for the development of groundwater dynamics in LSMs, and *van Dijk et al*. [2011] have used GRACE data to evaluate the Australian Water Resources Assessment system and recommended necessary improvements to the system, such as better precipitation forcing and the addition of groundwater dynamics. More applications of GRACE data in model development and evaluation can be found in a review paper by *Güntner* [2008]. This study uses the monthly GRACE TWS anomaly data, which have been processed into a 1° × 1° resolution gridded format [*Landerer and Swenson*, 2012; *Swenson and Wahr*, 2006] for easy comparison with LSM outputs and which can be publicly accessed on the Jet Propulsion Laboratory TELLUS website (http://grace.jpl.nasa.gov). The data are based on the CSR RL4.0 release by the Center for Space Research at the University of Texas at Austin. First, a destriping filter was applied to the data to minimize the systematic errors, which manifest as north-south oriented "stripes" in the GRACE TWS maps; then a 300 km wide Gaussian filter was applied to reduce random errors in higher-degree spherical harmonic coefficients not removed by the previous filter; and lastly, a spherical harmonic filter cutoff at 60° was applied. During the filtering process, because GRACE TWS was spatially averaged, signals were attenuated by showing smaller root-mean-square variability. To restore the signal attenuation, a gain factor, which was derived by using a simple least square regression to minimize the mismatch between the unfiltered, true, and filtered storage time series, was applied to each of the 1° × 1° grids. More information about the data processing can be found in *Landerer and Swenson* [2012], *Chen et al*. [2006], and *Swenson and Wahr* [2006].

2.5. Evaluation Statistics

[22] The agreement between the values predicted by a model and the values actually observed is measured using the following statistics: mean, root-mean-square error (RMSE), square of the correlation coefficient (R^2), and Nash-Sutcliffe

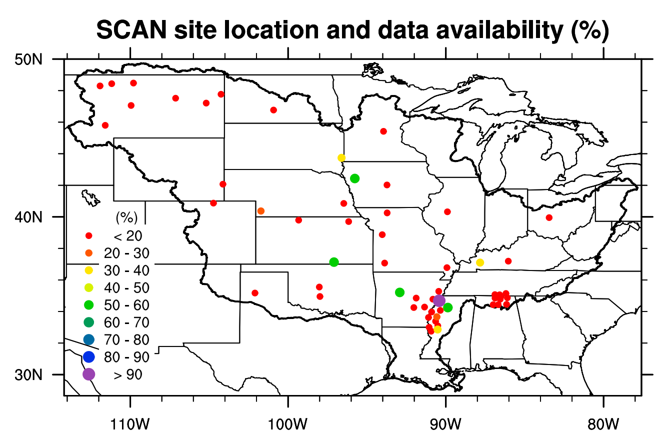

Figure 2. Map of the 60 SCAN stations and their data availability (percentage of the total number of months from 2002 to 2007 with observational data) in the Mississippi River Basin.

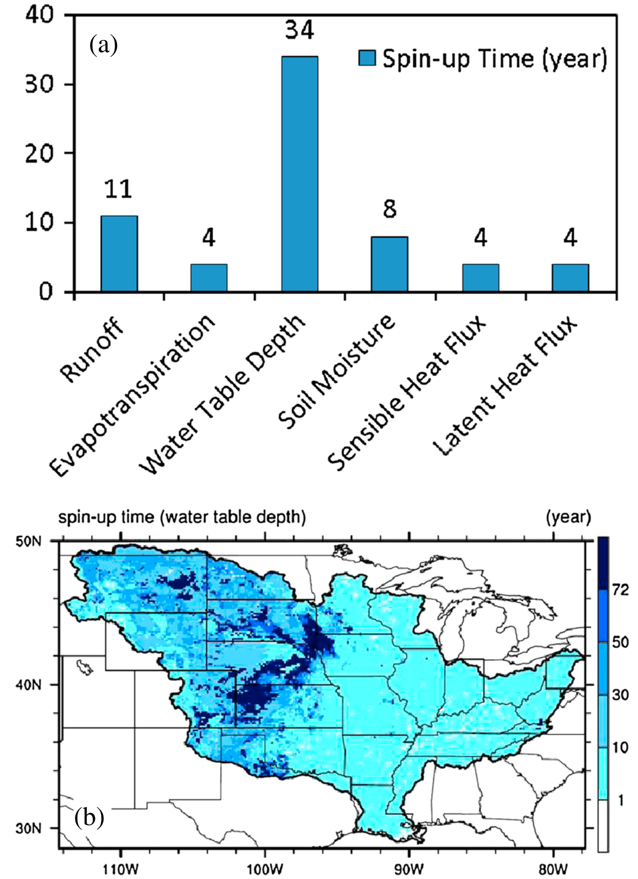

Figure 3. (a) Spin-up time (in years) for the individual variables based on averaged values for the entire Mississippi River Basin and (b) spatial distribution of the spin-up time (in years) for the water table depth.

efficiency (NSE) coefficient [*Nash and Sutcliffe*, 1970]. The NSE is calculated as

$$\mathrm{NSE} = 1 - \frac{\sum_{i=1}^{N} (M_i - O_i)^2}{\sum_{i=1}^{N} (O_i - \overline{O})^2} \tag{8}$$

where M_i and O_i are the predicted and measured values of the same variable, respectively, and $\overline{O}$ is the mean of the measured values. NSE ranges from minus infinity (poor fit) to 1 (perfect fit). In general, model prediction is considered to be satisfactory if NSE > 0.50 [*Moriasi et al.*, 2007].

3. Model Spin-Up, Sensitivity Tests, and Calibration

3.1. Model Spin-Up

[23] To allow some of the model variables with longer memories reach equilibrium, a numerical model must be properly initialized. When SIMGM, the groundwater component of Noah-MP, was introduced [*Niu et al.*, 2007], it was noted that it might take at least 250 years to spin-up the WTD in arid regions. Therefore, we are interested in examining the time span required to spin-up the model for this river basin. In this study, the spin-up is completed by running the model repeatedly through 1997 until each of the variables reaches equilibrium and the spin-up time is defined as year n, if

$$\left|\mathrm{Var}^{n+1} - \mathrm{Var}^{n}\right| < 0.001 \cdot \left|\mathrm{Var}^{n}\right| \tag{9}$$

where Var stands for each of the variables for the spin-up. This definition is as strict as the constraint by *Yang et al.* [1995]. The Var for the calculation in Figure 3a is spatially averaged, and for the calculation in Figure 3b, it is averaged per grid cell.

[24] The WTD requires the longest spin-up time, 34 years (Figure 3a), followed by runoff with 11 years and soil moisture (total soil column) with 8 years. This is consistent with previous studies of WTD [*Niu et al.*, 2007] and soil moisture [*Cosgrove et al.*, 2003; *Yang et al.*, 1995]. However, it is surprising to note that the spin-up time needed for runoff is longer than for soil moisture, which is not as commonly reported in literature. This may be due to the long spin-up time for WTD, which influences the runoff generation. The sensible heat fluxes and latent heat fluxes need shorter times to spin-up, approximately 4 years, because they are more influenced by surface soil and vegetation states and by the atmospheric forcing data, as indicated in equation (6).

[25] Regarding the spatial distribution of the time (in years) required for WTD to reach equilibrium (Figure 3b), for the wet region (east), less than 10 years is required to spin-up, whereas for the dry region (west), more than 72 years or even hundreds of years may be required for some small but extremely dry areas.

3.2. Parameter Sensitivity Tests

[26] Hydrological modeling involves significant efforts in parameter sensitivity testing and calibration, which were usually overlooked in the past. However, it is becoming a must when LSMs are more and more used in hydrological studies. Here we briefly describe how the model parameters are finalized before the model is ready for evaluation. Based on our modeling experience and previous studies [*Rosero et al.*, 2010], several sensitive parameters are selected for further analysis. However, only three parameters are identified as sensitive parameters for runoff simulation: surface dryness factor (α), saturated hydraulic conductivity (k), and saturated soil moisture (θ_{max}). Table 2 provides the definitions, units, and value ranges of the three parameters. Figure 4 shows how annual runoff varies with different values of the individual parameters. Spatially averaged annual mean runoff (1) decreases as the surface dryness factor increases, in a nearly linear relationship, (2) decreases dramatically as hydraulic

Table 2. Experimental Design for Parameter Calibration

Parameters	Controlling Process	Units	Min	Max	Default	Values	#
Surface dryness factor	Partitions of the surface hydrology	—	0	10	6.0	2.5, 3, 3.5, 4, 4.5, 5, 5.5, 6	8
Saturated hydraulic conductivity	Base flow in runoff simulation	$\mathrm{m\,s^{-1}}$	2E-9	7E-2	Vary	Multiply by 0.01, 0.05, 0.1, 0.5, 1, 5, 10	7
Saturation soil moisture content	Water flow between aquifer and soil	—	0.10	0.71	Vary	Multiply by 0.8, 0.85, 0.9, 0.95, 1, 1.05, and 1.1	7

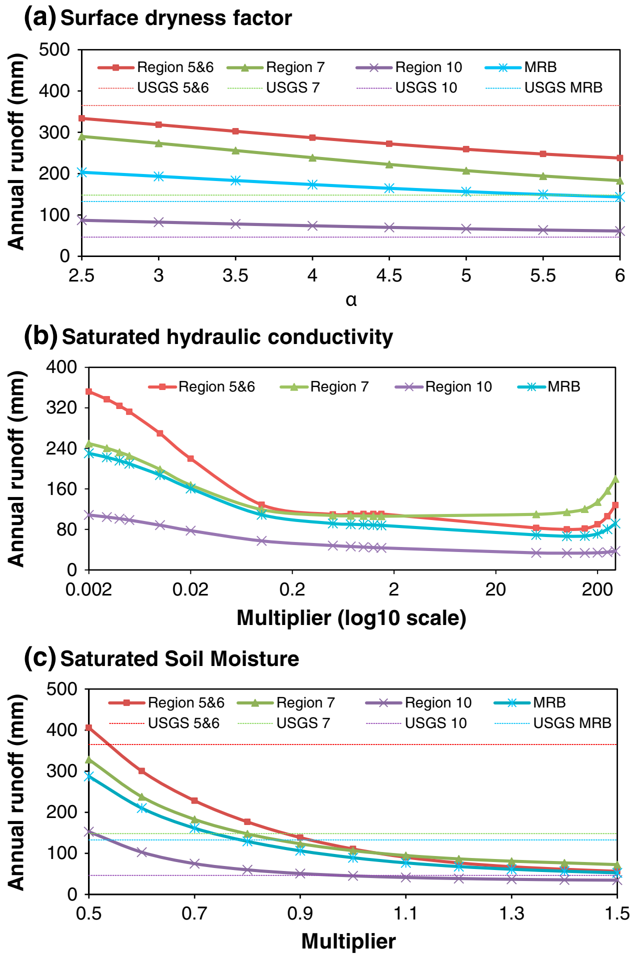

Figure 4. Sensitivity tests of (a) the surface dryness factor, (b) the saturated hydraulic conductivity, and (c) the maximum soil moisture (porosity).

conductivity increases when saturated hydraulic conductivity is less than 10% of its original values, and (3) also decreases dramatically as saturated soil moisture increases when the multiplier (a factor multiplied by the original values) is less than 0.9.

[27] Although the sensitivity tests are based on the changes in total runoff with different parameter values, they influence other hydrological variables more directly. For example, as indicated in equation (7), the surface dryness factor is a parameter that determines the soil surface resistance and hence controls soil evaporation; when the surface dryness factor increases, the soil evaporation increases correspondingly. Therefore, to maintain water balance, either or both transpiration and runoff have to decrease. In this case, annual runoff decreases as the surface dryness factor increases (Figure 4a). Total runoff is affected by the saturated hydraulic conductivity via its capability to control the subsurface runoff (base flow). Similarly, saturated soil moisture controls the storage capacity of the soil and hence affects evaporation and infiltration.

3.3. Model Calibration

[28] Based on the parameter sensitivity tests above, model calibration is conducted manually by obtaining the optimal combination of the three most sensitive parameters (α, k, and θ_{max}) for the entire MRB (i.e., called lumped calibration). As shown in Table 2, 392 experiments are designed and run. The three parameter values that produce the highest NSE for the entire MRB are selected. The calibrated hydrographs are shown in Figure 5 and the corresponding statistics are included in Table 3. In the hydrograph, the USGS-observed streamflow is from the station near the basin outlet and the Noah-MP runoff is aggregated from all the grids in each basin or subbasins. For the entire MRB, we can observe that the hydrograph is greatly improved from the default simulation to the calibrated simulation. Although the increase in R^2 is small (from 0.76 to 0.81), the decrease in RMSE is large (37%), which results in a large increase in NSE from 0.42 (less than the satisfactory threshold of 0.50) to 0.77. For the Ohio-Tennessee River Basin, all three statistics improve. For the Upper MRB and the Missouri River Basin, however, the NSE decreases (from 0.56 and 0.01 to 0.29 and -0.16, respectively) due to the increase in RMSE, although the R^2 increases.

[29] To improve the simulation for the subbasins, calibration is conducted by using different sets of the three parameters (α, k, and θ_{max}) for each subbasin (i.e., subbasin calibration). However, the improvement to the runoff simulation for the Upper MRB and the Missouri River Basin is very trivial (not shown here). The possible reasons are (1) the selected sensitive parameters are not applicable for these two subbasins, and/or (2) hydrological modeling for arid and semiarid areas (such as these two subbasins) is more difficult than that for humid areas—a well-recognized problem. Figure 5d shows that subbasin calibration does significantly improve the simulation for the Ohio-Tennessee River Basin. From the lumped calibration to the subbasin calibration, the NSE (R^2) increases from 0.36 (0.67) to 0.68 (0.81). In summary, it is worthwhile to calibrate models at the subbasin level for humid regions, whereas for arid and semiarid regions, the model structure and the selection of the sensitive parameters need further investigation.

4. Evaluation and Discussion

4.1. Runoff

[30] Figure 6 compares the spatial distributions of the UNH-GRDC composite runoff, the USGS hydrologic unit runoff, and the Noah-MP simulated runoff. Noah-MP is capable of capturing the observed general spatial pattern of the runoff, which is similar to the precipitation pattern shown in Figure 1c. For UNH-GRDC, the runoff in the red box is much lower than its surrounding area, which is not found in the Noah-MP simulation or the USGS observations. On the contrary, the Noah-MP-simulated runoff in that box is slightly higher than its surrounding area, whereas the USGS runoff follows the general transition pattern. The USGS stream gages in this region are very sparse compared with other regions (not shown here), which might explain the difference in the UNH-GRDC runoff data because the UNH-GRDC runoff requires runoff input from the USGS. The high Noah-MP runoff in that box corresponds to the high precipitation in Figure 1c. It is also notable that the UNH-GRDC runoff is less than 1 mm for most of the western portion of the basin (red shaded area in Figure 6b), whereas it is 5 mm to 25 mm in the Noah-MP runoff and 6 mm to 50 mm in the USGS runoff.

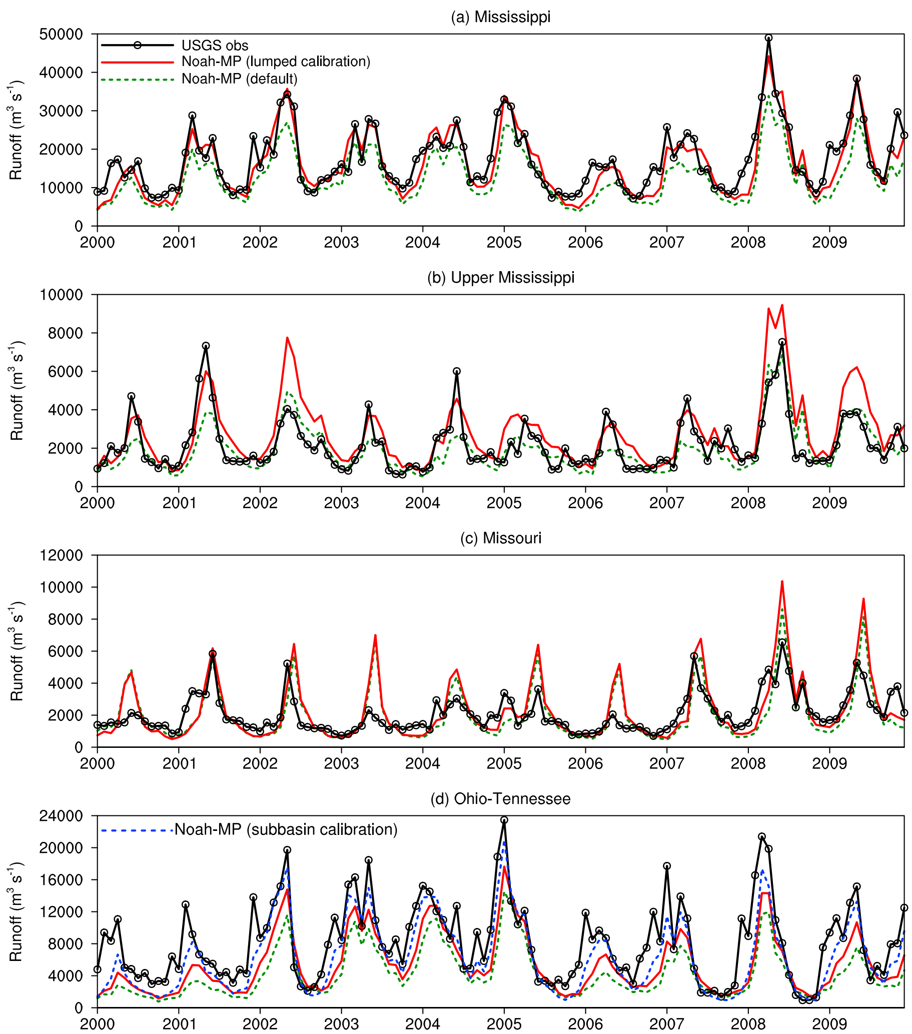

Figure 5. Comparisons of the USGS-observed and the Noah-MP- simulated (default, lumped calibrated, and subbasin calibrated) hydrographs for (a) the Mississippi River Basin, (b) the Upper Mississippi River Basin, (c) the Missouri River Basin, and (d) the Ohio-Tennessee River Basin.

[31] To examine how Noah-MP is improved from the baseline Noah LSM in terms of runoff simulation, we also present the comparison in Figure 6e, which indicates a substantial improvement from the Noah LSM to Noah-MP. The results from the default Noah-MP setting slightly underestimate the USGS observations; however, they are already much better than the baseline Noah LSM results. The results from the calibrated Noah-MP are further improved, as both the magnitudes and the temporal variations correspond closely with the USGS observations. For easy comparison with similar studies, monthly climatological runoff is also shown in Figure 6d. Compared with previous studies by *Falloon et al.* [2011], *Li et al.* [2011], and *Xia et al.* [2012a], Noah-MP performs as well as or better than other mainstream LSMs in runoff modeling.

[32] One must bear in mind that these improvements may be undermined by possible uncertainties in our comparison process and the models used. First of all, it is unfair to compare the model-simulated runoff directly (without river routing) with the USGS-gaged streamflow. However, the influence of the runoff routing on the comparison is relatively minor if we compare them at the monthly scale. Second, the USGS-gaged runoff is a direct measurement of the streamflow through a specific location without tracking its exact movement and distribution; therefore, great uncertainties evolve from human activities such as irrigation, tile drainage [*Li et al.*, 2010], water supply, and reservoir regulation, as the MRB is one of the river basins that involve intensive water consumption [*Murray et al.*, 2011]. The traditional method is to use gaged streamflow to retrieve natural runoff without human interference. However, streamflow naturalization requires significant data on water use and water resource

Table 3. Statistical Summary of Model Calibration for the Mississippi River Basin and Some of Its Subbasins

	Mississippi		Upper Mississippi		Missouri		Ohio-Tennessee		
	CTL[a]	LPC[b]	CTL	LPC	CTL	LPC	CTL	LPC	SBC[c]
RMSE	6128	3875	904	1147	1192	1290	5054	3919	2773
R^2	0.76	0.81	0.60	0.68	0.48	0.57	0.60	0.67	0.81
NSE	0.42	0.77	0.56	0.29	0.01	−0.16	−0.07	0.36	0.68

[a]CRT: default (control) model run.
[b]LPC: lumped calibration.
[c]SBC: calibration for specific subbasin.

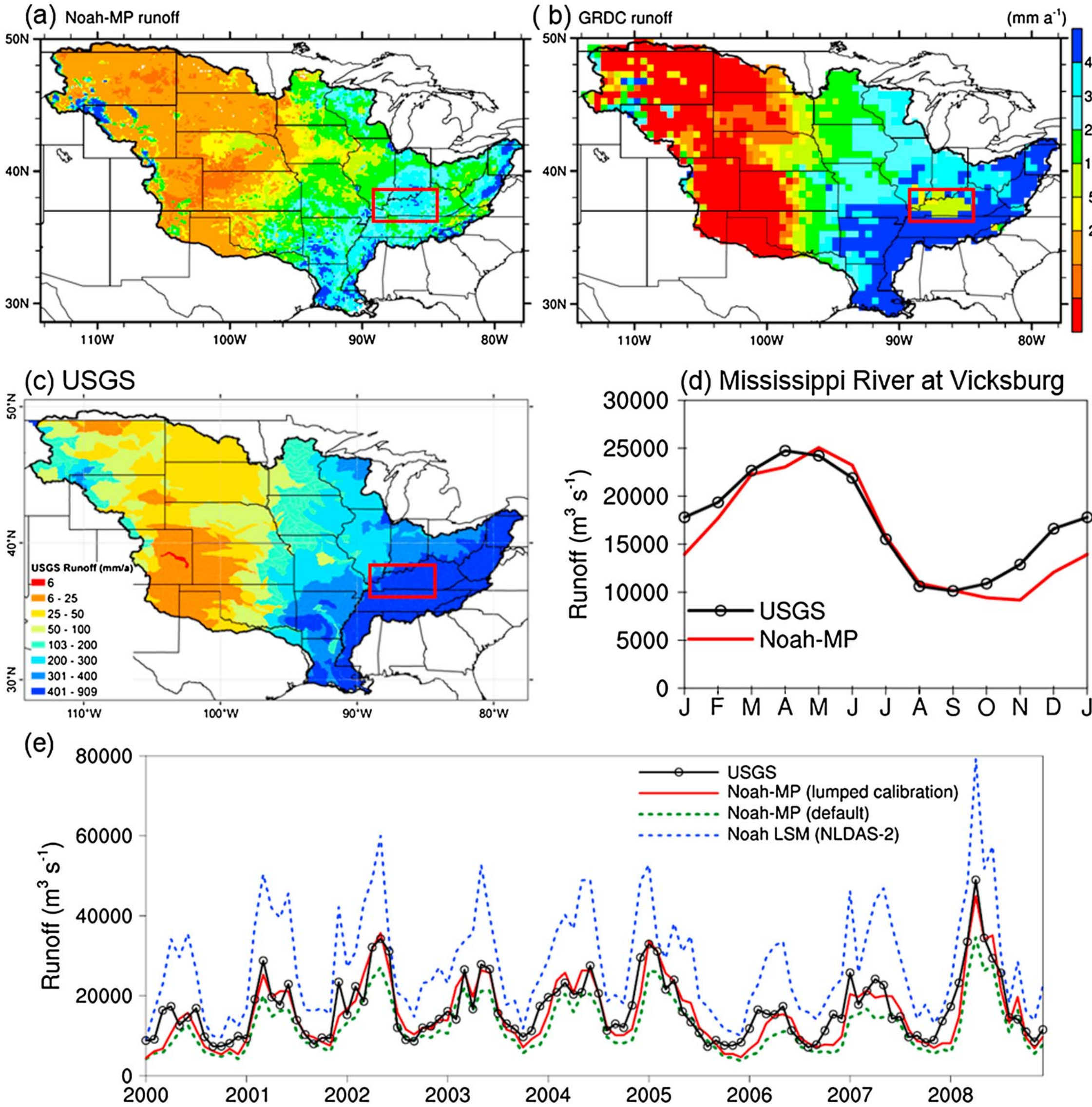

Figure 6. Climatological mean annual runoff from (a) Noah-MP (2000–2009), (b) UNH-GRDC (all available observational time periods [*Fekete et al*., 2002]), and (c) USGS hydrologic unit runoff (1901–2009). (d) Monthly climatological runoff (2000–2009) from USGS observation and Noah-MP simulation. (e) USGS-gaged and Noah-MP (default and calibrated) and Noah LSM simulated runoff (2000–2008) for the Mississippi River Basin. The region enclosed in the red box is discussed in the text.

management, which are difficult to collect. Third, Noah-MP does not include a process to represent the artificial tile drainage in the model, which is a very important mechanism in some of Midwest areas, not only in runoff generation but also in groundwater and soil moisture modeling [*Algoazany et al*., 2007; *Gentry et al*., 2009; *Goswami*, 2006; *Li et al*., 2010]. One possible solution is to improve the ability of Noah-MP to represent human activities (e.g., irrigation and tile drainage) in future model development.

4.2. Groundwater

[33] Groundwater dynamics have attracted increasing attention within the climate community [*Fan et al*., 2007; *Leung et al*., 2011; *Lo et al*., 2010; *Miguez-Macho et al*., 2007; *Niu et al*., 2007] for three reasons. First, groundwater directly influences soil moisture, which is an important variable in LSMs and climate models, an important indicator for drought detection, and a major controlling factor for the interaction between the land and the atmosphere [*Niu et al*., 2007]. Second, groundwater, which provides most of the water needed for ET during the dry season [*Gutowski et al*., 2002], also influences ET. Because ET is both a water flux and a heat flux term, the influence of groundwater is passed on throughout the surface energy and water balances. Third, the inclusion of groundwater dynamics in climate models provides a direct tool to evaluate the impact of climate change on groundwater systems, which is vital for research into climate change adaptation.

[34] Groundwater dynamics is one of the major improvements in Noah-MP; however, the employed SIMGM groundwater model was evaluated against the GRACE TWS anomaly data at the global scale [*Niu et al*., 2007] without

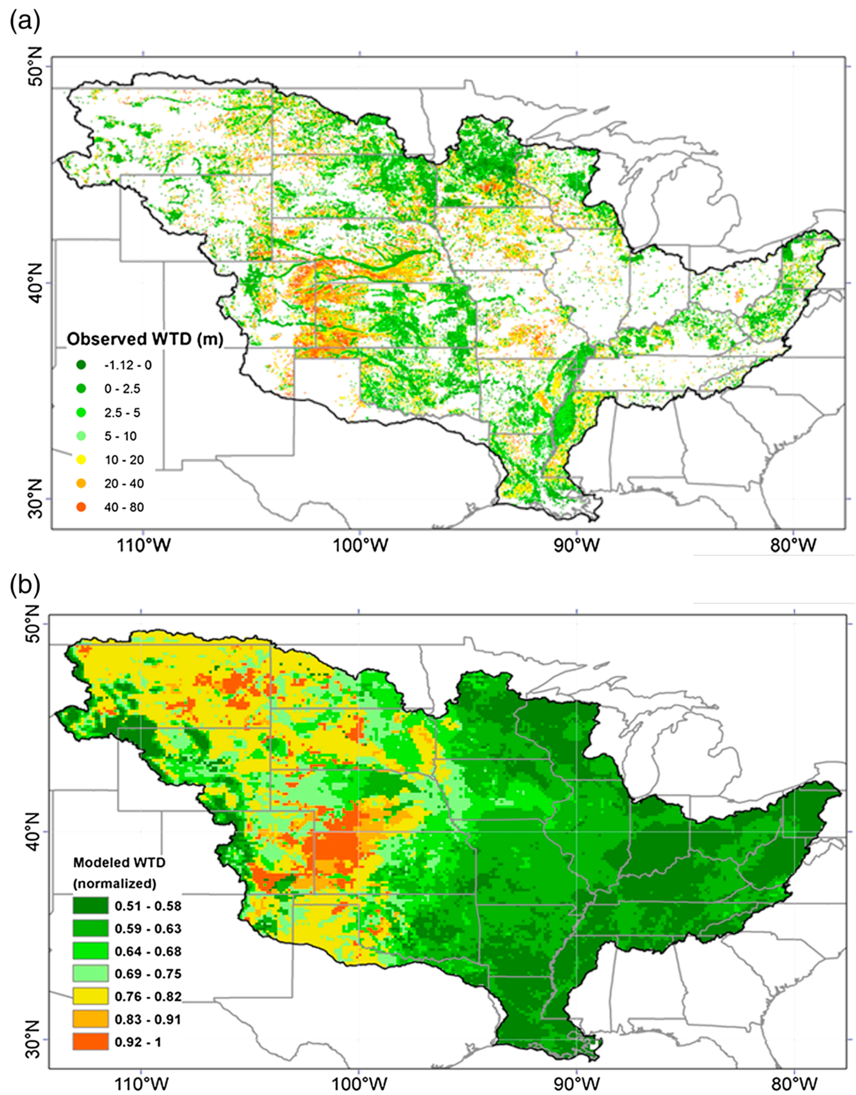

Figure 7. Climatological water table depth from (a) USGS measurements (all available observational time periods [*Fan and Miguez-Macho*, 2011]) and (b) Normalized Noah-MP simulations (2000–2009).

comparison with actual ground measurements. Therefore, this study compares the simulated WTD against the USGS-observed WTD, both spatially and temporally. Figure 7 shows that Noah-MP can simulate the climatological spatial pattern, in which the water table is shallower in the southeast and deeper in the northwest. Some small areas in the wet region with deep water tables are not well simulated by Noah-MP, which may be due to the coarse spatial resolution or the model structure. Temporal variation is also compared with the observed groundwater storage data (Figure 8). The simulated groundwater variations agree very well with the observations for the entire MRB, the Ohio-Tennessee River Basin, and the Upper MRB, with R^2 values of 0.75, 0.67, and 0.57, respectively. For the Ohio-Tennessee River Basin, the simulated anomalies are very similar to the observations. Because precipitation occurs frequently in this wet region, a small but very frequent fluctuation occurs in the observations, which Noah-MP fails to replicate. For the entire MRB and for the Upper MRB, the simulated anomalies are less than in the observations. For the Missouri River Basin, however, the simulated anomalies are much less than in the observations. The observed strong seasonal cycle is likely caused by the very shallow water table in this region (see Figure 7), which is because the aquifers are thin valley alluvium perched on top of the bedrock cores of the Rocky Mountains. These thin alluvial aquifers have very little storage, and thus, they are very responsive to seasonal snowmelt recharge (rises quickly) followed by efficient drainage into the deeply incised streams (i.e., the water level falls quickly). Models have difficulties representing these perched thin aquifers (Y. Fan, personal communication, 2011). For future model development, it would be helpful to collect the bedrock distribution data and include this process in LSMs.

[35] The modeled WTD is normalized here so that its spatial pattern is comparable to the observations. Indeed, Noah-MP-simulated WTD only ranges from 2 m to less than 14 m, whereas the observed WTD ranges from above the ground to greater than 80 m. The reason that WTD is limited to greater than 2 m in the model is to avoid a numerical computation problem. The simulated WTD cannot go deeper than 14 m, most likely due to the coarse spatial resolution. The range of WTD is very sensitive to the grid resolution:

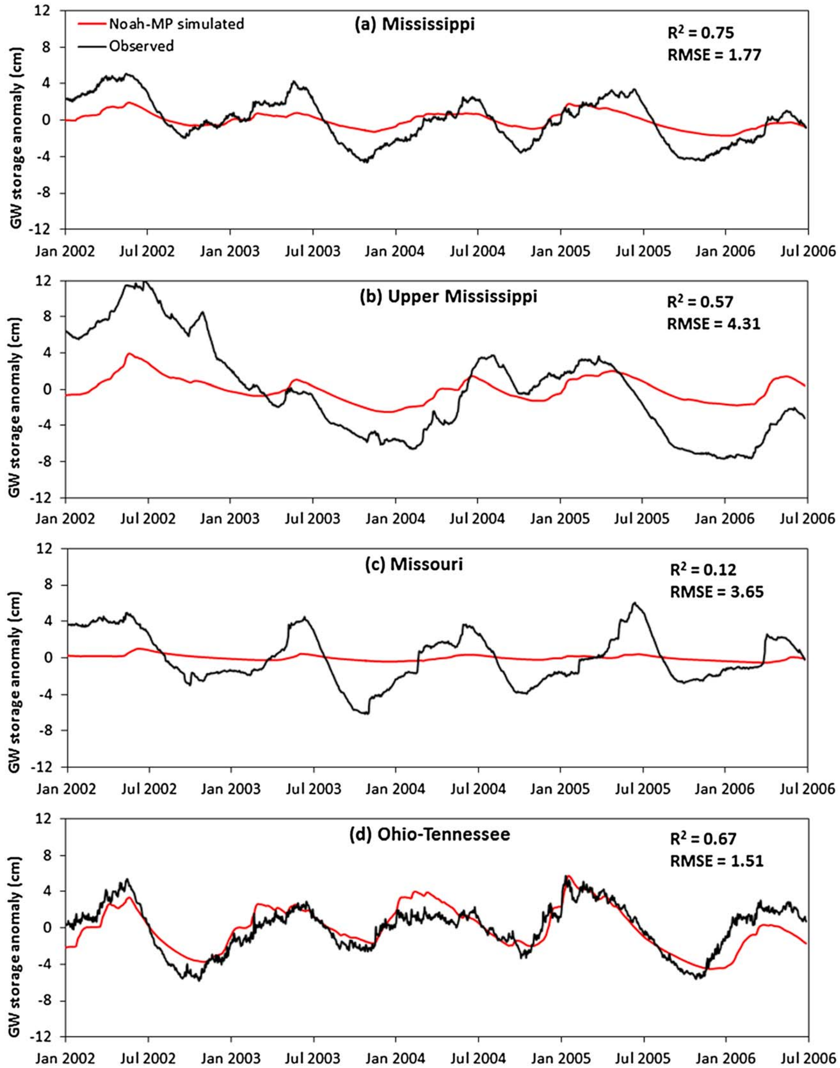

Figure 8. Comparison of the groundwater storage anomaly from observations [*Rodell et al.*, 2007] and model simulations for (a) the entire Mississippi River Basin, (b) the Upper Mississippi River Basin, (c) the Missouri River Basin, and (d) the Ohio-Tennessee River Basin.

The finer the grid, the larger the range of WTD because when the grids are finer, the steeper land slope can be represented in the model, which accelerates the drainage speed. The most prominent scale for groundwater divergence-convergence is from hilltops to valleys. When averaging over the valleys and hills, and thus only having regional gradients, we get regional WTD and groundwater flow, which have much smaller gradients and ranges (Y. Fan, personal communication, 2011). Therefore, improving the spatial resolution is another direction for groundwater model development in LSMs, but there is always a tradeoff between model resolution and computational affordability.

4.3. Evapotranspiration

[36] For the entire MRB, the simulated canopy evaporation, transpiration, and soil evaporation are 35.6 mm, 278.6 mm, and 323.5 mm, respectively, which account for 5.6%, 43.7%, and 50.7% of the total ET, respectively. To distinguish the effect of different vegetation types on ET, the 15 AmeriFlux tower sites are divided into four groups by their land cover types, with five grassland sites, five cropland sites, four forestland sites, and one shrubland site. Their climatological latent heat fluxes (the energy form of ET) from observations and from model simulations are presented in Figure 9. Among the four land cover types, forestlands and shrublands show significant improvements from Noah LSM to Noah-MP, in terms of better timing and more similar mean values, and grasslands also have improved timing. However, we find that the Noah-MP-simulated latent heat fluxes are slightly higher than the observations from AmeriFlux for all four land cover types, which is similar to the evaluation by *Blyth et al.* [2010], whereas Noah LSM underestimates latent heat fluxes for forestlands and shrublands, overestimates for croplands, and well estimates the mean value for grasslands. Interestingly, the three land cover types for which Noah-MP performs well are grassland, forestland, and shrubland, which are considered more naturally occurring, whereas the land cover type for which Noah-MP does not perform well is cropland, which involves more human activities. This is most likely due to the use of leaf dynamics in Noah-MP, which can capture the processes of natural growth but cannot capture anthropogenic crop growth; thus, its ET increases too quickly

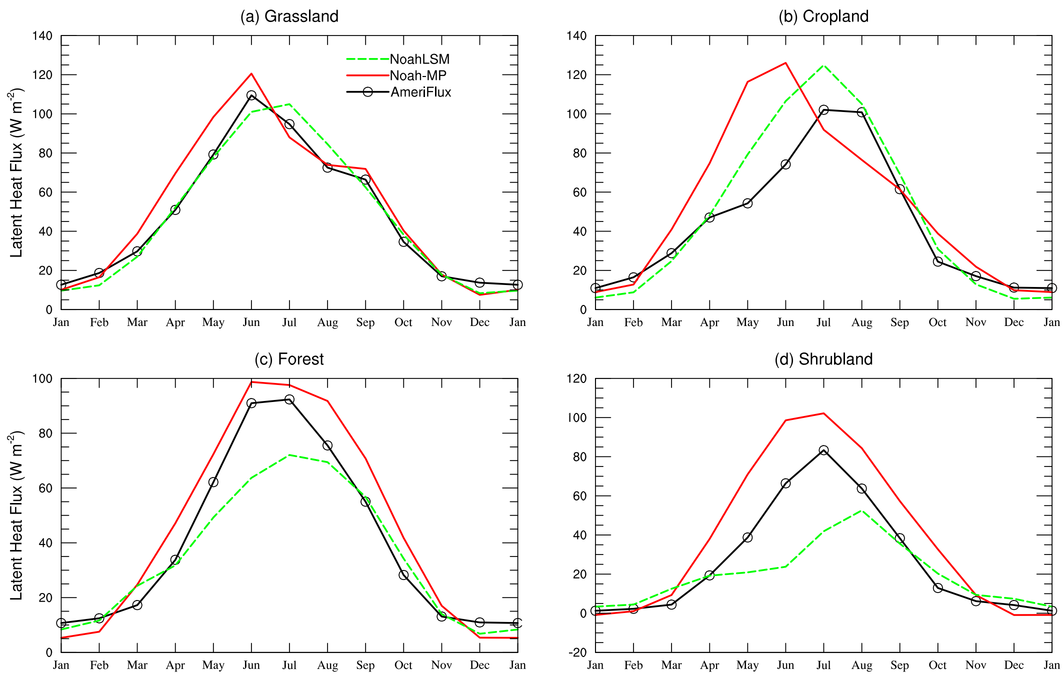

Figure 9. Comparison of the latent heat flux for the AmeriFlux observations and the model simulations for different land cover types. (a) Grassland (five sites), (b) cropland (five sites), (c) forestland (four sites), and (d) shrubland (one site).

during spring. In contrast, Noah LSM uses prescribed monthly LAI for various vegetation types and monthly GVF climatological values derived from NOAA/AVHRR, which better match anthropogenic crop growth. In addition to maintaining the strength in modeling natural vegetation dynamics, improvement in the simulation of the dynamic leaf model for cropland is highly recommended (this land type is expected to expand with the increasing population). One of the limitations of this study is that only runoff is calibrated; however, it is recommended that both runoff and ET be calibrated at the same time.

4.4. Soil Moisture

[37] Studies [e.g., *Entin et al.*, 2000] have shown that soil moisture measured at one location can represent the temporal variation for the surrounding area, up to 500 km in radius. Therefore, it is reasonable to use station-measured soil moisture to evaluate model-simulated soil moisture. Figure 10 compares the Noah-MP-simulated and SCAN-observed soil moistures for individual soil layers (the top 10 cm, 10–40 cm, 40–100 cm, and 100–200 cm). For the top layer (the top 10 cm), the Noah-MP-simulated soil moisture values are nearly identical to the SCAN observations, with an R^2 of 0.923 and an RMSE of 0.016. Because the top layer plays an important role in the water and energy exchanges between the land surface and the atmosphere, Noah-MP shows its high potential ability to study land-atmosphere coupling. For the second layer (10–40 cm), although the discrepancy is slightly greater than the top layer (RMSE is 0.025), the comparison has an even higher R^2 value (0.933). A larger discrepancy is found in the deep layers, with R^2 of 0.624 and RMSE of 0.077 for the third layer (40–100 cm) and R^2 of 0.574 and RMSE of 0.035 for the bottom layer (100–200 cm). In terms of R^2, the results for the two deep layers are still acceptable; however, the Noah-MP-simulated soil moisture values largely underestimate the SCAN observations for the third layer, particularly in the summer and fall.

[38] Why does Noah-MP underestimate the SCAN soil moisture for the third soil layer? This question can be answered by the connection with the ET comparison in the previous section, where it was demonstrated that Noah-MP-simulated ET values increased more quickly than did the AmeriFlux observations in the spring (Figure 9). Due to the high values of the simulated ET, more water is extracted from soil, which very likely leads to the low values of the simulated soil moisture. Furthermore, the variation in ET is dominated by transpiration (approximately twice that of soil evaporation). The transpiration rate from each soil layer is determined by the soil moisture factor controlling stomatal resistance, β_i (the higher value, the larger fraction of water for transpiration from the layer), which is parameterized as

$$\beta_i = \frac{\Delta z_i}{z_{\text{root}}} \min\left(1.0, \frac{\theta_{\text{liq},i} - \theta_{\text{wilt}}}{\theta_{\text{ref}} - \theta_{\text{wilt}}}\right) \Big/ \sum_{i=1}^{N_{\text{root}}} \left(\frac{\Delta z_i}{z_{\text{root}}} \min\left(1.0, \frac{\theta_{\text{liq},i} - \theta_{\text{wilt}}}{\theta_{\text{ref}} - \theta_{\text{wilt}}}\right)\right) \tag{10}$$

where Δz_i is the thickness of the ith soil layer (m), z_{root} is the total depth of root zone (m), $\theta_{\text{liq},i}$ is the liquid soil moisture in the ith soil layer ($m^3\,m^{-3}$), θ_{wilt} is the soil moisture at wilting point ($m^3\,m^{-3}$), and θ_{ref} is the reference soil moisture (close to field capacity) ($m^3\,m^{-3}$). In our comparison, most of the model grids with good availability of SCAN data (Figure 2) are croplands, grasslands, and shrublands, which have shallow roots, so that root depth may only reach the third soil layer; only a few model grids with poor availability of SCAN data are forestlands, which have deep root depths that reach the bottom layer. Because of the great thickness of the third soil layer (0.6 m), the water supply for transpiration is

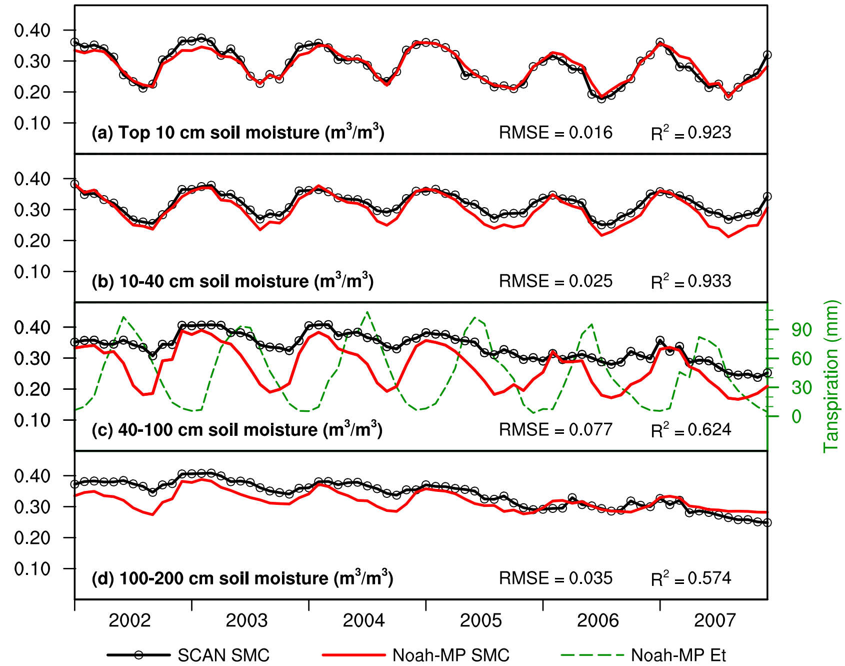

Figure 10. SCAN-observed and Noah-MP-simulated monthly soil moisture (SMC) for the Mississippi River Basin at a depth of (a) top 10 cm, (b) 10–40 cm, (c) 40–100 cm, and (d) 100–200 cm. Figure 10c also shows the Noah-MP simulated transpiration (ET). For the Noah-MP simulation, only those grids with a SCAN site are included, and for each grid, only those months with observed values are used.

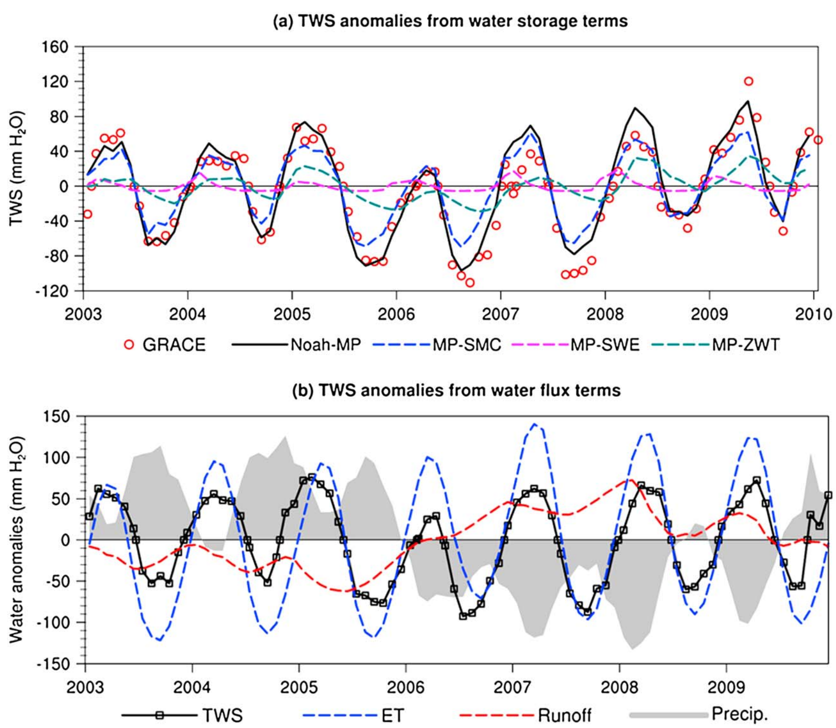

Figure 11. TWS anomalies for the Mississippi River Basin calculated from (a) the water storage terms and their contributing components and (b) the water flux terms and their contributing components. TWS anomalies are the cumulative anomalies of (Precipitation – ET – Runoff), which are compared concurrently with the anomalies of the individual terms –ET and –Runoff. Note that the ET and runoff anomalies are shown as the negative of the original anomalies.

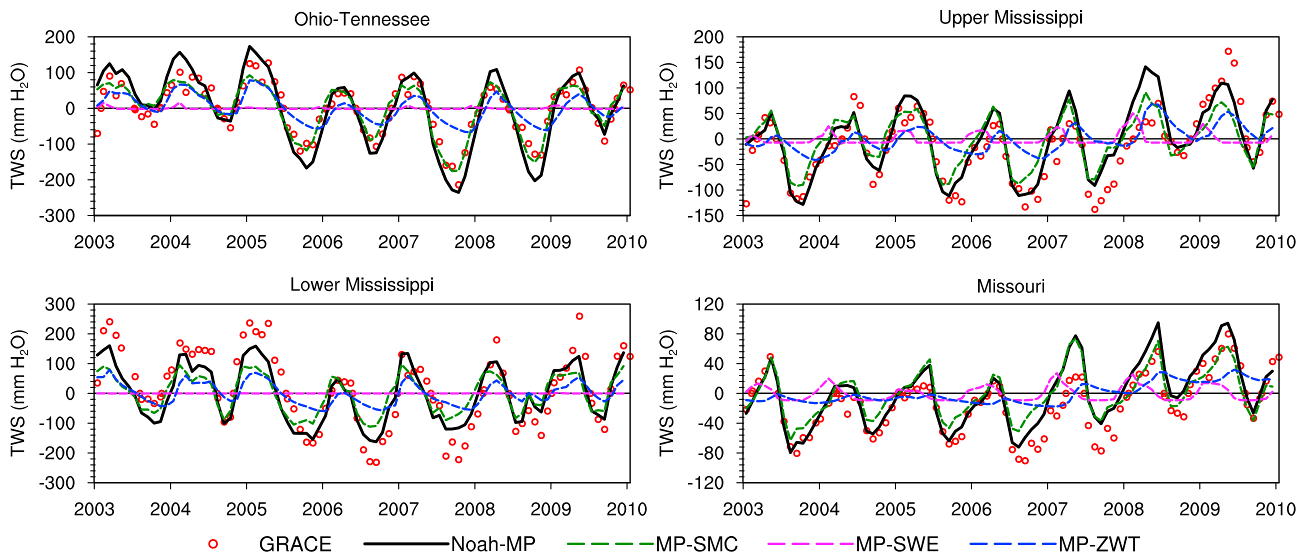

Figure 12. Comparison of the TWS anomalies from the GRACE-based measurements and the Noah-MP simulations from the water storage terms and their contributing components for the four subbasins.

heavily from the third layer. However, this may not be true in reality for croplands and grasslands, where roots may only reach the upper portion of the third layer not the entire third layer. Therefore, the water extraction for transpiration from the third layer is overestimated.

[39] Figure 10c shows the strong annual cycle of transpiration, where high transpiration rates correspond to low soil moisture and low transpiration rates correspond to high soil moisture in the third soil layer. There are approximately 2 months of phasing difference between the transpiration and soil moisture, which is because after its peak, transpiration remains high and continues to dry the soil. To improve the soil moisture estimation for the third layer, adjustments (parameters or parameterization) to the dynamic leaf model in Noah-MP are needed to limit the increase in the ET rate in the spring.

4.5. Terrestrial Water Storage

[40] The Noah-MP-simulated TWS anomaly is compared with the GRACE-based TWS anomaly for the entire MRB in Figure 11, in which the contributing components of the simulated TWS anomaly are also presented. There are several notable points from this comparison.

[41] 1. Noah-MP agrees well with GRACE in terms of the TWS anomaly, indicating that Noah-MP can capture the overall water cycle, including both the timing and the magnitude of water fluctuation. Although it may still involve great uncertainties from each of these components, Noah-MP captures the most important components such as soil moisture, groundwater, and snow.

[42] 2. Because Noah-MP does not simulate the water storage of ice, lakes, rivers, and biomass, from the water balance point of view, TWS has to be balanced by soil moisture, groundwater, and snow, which are the water storage terms that are simulated in the model. In this particular region, soil moisture contributes the most to the TWS anomaly, followed by groundwater, and then snow. Although Noah-MP has difficulty in capturing the absolute values of WTD (see section 4.2), it is quite capable of capturing the annual groundwater fluctuation. Compared with the original Noah LSM, in which TWS is only balanced by soil moisture and snow, Noah-MP obtains a great improvement by including the second largest component of the TWS anomaly—groundwater.

[43] 3. Noah-MP still simulates only the natural part of the TWS anomaly, without considering human activities; therefore, Noah-MP has difficulty reflecting human interference. We can clearly observe that in the GRACE-based TWS anomaly curve, where there are two peaks in approximately half of the years, which cannot be observed in the Noah-MP curve. This is very likely due to human activities, for example, high irrigation rates in the spring. Some researchers have attempted to include irrigation in LSMs [*Ozdogan et al.*, 2010; *Pokhrel et al.*, 2012; *Sorooshian et al.*, 2012].

[44] 4. In Figure 11b, the variation amplitude of the cumulative anomaly of ET is much higher than in the other fluxes (precipitation and runoff); hence, ET is the dominant water flux driving the TWS anomaly.

[45] As shown in Figure 12, we also compare the Noah-MP-simulated TWS anomaly with the GRACE-based TWS anomaly at the subbasin level. In all four regions, soil moisture is always the largest contributor to the TWS anomaly; groundwater is the second largest contributor in the Ohio-Tennessee, Upper Mississippi, and Lower Mississippi regions, but in the cold Missouri region, snow contributes as much as groundwater to the TWS anomaly. In the smaller subbasins, such as the Upper and Lower Mississippi, the agreement between the model and the GRACE TWS anomalies is not as good as the agreement at the level of the entire MRB.

5. Conclusions

[46] In line with the GEWEX GLASS for LSM benchmarking or better model evaluation, we evaluated the model at the continental basin scale, specifically for the MRB. We began our evaluation with model spin-up, parameter sensitivity tests, and model calibration, and then the calibrated results were compared with a number of traditional and recently available observational data sets. From this study, we have

reached several conclusions that may be of interest to LSM developers and users.

[47] With groundwater dynamics included in Noah-MP, it takes longer for WTD to reach equilibrium than without groundwater dynamics. This long WTD spin-up time would influence the spin-up times of other variables because when the water table is far from an equilibrium state, other variables such as runoff, ET, and soil moisture need to be adjusted to help WTD reach equilibrium. For the entire MRB, at least 34 years is required for the model to spin-up. For some mountain regions with very deep water tables, hundreds of years may be required for the model to spin-up.

[48] Runoff is found to be sensitive to three parameters: the surface dryness factor (α), the saturated hydraulic conductivity (k), and the saturated soil moisture ($\theta_{\max}$); these three factors are selected for model calibration to improve runoff simulation. Although lumped calibration can improve model performance, distributed calibration is needed to obtain the best parameter values for some wet regions. If time and resources are limited for conducting automatic calibration (e.g., in this study), a better understanding of model physics and more analyses of the observational data would shorten the calibration time and benefit the model performance.

[49] Noah-MP has shown significant improvements in hydrological modeling.

[50] 1. The Noah-MP-simulated runoff is significantly improved compared with the baseline Noah LSM output in the NLDAS-2 framework. The spatial pattern of the Noah-MP simulated runoff matches fairly well with both the UNH-GRDC runoff and the USGS hydrologic unit runoff. We believe that this is the first time the USGS hydrologic unit runoff has been used in LSM evaluation and found to be very reasonable.

[51] 2. Groundwater evaluation indicated that Noah-MP captures the general spatial pattern of the climate conditions and captures the temporal patterns for the wet regions. However, it fails in simulating the absolute values and the temporal variation in the water table for the dry regions.

[52] 3. The addition of leaf dynamics to Noah-MP has improved its performance in ET simulation for natural land cover types.

[53] 4. One of the highlights of the study is that Noah-MP produces soil moisture values consistent with the SCAN observations for the top two soil layers (0–10 cm and 10–40 cm), which indicates its great potential for use in studying land-atmosphere coupling.

[54] 5. The Noah-MP-simulated TWS anomaly agrees very well with the GRACE observations, which may partly benefit from the inclusion in the model of groundwater dynamics, considered the second largest component of the TWS anomaly for most of the MRB.

[55] **Acknowledgments.** This work is supported by the NASA Interdisciplinary Science Program, award NNX11AJ43G and NSFC grant 41375088. The authors would like to thank Seungbum Hong for his help in model setup, Ying Fan for providing observed water table depth data and sharing her insight into these data, Youlong Xia for sharing the SCAN-observed soil moisture data, and Sean C. Swenson for providing the GRACE data and for his help in the calculation of modeled TWS. The first author is grateful to Jiangfeng Wei for his invaluable comments.

References

Abramowitz, G. (2012), Towards a public, standardized, diagnostic benchmarking system for land surface models, *Geosci. Model Dev.*, *5*(3), 819–827, doi:10.5194/gmd-5-819-2012.

Algoazany, A. S., P. K. Kalita, G. F. Czapar, and J. K. Mitchell (2007), Phosphorus transport through subsurface drainage and surface runoff from a flat watershed in east central Illinois, USA, *J. Environ. Qual.*, *36*(3), 681–693, doi:10.2134/Jeq2006.0161.

Baldocchi, D., et al. (2001), FLUXNET: A new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities, *Bull. Am. Meteorol. Soc.*, *82*(11), 2415–2434, doi:10.1175/1520-0477(2001)082<2415:Fantts>2.3.Co;2.

Blyth, E., J. Gash, A. Lloyd, M. Pryor, G. P. Weedon, and J. Shuttleworth (2010), Evaluating the JULES land surface model energy fluxes using FLUXNET data, *J. Hydrometeorol.*, *11*(2), 509–519, doi:10.1175/2009jhm1183.1.

Bonan, G. B. (2008), *Ecological Climatology: Concepts and Applications*, 2nd ed., pp. 550, Cambridge Univ. Press, Cambridge, U. K.

Brakebill, J. W., D. M. Wolock, and S. E. Terziotti (2011), Digital hydrologic networks supporting applications related to spatially referenced regression modeling, *J. Am. Water Resour. Assoc.*, *47*(5), 916–932, doi:10.1111/j.1752-1688.2011.00578.x.

Chen, F., Z. Janjic, and K. Mitchell (1997), Impact of atmospheric surface-layer parameterizations in the new land-surface scheme of the NCEP mesoscale Eta model, *Bound.-Lay. Meteorol.*, *85*(3), 391–421, doi:10.1023/A:1000531001463.

Chen, J. L., C. R. Wilson, and K. W. Seo (2006), Optimized smoothing of Gravity Recovery and Climate Experiment (GRACE) time-variable gravity observations, *J. Geophys. Res.*, *111*, B06408, doi:10.1029/2005JB004064.

Clark, M. P., D. Kavetski, and F. Fenicia (2011), Pursuing the method of multiple working hypotheses for hydrological modeling, *Water Resour. Res.*, *47*, W09301, doi:10.1029/2010WR009827.

Cosgrove, B. A., et al. (2003), Land surface model spin-up behavior in the North American Land Data Assimilation System (NLDAS), *J. Geophys. Res.*, *108*(D22), 8845, doi:10.1029/2002JD003316.

Daly, C., R. P. Neilson, and D. L. Phillips (1994), A statistical-topographic model for mapping climatological precipitation over mountainous terrain, *J. Appl. Meteorol.*, *33*(2), 140–158, doi:10.1175/1520-0450(1994)033<0140:Astmfm>2.0.Co;2.

Dickinson, R. E., M. Shaikh, R. Bryant, and L. Graumlich (1998), Interactive canopies for a climate model, *J. Clim.*, *11*(11), 2823–2836.

Ek, M. B., K. E. Mitchell, Y. Lin, E. Rogers, P. Grunmann, V. Koren, G. Gayno, and J. D. Tarpley (2003), Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model, *J. Geophys. Res.*, *108*(D22, 8851), doi:10.1029/2002JD003296.

Ek, M. B., et al. (2011), North American Land Data Assimilation System Phase 2 (NLDAS-2): Development and applications, *GEWEX News*, *21*(2), 6–7.

Entin, J. K., A. Robock, K. Y. Vinnikov, S. E. Hollinger, S. X. Liu, and A. Namkhai (2000), Temporal and spatial scales of observed soil moisture variations in the extratropics, *J. Geophys. Res.*, *105*(D9), 11,865–11,877, doi:10.1029/2000JD900051.

Falloon, P., R. Betts, A. Wiltshire, R. Dankers, C. Mathison, D. McNeall, P. Bates, and M. Trigg (2011), Validation of river flows in HadGEM1 and HadCM3 with the TRIP river flow model, *J. Hydrometeorol.*, *12*(6), 1157–1180, doi:10.1175/2011jhm1388.1.

Fan, Y., and G. Miguez-Macho (2011), A simple hydrologic framework for simulating wetlands in climate and earth system models, *Clim. Dyn.*, *37*(1–2), 253–278, doi:10.1007/s00382-010-0829-8.

Fan, Y., G. Miguez-Macho, C. P. Weaver, R. Walko, and A. Robock (2007), Incorporating water table dynamics in climate modeling: 1. Water table observations and equilibrium water table simulations, *J. Geophys. Res.*, *112*, D10125, doi:10.1029/2006JD008111.

Fekete, B. M., C. J. Vorosmarty, and W. Grabs (2002), High-resolution fields of global runoff combining observed river discharge and simulated water balances, *Global Biogeochem. Cycle*, *16*(3), 1042, doi:10.1029/1999GB001254.

Gentry, L. E., M. B. David, F. E. Below, T. V. Royer, and G. F. Mclsaac (2009), Nitrogen mass balance of a tile-drained agricultural watershed in east-central Illinois, *J. Environ. Qual.*, *38*(5), 1841–1847, doi:10.2134/Jeq2008.0406.

Goswami, D. (2006), Analysis of flow and water quality components in drainage channels in tile drained watersheds, PhD thesis, 154 pp, University of Illinois at Urbana-Champaign.

Güntner, A. (2008), Improvement of global hydrological models using GRACE data, *Surv. Geophys.*, *29*(4-5), 375–397, doi:10.1007/s10712-008-9038-y.

Gutman, G., and A. Ignatov (1998), The derivation of the green vegetation fraction from NOAA/AVHRR data for use in numerical weather prediction models, *Int. J. Remote Sens.*, *19*(8), 1533–1543, doi:10.1080/014311698215333.

Gutowski, W. J., C. J. Vorosmarty, M. Person, Z. Otles, B. Fekete, and J. York (2002), A coupled land-atmosphere simulation program (CLASP): Calibration and validation, *J. Geophys. Res.*, *107*(D16), 4283, doi:10.1029/2001JD000392.

Hogue, T. S., L. A. Bastidas, H. V. Gupta, and S. Sorooshian (2006), Evaluating model performance and parameter behavior for varying levels of land surface model complexity, *Water Resour. Res.*, *42*, W08430, doi:10.1029/2005WR004440.

Kumar, S., and V. Merwade (2011), Evaluation of NARR and CLM3.5 outputs for surface water and energy budgets in the Mississippi River Basin, *J. Geophys. Res.*, *116*, D08115, doi:10.1029/2010JD014909.

Kumar, S. V., C. D. Peters-Lidard, J. Santanello, K. Harrison, Y. Liu, and M. Shaw (2012), Land surface Verification Toolkit (LVT) – A generalized framework for land surface model evaluation, *Geosci. Model Dev.*, *5*(3), 869–886, doi:10.5194/gmd-5-869-2012.

Landerer, F. W., and S. C. Swenson (2012), Accuracy of scaled GRACE terrestrial water storage estimates, *Water Resour. Res.*, *48*, W04531, doi:10.1029/2011WR011453.

Leung, L. R., M. Y. Huang, Y. Qian, and X. Liang (2011), Climate–soil–vegetation control on groundwater table dynamics and its feedbacks in a climate model, *Clim. Dyn.*, *36*(1–2), 57–81, doi:10.1007/s00382-010-0746-x.

Li, H. Y., M. Y. Huang, M. S. Wigmosta, Y. H. Ke, A. M. Coleman, L. R. Leung, A. H. Wang, and D. M. Ricciuto (2011), Evaluating runoff simulations from the Community Land Model 4.0 using observations from flux towers and a mountainous watershed, *J. Geophys. Res.*, *116*, D24120, doi:10.1029/2011JD016276.

Li, H., M. Sivapalan, F. Tian, and D. Liu (2010), Water and nutrient balances in a large tile-drained agricultural catchment: A distributed modeling study, *Hydrol. Earth Syst. Sci.*, *14*(11), 2259–2275, doi:10.5194/hess-14-2259-2010.

Liu, Q., R. H. Reichle, R. Bindlish, M. H. Cosh, W. T. Crow, R. de Jeu, G. J. M. De Lannoy, G. J. Huffman, and T. J. Jackson (2011), The contributions of precipitation and soil moisture observations to the skill of soil moisture estimates in a land data assimilation system, *J. Hydrometeorol.*, *12*(5), 750–765, doi:10.1175/Jhm-D-10-05000.1.

Lo, M. H., J. S. Famiglietti, P. J. F. Yeh, and T. H. Syed (2010), Improving parameter estimation and water table depth simulation in a land surface model using GRACE water storage and estimated base flow data, *Water Resour. Res.*, *46*, W05517, doi:10.1029/2009WR007855.

Miguez-Macho, G., Y. Fan, C. P. Weaver, R. Walko, and A. Robock (2007), Incorporating water table dynamics in climate modeling: 2. Formulation, validation, and soil moisture simulation, *J. Geophys. Res.*, *112*, D13108, doi:10.1029/2006JD008112.

Mitchell, K. E., et al. (2004), The multi-institution North American Land Data Assimilation System (NLDAS): Utilizing multiple GCIP products and partners in a continental distributed hydrological modeling system, *J. Geophys. Res.*, *109*, D07S90, doi:10.1029/2003JD003823.

Moriasi, D. N., J. G. Arnold, M. W. Van Liew, R. L. Bingner, R. D. Harmel, and T. L. Veith (2007), Model evaluation guidelines for systematic quantification of accuracy in watershed simulations, *Trans. ASABE*, *50*(3), 885–900.

Murray, S. J., P. N. Foster, and I. C. Prentice (2011), Evaluation of global continental hydrology as simulated by the Land-surface Processes and eXchanges Dynamic Global Vegetation Model, *Hydrol. Earth Syst. Sci.*, *15*(1), 91–105, doi:10.5194/hess-15-91-2011.

Murray, S. J., I. M. Watson, and I. C. Prentice (2013), The use of dynamic global vegetation models for simulating hydrology and the potential integration of satellite observations, *Prog. Phys. Geogr.*, *37*(1), 63–97, doi:10.1177/0309133312460072.

Nash, J. E., and J. V. Sutcliffe (1970), River flow forecasting through conceptual models, Part 1. A discussion of principles, *J. Hydrol.*, *10*(3), 282–290, doi:10.1016/0022-1694(70)90255-6.

Niu, G.-Y., Z.-L. Yang, R. E. Dickinson, and L. E. Gulden (2005), A simple TOPMODEL-based runoff parameterization (SIMTOP) for use in global climate models, *J. Geophys. Res.*, *110*, D21106, doi:10.1029/2005JD006111.

Niu, G.-Y., Z.-L. Yang, R. E. Dickinson, L. E. Gulden, and H. Su (2007), Development of a simple groundwater model for use in climate models and evaluation with Gravity Recovery and Climate Experiment data, *J. Geophys. Res.*, *112*, D07103, doi:10.1029/2006JD007522.

Niu, G.-Y., et al. (2011), The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements, *J. Geophys. Res.*, *116*, D12109, doi:10.1029/2010JD015139.

Ozdogan, M., M. Rodell, H. K. Beaudoing, and D. L. Toll (2010), Simulating the effects of irrigation over the United States in a land surface model based on satellite-derived agricultural data, *J. Hydrometeorol.*, *11*(1), 171–184, doi:10.1175/2009jhm1116.1.

Pokhrel, Y., N. Hanasaki, S. Koirala, J. Cho, P. J. F. Yeh, H. Kim, S. Kanae, and T. Oki (2012), Incorporating anthropogenic water regulation modules into a land surface model, *J. Hydrometeorol.*, *13*(1), 255–269, doi:10.1175/Jhm-D-11-013.1.

Roads, J., et al. (2003), GCIP water and energy budget synthesis (WEBS), *J. Geophys. Res.*, *108*(D16), 8609, doi:10.1029/2002JD002583.

Rodell, M., J. L. Chen, H. Kato, J. S. Famiglietti, J. Nigro, and C. R. Wilson (2007), Estimating groundwater storage changes in the Mississippi River basin (USA) using GRACE, *Hydrogeol. J.*, *15*(1), 159–166, doi:10.1007/s10040-006-0103-7.

Rosero, E., Z.-L. Yang, T. Wagener, L. E. Gulden, S. Yatheendradas, and G.-Y. Niu (2010), Quantifying parameter sensitivity, interaction, and transferability in hydrologically enhanced versions of the Noah land surface model over transition zones during the warm season, *J. Geophys. Res.*, *115*, D03106, doi:10.1029/2009JD012035.

Running, S. W., D. D. Baldocchi, D. P. Turner, S. T. Gower, P. S. Bakwin, and K. A. Hibbard (1999), A global terrestrial monitoring network integrating tower fluxes, flask sampling, ecosystem modeling and EOS satellite data, *Remote Sens. Environ.*, *70*(1), 108–127, doi:10.1016/S0034-4257(99)00061-9.

Schaake, J. C., V. I. Koren, Q. Y. Duan, K. Mitchell, and F. Chen (1996), Simple water balance model for estimating runoff at different spatial and temporal scales, *J. Geophys. Res.*, *101*(D3), 7461–7475, doi:10.1029/95JD02892.

Schaefer, G. L., M. H. Cosh, and T. J. Jackson (2007), The USDA Natural Resources Conservation Service Soil Climate Analysis Network (SCAN), *J. Atmos. Oceanic Technol.*, *24*(12), 2073–2077, doi:10.1175/2007jtecha930.1.

Sellers, P. J., M. D. Heiser, and F. G. Hall (1992), Relations between surface conductance and spectral vegetation indices at intermediate (100 m^2 to 15 km^2) length scales, *J. Geophys. Res.*, *97*(D17), 19,033–19,059, doi:10.1029/92JD01096.

Sorooshian, S., J. L. Li, K. L. Hsu, and X. G. Gao (2012), Influence of irrigation schemes used in regional climate models on evapotranspiration estimation: Results and comparative studies from California's Central Valley agricultural regions, *J. Geophys. Res.*, *117*, D06107, doi:10.1029/2011JD016978.

Stöckli, R., D. M. Lawrence, G.-Y. Niu, K. W. Oleson, P. E. Thornton, Z.-L. Yang, G. B. Bonan, A. S. Denning, and S. W. Running (2008), Use of FLUXNET in the Community Land Model development, *J. Geophys. Res.*, *113*, G01025, doi:10.1029/2007JG000562.

Swenson, S., and J. Wahr (2006), Post-processing removal of correlated errors in GRACE data, *Geophys. Res. Lett.*, *33*, L08402, doi:10.1029/2005GL025285.

Swenson, S., and P. C. D. Milly (2006), Climate model biases in seasonality of continental water storage revealed by satellite gravimetry, *Water Resour. Res.*, *42*, W03201, doi:10.1029/2005WR004628.

van den Hurk, B., M. Best, P. Dirmeyer, A. Pitman, J. Polcher, and J. Santanello (2011), Acceleration of land surface model development over a decade of GLASS, *Bull. Am. Meteorol. Soc.*, *92*(12), 1593–1600, doi:10.1175/Bams-D-11-00007.1.

van Dijk, A. I. J. M., L. J. Renzullo, and M. Rodell (2011), Use of Gravity Recovery and Climate Experiment terrestrial water storage retrievals to evaluate model estimates by the Australian water resources assessment system, *Water Resour. Res.*, *47*, W11524, doi:10.1029/2011WR010714.

Verdin, K., and S. Jenson (1996), Development of continental scale DEMs and extraction of hydrographic features, in Proceedings of the Third International Conference/Workshop on Integrating GIS and Environmental Modeling, edited, Santa Fe, NM, January 21-26, 1996, National Center for Geographic Information and Analysis, Santa Barbara, CA.

Wahr, J., S. Swenson, V. Zlotnicki, and I. Velicogna (2004), Time-variable gravity from GRACE: First results, *Geophys. Res. Lett.*, *31*, L11501, doi:10.1029/2004GL019779.

Wood, E. F., et al. (2011), Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth's terrestrial water, *Water Resour. Res.*, *47*, W05301, doi:10.1029/2010WR010090.

Xia, Y., et al. (2012a), Continental-scale water and energy flux analysis and validation for North American Land Data Assimilation System project phase 2 (NLDAS-2): 2. Validation of model-simulated streamflow, *J. Geophys. Res.*, *117*, D03110, doi:10.1029/2011JD016051.

Xia, Y., et al. (2012b), Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. Intercomparison and application of model products, *J. Geophys. Res.*, *117*, D03109, doi:10.1029/2011JD016048.

Yang, Z.-L., and G.-Y. Niu (2003), The versatile integrator of surface and atmosphere processes: Part 1. Model description, *Global Planet. Change*, *38*(1–2), 175–189, doi:10.1016/S0921-8181(03)00028-6.

Yang, Z.-L., and R. E. Dickinson (1996), Description of the biosphere-atmosphere transfer scheme (BATS) for the soil moisture workshop and evaluation of its performance, *Global Planet. Change*, *13*(1–4), 117–134, doi:10.1016/0921-8181(95)00041-0.

Yang, Z.-L., R. E. Dickinson, A. Henderson-Sellers, and A. J. Pitman (1995), Preliminary study of spin-up processes in land surface models with the first stage data of Project for Intercomparison of Land Surface Parameterization Schemes Phase 1(a), *J. Geophys. Res.*, *100*(D8), 16,553–16,578, doi:10.1029/95JD01076.

Yang, Z.-L., et al. (2011), The community Noah land surface model with multiparameterization options (Noah-MP): 2. Evaluation over global river basins, *J. Geophys. Res.*, *116*, D12110, doi:10.1029/2010JD015140.