Uncertainty in evapotranspiration from land surface modeling, remote sensing, and GRACE satellites

Di Long, Laurent Longuevergne, Bridget R. Scanlon

AGU PUBLICATIONS

Water Resources Research

RESEARCH ARTICLE

10.1002/2013WR014581

Key Points:
- ET from LSMs, remote sensing and GRACE is evaluated
- Water budget closure using a range of LSM and RS products is performed
- Methods of driving TWSC from GRACE original TWSA are compared

Supporting Information:
- Texas GRACE ET
- Figures S1, S2, S3, S4, S5, S6, S7, S8, S9

Correspondence to:
D. Long,
di.long@beg.utexas.edu

Citation:
Long, D., L. Longuevergne, and B. R. Scanlon (2014), Uncertainty in evapotranspiration from land surface modeling, remote sensing, and GRACE satellites, *Water Resour. Res.*, *50*, doi:10.1002/2013WR014581.

Received 13 AUG 2013
Accepted 12 JAN 2014
Accepted article online 17 JAN 2014

Uncertainty in evapotranspiration from land surface modeling, remote sensing, and GRACE satellites

Di Long[1], Laurent Longuevergne[2], and Bridget R. Scanlon[1]

[1]Bureau of Economic Geology, Jackson School of Geosciences, University of Texas at Austin, Austin, Texas, USA, [2]Geosciences Rennes, UMR CNRS 6118, Université de Rennes 1, Rennes, France

Abstract Proliferation of evapotranspiration (ET) products warrants comparison of these products. The study objective was to assess uncertainty in ET output from four land surface models (LSMs), Noah, Mosaic, VIC, and SAC in NLDAS-2, two remote sensing-based products, MODIS and AVHRR, and GRACE-inferred ET from a water budget with precipitation from PRISM, monitored runoff, and total water storage change (TWSC) from GRACE satellites. The three cornered hat method, which does not require a priori knowledge of the true ET value, was used to estimate ET uncertainties. In addition, TWSC or total water storage anomaly (TWSA) from GRACE was compared with water budget estimates of TWSC from a flux-based approach or TWSA from a storage-based approach. The analyses were conducted using data from three regions (humid-arid) in the South Central United States as case studies. Uncertainties in ET are lowest in LSM ET (~5 mm/mo), moderate in MODIS or AVHRR-based ET (10–15 mm/mo), and highest in GRACE-inferred ET (20–30 mm/month). There is a trade-off between spatial resolution and uncertainty, with lower uncertainty in the coarser-resolution LSM ET (~14 km) relative to higher uncertainty in the finer-resolution (~1–8 km) RS ET. Root-mean-square (RMS) of uncertainties in water budget estimates of TWSC is about half of RMS of uncertainties in GRACE-derived TWSC for each of the regions. Future ET estimation should consider a hybrid approach that integrates strengths of LSMs and satellite-based products to constrain uncertainties.

1. Introduction

Evapotranspiration (ET), the highest outgoing water flux in the hydrological cycle at the global scale, is a primary determinant of water availability together with precipitation [*Long and Singh*, 2012a; *McCabe and Wood*, 2006; *Wang and Dickinson*, 2012]. Accurate knowledge of ET is the linchpin to developing a greater understanding of the water and energy balance of a region. Limited networks of ET monitoring stations (e.g., weighing lysimeters, energy balance Bowen ratio systems, or eddy covariance (EC) towers) globally limit quantification of ET across large areas. Satellite remote sensing (RS) provides an unprecedented opportunity to monitor spatiotemporal variability in ET using two basic approaches: (1) vegetation index-based data: Leaf Area Index (LAI) or Normalized Difference Vegetation Index (NDVI), and (2) land surface temperature (LST). Remotely sensed LAI or NDVI are used with surface resistance in the Penman-Monteith equation to provide global estimates of ET [e.g., *Mu et al*., 2011; *Zhang et al*., 2010]. Issues of spatial scale mismatch between finer vegetation data and coarser meteorological forcing remain unresolved and can result in large uncertainties in ET retrievals [*Yang et al*., 2013]. Remotely sensed LST is used in many algorithms to estimate ET, including the spatial variability models such as SEBAL and triangle models [*Bastiaanssen et al*., 1998; *Jiang and Islam*, 2001], and physically based one-source or two-source models [*Anderson et al*., 2007a; *Long and Singh*, 2012b; *Su*, 2002]. LST-based models have been used to generate ET for the United States [*Anderson et al*., 2007a, 2007b] and for the world [*Vinukollu et al*., 2011], but these ET estimates generally cover a limited time due mostly to cloud cover issues. Estimation of ET using LST requires cloud-free images, which are sometimes difficult to obtain [e.g., *Long and Singh*, 2010; *Nishida et al*., 2003; *Yang and Shang*, 2013]. Many of these algorithms using LST also require a strong contrast in ET within each image [*Long and Singh*, 2013], and are mostly applied to irrigated agricultural regions in semiarid areas [e.g., *Bastiaanssen et al*., 2005; *Tang et al*., 2009].

ET at global to grid scales can also be simulated using land surface models (LSMs), e.g., Noah [*Ek et al*., 2003], Mosaic [*Koster and Suarez*, 1994, 1996], Variable Infiltration Capacity (VIC) [*Liang et al*., 1994],

Sacramento Soil Moisture Accounting (SAC) [*Burnash et al*., 1973], and Common Land Model (CLM) [*Dai et al*., 2003]. The North American Land Data Assimilation System (NLDAS-2) [*Xia et al*., 2012a, 2012b] LSMs (Noah, Mosaic, VIC, and SAC) provide data sets on ET and other fluxes (e.g., runoff and drainage) and state variables (e.g., soil moisture and temperature) at a resolution of 0.125° (~14 km) across the United States from 1979 to present. NLDAS-2 improved the accuracy and consistency of surface forcing data and upgraded the LSM code and parameters from the first phase of NLDAS (NLDAS-1) [*Mitchell et al*., 2004]. The Global Land Data Assimilation System (GLDAS) [*Ek et al*., 2003; *Rodell et al*., 2004b] provides land surface states globally at a spatial resolution of 1° (~111 km) and a near-real-time scale, including output of Noah, Mosaic, VIC, and CLM. Note that the same model (e.g., Noah, VIC) in different data assimilation systems represents different versions of the codes with different forcing and associated parameters for soils and vegetation [*Long et al*., 2013].

Some studies have used Gravity Recovery and Climate Experiment (GRACE) satellite-derived total water storage change (TWSC, dS/dt) to estimate ET as a residual in the water budget equation:

$$\mathrm{ET} = P - R - dS/dt \tag{1}$$

where P is precipitation and R is streamflow [*Ramillien et al*., 2006]. GRACE satellites provide information on changes in the gravity field which are controlled primarily by variations in water distribution and are used to derive TWSC at a spatial resolution of $\geq$200,000 km^2, providing an opportunity to better constrain the water budget equation. TWSC includes changes in surface water, soil moisture, and groundwater storage [*Tapley et al*., 2004].

Water budgets provide a check on the RS products and assess the partitioning of water among different components to evaluate water availability. *Sheffield et al*. [2009] evaluated water budget closure over the Mississippi River basin using various satellite products, including precipitation from Tropical Rainfall Measuring Mission (TRMM), Multisatellite Precipitation Analysis (TMPA), ET from the Penman-Monteith equation applied to MODIS LAI data [*Mu et al*., 2007], and GRACE-derived TWSC. Calculated runoff greatly overestimated monitored runoff because of large positive bias in precipitation [*Gao et al*., 2010; *Sheffield et al*., 2009]. *Sahoo et al*. [2011] and *Pan et al*. [2011] developed water budgets over 10 large river basins globally using a range of satellite and/or ground-based monitoring of P, R, ET, and GRACE-derived TWSC; however, large water budget nonclosure errors ranging from 5% to 25% of P were attributed primarily to errors in satellite-derived P and TWSC.

A variety of approaches are used to evaluate ET products, including comparison with ground-based measurements and comparison of multiple LSM ET products. Comparison of RS ET with ground-based ET monitoring is complicated because ground-based ET often has energy budget nonclosure of up to 30% [*Twine et al*., 2000]. Scale mismatch between satellite or LSM-based pixel resolution (e.g., 1 km for MODIS or ~14 km for LSM output) and the footprint of micrometeorological flux measurements (e.g., typically 100 m for EC towers) makes ground referencing of ET from RS or LSM output difficult [*Gao and Long*, 2008; *McCabe and Wood*, 2006]. ET is also evaluated as a residual of equation (1) (ET = P – R) at multiyear scales with TWSC assumed to be negligible. However, this approach cannot be used to evaluate ET at finer temporal scales (e.g., monthly). Comparison of RS ET from the Penman-Monteith algorithm [*Mu et al*., 2007], Priestley-Taylor algorithm [*Fisher et al*., 2008], and the SEBS model [*Su*, 2002] with 12 EC towers across the United States (2003–2006) showed that the Penman-Monteith ET algorithm was biased low relative to the EC values (bias: ~−18 W m^{-2}), with biases for the Priestley-Taylor algorithm of ~−6 W m^{-2} and a bias for SEBS of ~−8 W m^{-2} [*Vinukollu et al*., 2011]. *Mu et al*. [2011] showed that the Penman-Monteith algorithm reproduced daily ET at 46 EC tower scales globally with a mean absolute percentage difference (MAPD) of ~25%. *Zhang et al*. [2010] generated a global NDVI-based monthly ET product with a MAPD of ~28% relative to 48 EC towers globally. However, information regarding reliability and uncertainties in LSM or RS-based ET products at river basin scales and monthly time scales is limited.

The primary objective of this study is to quantify uncertainties in ET from LSMs in NLDAS-2, from RS (i.e., MODIS and AVHRR) and from GRACE using the three cornered hat method [TCH, *Premoli and Tavella*, 1993], and to evaluate the water budget using LSM, RS, and ground-based monitoring data. TWSC from GRACE was also compared with water budget estimates of TWSC using monitored P, R, and the different ET products investigated in this study. The TCH method ranks uncertainties in the different ET products without any a priori knowledge of the true value of ET or input variables.

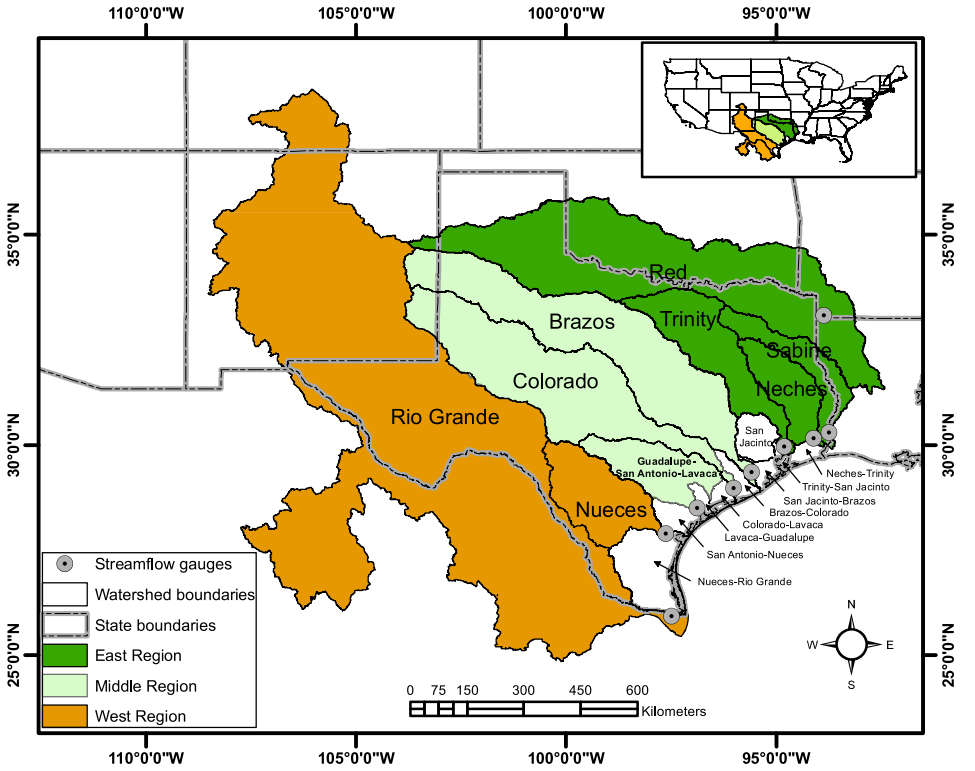

Figure 1. Three regions and USGS streamflow gauges in Texas that are used to estimate outflow of the three regions being investigated.

We compared two approaches of evaluating consistency between GRACE-derived TWSC or total water storage anomaly (TWSA) and water budget estimates of TWSC or TWSA. The first one is called a flux-based approach in which different ET is inserted in equation (1) and *P*-*R*-ET is compared with GRACE-derived TWSC. The second one is called a storage-based approach in which detrended integral of *P*-*R*-ET is compared with detrended GRACE-derived TWSA (details of the storage-based method are given in section 3.3).

Two different sources of GRACE data were used (i.e., Center for Space Research (CSR) at the University of Texas at Austin and the Groupe de Recherche de Géodésie Spatiale (GRGS) analysis center). The latest release (RL05) from CSR with marked reductions in errors relative to RL04 [*Long et al*., 2013] was used. Comprehensive comparisons of the various ET products and quantification of their uncertainties should be valuable for selection of appropriate ET products for water management and for guiding future research in ET estimation.

2. Study Regions and Data Sources

2.1. Study Regions

This study was conducted using data for three major river basins with a broad range of climates and land cover types in the South Central United States (primarily in Texas and surrounding states, Figures 1 and S1). Variability in climate and land cover in the South Central United States makes it a promising test bed for evaluating ET products under varying climate and land cover conditions. Climate ranges from humid in the E [mean annual precipitation, MAP ~940 mm, 1895–2012 climatology from PRISM, *Daly et al*., 2008] composed of the Red, Trinity, Sabine, and Neches River basins (area totaling ~272,000 km^2), semihumid in the M (MAP ~640 mm) comprising the Brazos, Colorado, Guadalupe, San Antonio, and Lavaca River basins (~261,000 km^2), and semiarid in the W (MAP of the U.S. portion of basin ~380 mm) comprising the Rio Grande and Nueces River basins (~598,000 km^2 with 62% in the United States and the remaining in Mexico). Land cover from east to west is characterized by decreasing forestland (E: ~24%, W: 11%) and increasing shrubland (E: ~13%, W: ~64%, supporting information section S1). Coastal river basins totaling nine

Table 1. USGS Gauges Used at the Outlets of Major River Basins in Texas and Related Information

Major River Basins	River Basins	Gauge Number	Latitude	Longitude	Drainage Area (km^2)	Area in Texas (km^2)
East	Red River	07344370	33°05′22″	−93°51′34″	12,26,301	62,963
	Sabine River	08030500	30°18′13″	−93°44′37″	24,162	19,616
	Neches River	08041780	30°09′24″	−94°06′51″	25,353	25,751
	Trinity River	08067070 (Canal)	29°57′40″	−94°48′36″	45,242	46,418
Middle	Brazos River	08116650	29°20′58″	−95°34′56″	117,428	1,11,077
	Colorado River	08162500	28°58′26″	−96°00′44″	109,401	1,02,172
	San Antonio River and Guadalupe River	08188800	28°30′20″	−96°53′04″	26,231	26,259
West	Nueces River	08211500	27°52′58″	−97°37′30″	43,211	43,276
	Rio Grande	Rio Grande near Brownsville, TX and Matamoros, Tamaulipas	25°55′49″	−97°29′4″	598,200	Area in United States: 370,200

(Figure 1) were excluded from this analysis because complete streamflow data for calculating outflows are not available. Partitioning of the South Central United States into the three study regions also considered the footprint of GRACE signals of ∼200,000 km^2 (∼4° × 4° or 440 km × 440 km) [*Longuevergne et al*., 2010].

2.2. Data Sources

2.2.1. Precipitation and Streamflow

Different precipitation sources were available to check their consistency and reliability of deriving ET from GRACE and comparison of water budget estimates of TWSC or TWSA with those from GRACE. Precipitation for the river basins was derived from three data sets: (1) monthly PRISM precipitation data at 4 km (2.5′) resolution [*Daly et al*., 2008], (2) NOAA Climate Prediction Center (CPC) 0.25° × 0.25° (∼28 km) daily U.S. unified precipitation data [*Higgins et al*., 2000], and (3) TMPA (3B-43) computed at monthly intervals at 0.25° × 0.25° resolution [*Huffman et al*., 2007]. Daily NOAA CPC data were aggregated into monthly values. PRISM data were unavailable for the Mexican part of the Rio Grande basin; therefore, TMPA was used in this region.

Streamflow data for several gauges at the outlets of the three regions were obtained from the USGS (Table 1). Outflow data of the Rio Grande are obtained from the website (http://www.ibwc.gov/wad/DDQBROWN.htm). The Rio Grande has much lower outflow at the gauge station in Brownsville, TX (the nearest gauge station to the Gulf of Mexico) and does not even reach the Gulf of Mexico as a result of a combined effect of climate impacts (e.g., drought) and human activities (e.g., over appropriation of water) [*Wurbs*, 2006].

2.2.2. Evapotranspiration

Three different ET products were evaluated in this study, i.e., (1) LSM ET, (2) RS ET, and (3) GRACE-inferred ET. Four ET outputs from LSMs in NLDAS-2 were used with subscripts indicating the LSMs, including ET_{Noah}, ET_{Mosaic}, ET_{VIC}, and ET_{SAC}. All of the LSMs characterize ET using soil moisture stress impacts on evaporation from the top layer of the soil profile and vegetation transpiration. For instance, evaporation rates from the first layer of soil in the Noah model can proceed only at the rate at which the top soil layer can transfer water upward from below [*Chen et al*., 1996]. Under extremely wet conditions, evaporation from the soil surface can take place at the potential rate. Vegetation transpiration from the root zone is constrained by: (1) canopy interception and (2) canopy resistance that is explicitly parameterized by constraints of soil moisture and the ambient environment, e.g., air temperature, vapor pressure deficit, and solar radiation [*Koster and Suarez*, 1994]. These LSMs models have different versions, structures, and associated parameters embedded in the codes, e.g., the Noah model in NLDAS-2 has four soil layers with spatially invariant thicknesses of 10, 30, 60, and 100 cm. The first three layers form the root zone in nonforested regions, with the fourth layer included in forest regions. However, the Mosaic model in NLDAS-2 accounts for the subgrid heterogeneity of vegetation and soil moisture with a tiling approach. Up to 10 tiles can be used in the current configuration of Mosaic. The model has three layers with thicknesses of 10, 30, and 160 cm, and the first two of which fall within the root zone. Mosaic has a greater ability to transfer water from the deep layer to the surface, and therefore shows higher ET rates under normal conditions [*Long et al*., 2013]. More details about these LSMs in NLDAS-2 can be found in *Wei et al*. [2013] and *Xia et al*. [2012a, 2012b]. Note that the aim of this study is not to investigate performance of different LSMs, but to provide objective information on uncertainties in their ET output.

RS ET products include ET_{MODIS} (http://www.ntsg.umt.edu/project/mod16#data-product) and ET_{AVHRR} (http://www.ntsg.umt.edu/project/et#data-product). $ET_{MODIS,}$ spanning from 2000 to 2012, was derived from MODIS-based phenological and surface variables (e.g., LAI, fraction of absorbed photosynthetically active radiation, enhanced vegetation index, surface albedo, and land cover) and a daily meteorological reanalysis data set (e.g., radiation and air temperature) from NASA's Global Modeling and Assimilation Office (GMAO) in combination with the Penman-Monteith equation [*Mu et al*., 2011]. $ET_{AVHRR,}$ spanning from 1983 to 2006, was derived from NOAA-AVHRR Global Inventory Modeling and Mapping Studies (GIMMS) NDVI, NCEP/NCAR Reanalysis (NNR) daily surface meteorology, and NASA/Global Energy and Water-cycle Experiment (GEWEX) Surface Radiation Budget Release-3.0 solar radiation in combination with a modified Penman-Monteith equation [*Zhang et al*., 2010]. Both ET_{MODIS} and ET_{AVHRR} relate canopy resistance to LAI [*Mu et al*., 2011] or NDVI [*Zhang et al*., 2010] based on the framework proposed by *Fisher et al*. [2008]. ET_{AVHRR} since 2007 was generated using ET_{MODIS} time series and mean monthly ET ratios (spatially averaged first) between ET_{AVHRR} and ET_{MODIS} for the overlapping period from 2000 to 2006.

GRACE-inferred ET (six outputs) was derived from water budgets (equation (1)) using *P* from PRISM, *R* from USGS monitoring data, and TWSC from the first-order derivative of GRACE-derived TWSA (see section 3.2). Regional-scale LSM or RS ET was derived from aggregation of distributed ET values across a basin; however, GRACE-inferred ET is an integrated value at the regional scale that does not contain detailed information on spatial distribution of ET.

2.2.3. Total Water Storage Change

GRACE measures TWSA in which the reference is the mean gravity field of a study period. Therefore, GRACE TWSA depends on the reference related to a time period [*Yeh et al*., 2006]. TWSC is the time derivative of TWSA, which is referred specifically to as *dS*/*dt* at the monthly scale. GRACE satellite data from CSR and GRGS analysis centers were used to derive TWSA and TWSC at a monthly scale from January 2003 to September 2012 for the three regions. CSR and GRGS represent two end-members for GRACE processing. CSR is one of the least constrained solutions [*Bettadpur*, 2007] and GRGS is one of the most constrained solutions [*Bruinsma et al*., 2010]. Use of both products provides valuable information on uncertainty in GRACE TWSA. The latest release of CSR data (RL05) was used in the analysis. Details of the processing are provided in supporting information section S2. Uncertainties in GRACE-derived TWSA include: (1) inherent uncertainty in GRACE data and (2) propagation of bias/leakage correction using the additive correction method [*Longuevergne et al*., 2010] due to uncertainties in a priori soil moisture storage (SMS) changes in GLDAS-1. Uncertainties in SMS changes were estimated from the standard deviation of SMS changes among four LSMs (i.e., Noah, Mosaic, VIC, and CLM) in GLDAS-1. All data used in this study are listed in Table 2.

3. Methods

3.1. Three Cornered Hat Method

Traditional methods of evaluating uncertainties in heat fluxes assume errors in input variables (e.g., LST or NDVI) to be normally distributed. The total uncertainty (standard deviation) in heat fluxes is therefore the root of the sum of the square of partial derivative of heat fluxes relative to an input variable multiplied by the uncertainty in the variable [e.g., *Marx et al*., 2008].

A variety of ET products are available (e.g., RS, LSM, and GRACE); however, the true value of ET is unknown. Uncertainties in RS or LSM output are often conservatively estimated by evaluating variability among different outputs (standard deviation of different outputs), e.g., soil moisture storage from LSMs [e.g., *Longuevergne et al*., 2010; *Rodell et al*., 2009]. The three cornered hat method (TCH) [*Premoli and Tavella*, 1993], also called the Grubb's estimator [*Grubbs*, 1948], can be used to estimate relative uncertainties in different products without a priori knowledge of actual ET when at least three different sets of data (e.g., LSM, RS, and GRACE-inferred ET in this study) are available. The theory of the TCH is described in the following and based on the assumption that observational errors are normally distributed. A given set of observations obs_i consists of two components: the true value, *x*, and an associated measurement error, e_i:

$$obs_i = x + e_i \tag{2}$$

Given a set of three pairs of observations (*i*, *j*, *k*), the difference between observations (*i*, *j*) can be written as:

Table 2. Data Sources of Hydrological Fluxes and State Variables Used in This Study

Flux/State Variables	Sources	Spatial Resolution	Spatial Extent	Temporal Resolution	Time Span	References
P	1. PRISM	4 km (~2.5′)	Contiguous United States (24° to 50° Latitude; −125° to −66.5° Longitude)	Monthly	1890–present	*Daly et al.* [2008]
	2. NOAA CPC	0.25°	Contiguous United States (20° to 50° Latitude; −130° to −55° Longitude)	Daily: 1948–2006 Real time: 2007–present	1948–present	
	3. TRMM 3B43	0.25°	Quasi –global (−50° to 50° Latitude; −180° to 180° Longitude)	Monthly	1998–present	*Huffman et al.* [2007]
ET	1. MODIS data-based using the Penman-Monteith approach	1 km, 0.005°, and 0.05°	Global (−60° to 80° Latitude; −180° to 180° Longitude)	Monthly	2000–2012	*Mu et al.* [2011]
	2. NOAA-AVHRR data-based using the Penman-Monteith approach	8 km and 1°	Global (25° to 53° Latitude; −63° to 89° Longitude)	Monthly	1983–2006, undergoing extension to 2011	*Zhang et al.* [2010]
	3. Output of Noah from the NLDAS-2	0.125°	Contiguous United States (25° to 53° Latitude; −125° to 67° Longitude)	Hourly/monthly	1979–present	NLDAS-2, *Xia et al.* [2012a, 2012b]
	4. Output of Mosaic from the NLDAS-2	0.125°	Contiguous United States (25° to 53° Latitude; −125° to 67° Longitude)	Hourly/monthly	1979–present	NLDAS-2, *Xia et al.* [2012a, 2012b]
	5. Output of VIC from the NLDAS-2	0.125°	Contiguous United States (25° to 53° Latitude; −125° to 67° Longitude)	Hourly/monthly	1979–present	NLDAS-2, *Xia et al.* [2012a, 2012b]
	6. Output of SAC from the NLDAS-2	0.125°	Contiguous United States (25° to 53° Latitude; −125° to 67° Longitude)	Hourly/monthly	1979–present	NLDAS-2, *Xia et al.* [2012a, 2012b]
TWS	1. GRACE CSR	~20,000 km^2	Global	Monthly	2003–present	*Bettadpur* [2007]
	2. GRACE GRGS	~20,000 km^2	Global	10 days	2003–present	*Bruinsma et al.* [2010]
R	1. USGS gauges stations	Basin integrated	United States	Daily	2003–present	http://waterdata.usgs.gov/nwis/sw

$$obs_i - obs_j = x + e_i - (x + e_j) = e_i - e_j \tag{3}$$

The associated variance of the differences can be written as:

$$\sigma_{ij}^2 = \sigma_{ei}^2 + \sigma_{ej}^2 - 2\mathrm{cov}\,(e_i, e_j) \tag{4}$$

If errors between estimates of *i* and *j* are independent, the cov(e_i, e_j) equals zero. Finally, the individual variances σ_{ei}^2 may be separated by:

$$\sigma_{ei}^2 = \frac{1}{2}\left(\sigma_{ij}^2 + \sigma_{ik}^2 - \sigma_{jk}^2\right) \tag{5}$$

Because the three observations in the TCH approach could be correlated (e.g., GRACE-inferred ET and LSMs ET have similar precipitation input), it is important to consider cross correlation among the three observations (i.e., cov(e_i, e_j) is not necessarily zero). When considering three signals, the standard deviation of the signal differences provides three equations. If cross correlation (uncorrelated errors) is not considered, one has only three unknowns that could be solved based on equation (5). If cross correlation is considered, there will be six unknown variables to determine. *Tavella and Premoli* [1991] proposed to add mathematical characteristics of the covariance matrices, i.e., its positive definiteness that the determinant is positive, or, in other words, the determined variances should be all positive so as to solve the problem. This generalization was further developed by *Premoli and Tavella* [1993] and *Tavella and Premoli* [1994]. In this study, we used the TCH approach that considers cross correlation among observations and does not require that the data sources are entirely independent [*Premoli and Tavella*, 1993].

3.2. Different Methods of Calculating Total Water Storage Change

ET inferred from GRACE TWSC was estimated using equation (1). As GRACE provides noisy monthly TWSA, the computation of monthly TWSC to approximate *dS/dt* is not straightforward and potentially affected with large noise. Three derivative methods were used to estimate TWSC at a monthly scale (mm/mo) by computing: (i) the time derivative of TWSA using the simple derivative in equation (6), (ii) the double difference derivative in equation (7) that includes some light numerical smoothing, or (iii) the smoothed differentiation filter that can be regarded as a low pass differentiation filter [*Luo et al*., 2004] with a time shift of 1 month:

$$\frac{dS}{dt} \approx \frac{dTWSA}{dt} \approx \frac{TWSA(t)-TWSA(t-1)}{\Delta t} \tag{6}$$

$$\frac{dS}{dt} \approx \frac{dTWSA}{dt} \approx \frac{TWSA(t+1)-TWSA(t-1)}{2\Delta t} \tag{7}$$

where Δt is the temporal sampling of GRACE TWSA. The two GRACE solutions (CSR and GRGS) times the three derivative methods resulted in six GRACE-inferred ET estimates. The TCH method was subsequently applied by selecting one ET estimate from each of these three ET estimates (i.e., output from (1) LSMs, (2) RS, and (3) GRACE-inferred ET). The analysis results in six GRACE-inferred ET estimates, four LSM ET estimates (Noah, Mosaic, VIC, and SAC in NLDAS-2), and two RS ET products (MODIS and AVHRR) for each of the three study regions used to evaluate uncertainties in these ET products.

3.3. Water Budget Calculation

In this study, two methods of calculating water budgets were tested. The first is a traditional flux-based method (equation (1)), in which *dS/dt* is computed by the three ways introduced in section 3.2. Details of *P*, *R*, and ET at monthly scales (mm/mo) are described in sections 2.2.1 and 2.2.2. The second method is a storage-based method in equation (8), in which the cumulative sum of *P*, *R*, and ET during the study period is calculated.

$$S(t)-S_0(t_0)=\int_{t=0}^{t=t}[P(t)-R(t)-\mathrm{ET}\,(t)]dt \tag{8}$$

where $S(t)$ or $S_0(t_0)$ is the total water storage at time t or t_0, respectively. LSM and RS ET examined in this study are inserted in equation (2) in combination with monitored *P* and *R*. Note that if *P* and *R* are considered unbiased, any systematic bias in ET will result in a trend in $S(t)$-$S_0(t_0)$. The trend of the time series of $S(t)$-$S_0(t_0)$ is subsequently subtracted to derive TWSA for a study period that is compared with detrended GRACE-derived TWSA.

As the computation of the time derivative of noisy GRACE TWSA tends to amplify the high frequency noise, it is expected that the flux-based approach would provide ET directly but would be very noisy. Conversely, the storage-based approach requires temporal integration of fluxes (and ET), and any systematic bias in a flux would be converted into a trend in terms of storage, which could be adequately evaluated with GRACE.

3.4. Uncertainty Analysis in Water Budget

Uncertainty in water budget estimates of TWSC is quantified by assuming that errors in *P*, *R*, and ET are independent and normally distributed [*Rodell et al*., 2004a]:

$$\upsilon_{ds/dt}=\frac{\sqrt{\upsilon_P^2P^2+\upsilon_R^2R^2+\upsilon_{\mathrm{ET}}^2\,\mathrm{ET}^2}}{|P-R-\mathrm{ET}|} \tag{9}$$

where υ is relative uncertainty for each component (subscripts denote corresponding water components). Uncertainties in GRACE-derived TWSC were computed from uncertainties in GRACE TWSA for back and forward months added in quadrature. Uncertainties in both water budget estimates of TWSC and GRACE-derived TWSC are taken as the 95% confidence limits on TWSC, i.e., $\pm\upsilon_{ds/dt}$TWSC. A flow diagram showing

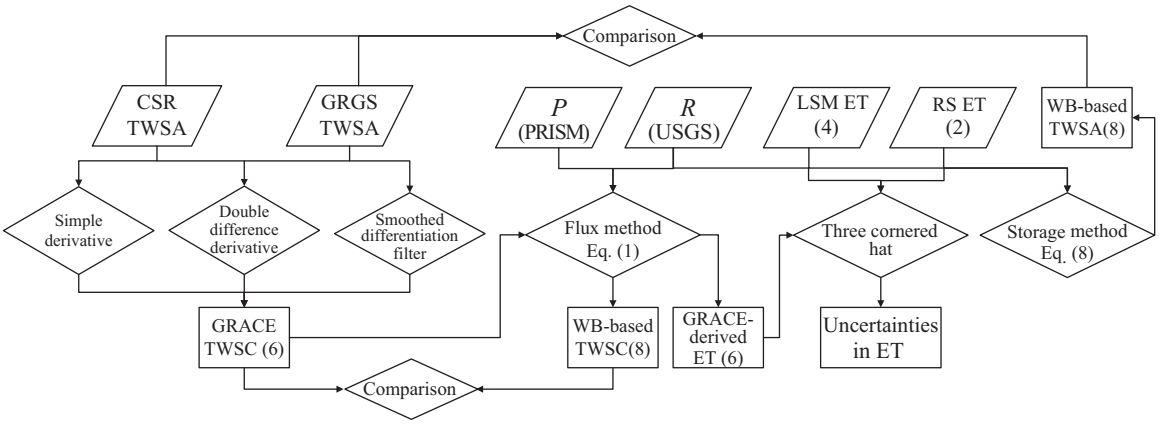

Figure 2. A flow diagram showing uncertainty analysis of different ET products being investigated and comparison of water budget estimates of TWSC/TWSA with GRACE-derived TWSC/TWSA. Trapezoids denote data sources, diamonds denote processing, and rectangles denote output. Numbers are the quantities of different products or outputs.

uncertainty analysis of the various ET products and comparison of water budget estimates of TWSC or TWSA with GRACE-based counterparts in this study is given in Figure 2.

4. Results and Discussion

Seasonal cycles of monitored *P* and *R* can be helpful for understanding seasonal cycles of ET for the three study regions being investigated. Details on seasonal cycles of *P* and a comparison among the three *P* products are provided in supporting information section S3 and Figure S2. In the following discussion, PRISM *P* was used in water budget calculations. TRMM *P* was only used for the Mexican portion of the west region.

4.1. Comparison of Land Surface Modeling and Remote Sensing Evapotranspiration Products

ET products within the same category (i.e., LSMs or RS) are similar in magnitude but differ markedly between categories, with values of RS ET generally lower than LSM ET (Figures 3 and 4). Mean annual RS ET from 2003 to 2011 is 19%, 27%, and 6% lower than LSM ET for the E, M, and W, respectively. Annual ET is highest for the Mosaic LSM model, and lowest for the MODIS sensor (Table 3). There are also appreciable differences in monthly and mean monthly ET estimates for the different ET products, with ET_{MODIS} being lowest and ET_{VIC} being highest during warm seasons (Figures 3, S3, and S4), e.g., for the M region, the largest difference in mean monthly ET estimates is ~33 mm in April ($ET_{MODIS} < ET_{VIC}$, Figure 4), and the largest difference in monthly ET time series between ET_{MODIS} and ET_{VIC} is up to ~50 mm in May 2005 (Figure S3). Differences in ET estimates from different products are generally larger in warm seasons than in cold seasons, and are larger in more humid regions (M and E) than in semiarid regions (W). RS ET is generally lower than LSM ET in most cases except under extremely dry conditions (e.g., the 2011 drought, see the grey background in Figures 3, S3, and S4). Seasonal cycles in ET are highly correlated with periodic variations in radiative energy in a year (Figure 4). In general, ET estimates from the four LSMs peak in May/June (in the E and M) or July/August (in the W) and are lowest in December. In May/June, surface net radiation is highest, maximizing potential ET that is only a function of meteorological variables (e.g., temperature and vapor pressure deficit) and is determined largely by radiation [*Long and Singh*, 2010].

ET_{MODIS} is lower than ET_{AVHRR}. Relatively lower magnitudes of ET_{MODIS} than other RS ET products (e.g., Priestley-Taylor ET, SEBS ET, International Satellite Cloud Climatology Project ET) were also observed across 10 large river basins globally [*Sahoo et al.*, 2011]. In addition, significant negative bias of RS ET estimates (lower than the ground-based observations) from the modified Penman-Monteith algorithm [*Mu et al.*, 2007] was found over 12 EC towers from the FLUXNET global network [*Vinukollu et al.*, 2011]. Relatively larger magnitudes of ET_{Mosaic} than other LSM ET in the cold seasons could be ascribed to greater diffusion of water from deeper soil layers to the shallow root zone [*Mitchell et al.*, 2004]. Few studies report the contrast between LSM ET and RS ET during extremely dry or wet conditions. These differences could be related to energy balance or water balance constraints. LSMs are based on SVAT schemes, in which actual ET is constrained directly by soil moisture simulations in addition to meteorological forcing. In contrast, RS ET is

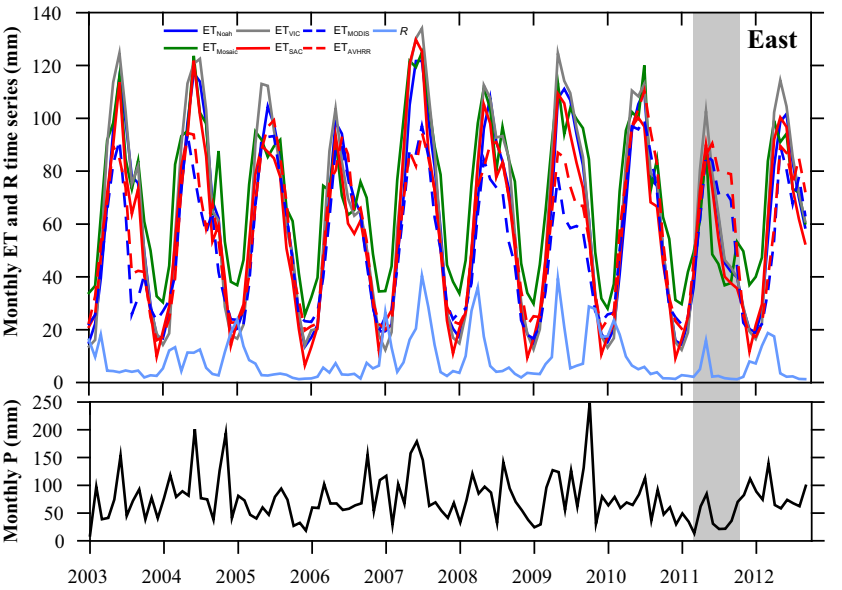

Figure 3. Time series of monthly ET from four land surface models (i.e., Noah, Mosaic, VIC, and SAC) and two satellite-based models (i.e., MODIS and NOAA-AVHRR), and corresponding precipitation and outflow for the east region from January 2003 to September 2012. The gray background shows the drought period in 2011 defined in *Long et al.* [2013].

determined largely by radiation, without explicit soil moisture constraints based on water balance, though soil moisture effects are implicitly accommodated by LAI or NDVI and atmospheric variables. During droughts, soil moisture is often greatly depleted and exercises constraints on simulated ET from LSMs; however, net radiation remains relatively high, thereby resulting in higher actual ET estimates from RS than from LSM ET and a time lag of decreases in RS ET compared with LSM ET. Decreases in RS ET during the drought could also be related to depleted soil moisture that is reflected in decreases in LAI/NDVI and increases in LST. In addition, increases in ET due to irrigation could be reflected in RS ET corresponding to relatively higher LAI/NDVI for irrigated areas than nonirrigated areas. However, this is not accounted for and simulated by LSMs.

The RS-based ET products are less responsive to precipitation than those from LSMs, except Mosaic, generally resulting in a more dampened seasonal cycle than the LSM ET estimates (e.g., coefficients of variation (CV) for the E region: ET_{Noah} (0.55); ET_{Mosaic} (0.39); ET_{VIC} (0.58); ET_{SAC} (0.55); ET_{MODIS} (0.48); ET_{AVHRR} (0.48)). Correlation coefficients (r) between P and the various ET products at annual and monthly scales indicate that LSM ET is more highly correlated with P than RS ET, e.g., r between monthly LSM ET and P is $\geq$0.75, but r between RS ET and P is $\leq$0.57 in the E (Table 4). This may result from different algorithms, inputs, and assumptions used to estimate ET. RS-based ET estimates are determined primarily by meteorological factors (e.g., vapor pressure

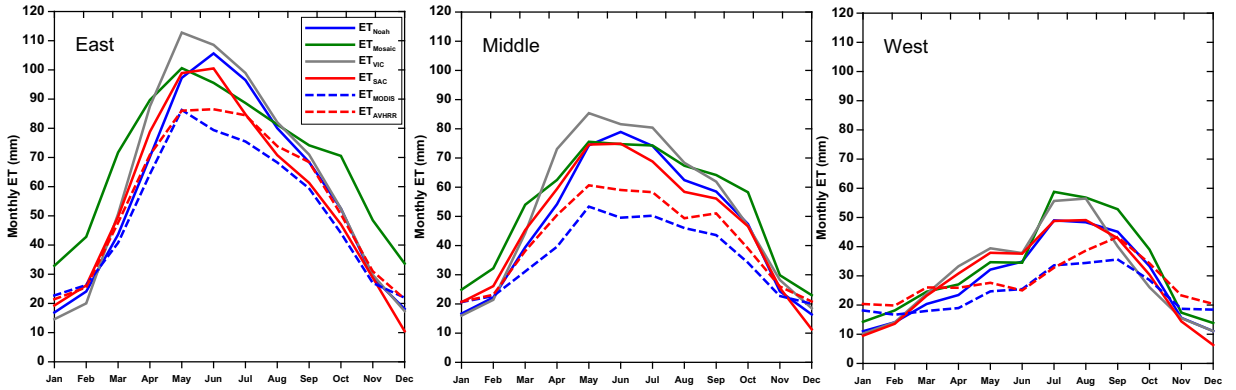

Figure 4. Mean monthly ET of east, middle, and west regions from four land surface models (i.e., Noah, Mosaic, VIC, and SAC) in NLDAS-2 and two satellite-based models (i.e., MODIS and NOAA-AVHRR) derived from January 2003 to September 2012.

Table 3. Annual ET Estimates From Four LSMs and Two Satellite ET Products for the East (E), Middle (M), and West (W) Regions Over the Central South United States, with Showing Annual *P* from PRISM of the Three Regions[a]

Region	Year	*P*	Noah	Mosaic	VIC	SAC	MODIS	AVHRR
E	2003	782	704	825	760	663	630	664
	2004	1197	777	912	811	752	684	685
	2005	673	683	806	724	641	602	700
	2006	918	657	740	667	588	539	642
	2007	1092	792	978	852	811	726	782
	2008	927	708	885	769	701	618	666
	2009	1136	763	903	810	731	609	656
	2010	799	703	861	747	676	688	742
	2011	618	541	578	577	526	460	490
	Average	*905*	*703*	*832*	*746*	*677*	*617*	*670*
	ET/P	–	**0.78**	**0.92**	**0.82**	**0.75**	**0.68**	**0.74**
M	2003	558	574	639	634	548	445	494
	2004	1008	695	829	722	697	522	549
	2005	545	608	706	704	619	445	530
	2006	571	519	554	557	487	364	473
	2007	932	697	841	777	727	574	655
	2008	502	537	588	597	540	398	454
	2009	664	552	609	607	546	393	447
	2010	686	615	711	676	614	521	595
	2011	325	358	341	395	359	265	299
	Average	*643*	*573*	*646*	*630*	*571*	*436*	*500*
	ET/P	–	**0.89**	**1.00**	**0.98**	**0.89**	**0.68**	**0.78**
W	2003	349	352	406	379	347	303	330
	2004	565	427	535	421	406	356	369
	2005	369	362	424	410	391	301	351
	2006	388	314	357	340	321	271	348
	2007	514	406	478	436	414	374	438
	2008	360	323	383	357	341	287	338
	2009	334	318	347	338	321	248	293
	2010	417	372	453	403	381	347	405
	2011	212	216	216	229	216	181	213
	Average	*390*	*343*	*400*	*368*	*349*	*296*	*343*
	ET/P	–	**0.88**	**1.03**	**0.94**	**0.89**	**0.76**	**0.88**

[a]Bold rows indicate the ratio of mean annual ET to mean annual precipitation for different regions being studied.

deficit and air temperature) and phenological parameters which are physically and mathematically related to remotely sensed LAI or NDVI. The effect of soil moisture on evaporation from the soil surface and vegetation transpiration is not explicitly depicted in the ET algorithms, but is assumed to be encapsulated in vegetation greenness and meteorological variables through a complementary relationship [*Fisher et al.*, 2008; *Mu et al.*, 2011]. Apparently, meteorological factors and LAI or NDVI are indirectly related to precipitation. In contrast, LSM ET is parameterized within the SVAT schemes and constrained by the water balance equation. Evaporation from the soil surface in LSMs is explicitly linked to soil moisture content by the soil diffusivity formulation or fraction of soil moisture saturation in the upper soil layer [*Ek et al.*, 2003]. Surface conductance for vegetation transpiration explicitly accounts for soil moisture stress, in addition to ambient environmental and phenological factors. Therefore, the LSM ET estimates tend to be more responsive to soil moisture and consequently precipitation compared with those from RS-based approaches.

4.2. Uncertainties in ET Products Using the Three Cornered Hat Method

4.2.1. GRACE TWSA and TWSC for the Three Study Regions

ET from GRACE depends largely on reliability of TWSC. TWSA from CSR and GRGS centers are highly correlated (E: 0.96; M: 0.93; and W: 0.87), providing confidence in use of GRACE to derive monthly TWSC. ET was subsequently inferred using GRACE-derived TWSC and equation (1). GRACE TWSC appears to be noisier and have lower amplitudes than GRACE TWSA because GRACE TWSA is the primitive integral of TWSC (Figure 5). The magnitudes of uncertainty in GRACE-derived TWSC (RMS error: E ~34 mm/mo, M ~28 mm/mo, and W ~20 mm/mo) are amplified compared with those in GRACE-derived TWSA (RMS errors: E ~24 mm, M ~20 mm, and W ~14 mm). The signal-to-noise ratio (STNR), calculated as the standard deviation of the TWSC time series divided by the standard deviation of the error time series, is highest in the E (4.1), lowest in the W (2.9), and moderate (3.3) in the M. This means that GRACE signals and consequently the inferred ET could be more reliable in relatively humid regions than in relatively dry regions.

Table 4. Correlation Coefficient Matrixes of Time Series of *P* and Different ET Products for the East (E), Middle (M), and West (W) Region[a]

Region	Variable	*P*	Noah	Mosaic	VIC	SAC	MODIS	AVHRR
E	*P*	1.00	0.83	0.75	0.77	0.80	0.57	0.45
	Noah	0.32	1.00	0.98	0.98	0.96	0.89	0.82
	Mosaic	0.46	0.92	1.00	0.99	0.97	0.92	0.87
	VIC	0.31	0.98	0.94	1.00	0.98	0.90	0.81
	SAC	0.37	0.97	0.95	0.98	1.00	0.90	0.79
	MODIS	0.31	0.97	0.93	0.97	0.96	1.00	0.92
	AVHRR	0.27	0.97	0.92	0.97	0.95	0.98	1.00
M	*P*	1.00	0.89	0.89	0.81	0.89	0.85	0.78
	Noah	0.58	1.00	1.00	0.98	0.99	0.96	0.93
	Mosaic	0.70	0.94	1.00	0.98	0.99	0.96	0.93
	VIC	0.51	0.97	0.92	1.00	0.98	0.95	0.95
	SAC	0.62	0.98	0.96	0.97	1.00	0.96	0.93
	MODIS	0.59	0.94	0.92	0.92	0.93	1.00	0.97
	AVHRR	0.53	0.94	0.92	0.94	0.94	0.97	1.00
W	*P*	1.00	0.93	0.94	0.85	0.87	0.91	0.84
	Noah	0.86	1.00	0.99	0.97	0.97	0.96	0.87
	Mosaic	0.91	0.97	1.00	0.96	0.96	0.96	0.87
	VIC	0.75	0.94	0.89	1.00	1.00	0.96	0.92
	SAC	0.78	0.97	0.92	0.98	1.00	0.95	0.91
	MODIS	0.86	0.89	0.92	0.79	0.82	1.00	0.95
	AVHRR	0.77	0.82	0.85	0.72	0.77	0.93	1.00

[a]The upper triangular matrixes indicate correlation coefficients at the yearly scale, and lower triangular matrixes indicate correlation coefficients at the monthly scale.

4.2.2. Uncertainties in Varying ET Products

Uncertainties in ET quantified by the TCH approach are lowest in LSMs (~5 mm/mo in the three river basins), moderate in RS-based ET products (~10 to ~15 mm/mo), and highest in GRACE-inferred ET estimates (W ~20 mm/mo to E ~30 mm/mo) (Figure 6). Uncertainties in GRACE-inferred ET quantified in this study have a similar magnitude to those (~26 mm/mo) in *Rodell et al*. [2004a].

Regarding uncertainties in LSM ET in the E and M regions, ET_{Noah} has the lowest uncertainties (E: ~3 mm/mo; M: ~2 mm/mo) and ET_{VIC} has the highest uncertainties (E: ~9 mm/mo; M: ~7 mm/mo). In the W region, uncertainties in ET_{Noah} and ET_{Mosaic} (~6 mm/mo) seem to be higher than those in ET_{VIC} and ET_{SAC} (~4 mm/mo) (Figure 7). For RS-based ET products in the E and M regions, ET_{MODIS} has higher uncertainties (E: ~10.5 mm/mo; M: ~14 mm/mo) than ET_{AVHRR} (E: ~10 mm/mo; M: ~12 mm/mo), but in the W river basin ET_{MODIS} (~11 mm/mo) has slightly lower uncertainties than ET_{AVHRR} (~12 mm/mo) (Figure 8). Regarding the GRACE-derived ET estimates, the double difference derivative method results in the lowest uncertainty (~24 mm/mo) compared with the simple derivative (~28 mm/mo) and smoothed differentiation filter methods (~25) (Figure 9).

These results confirm analyses in section 4.1 that RS ET shows low magnitudes compared with LSM ET output, resulting in statistically significantly different uncertainties between the two types of ET products. The statistically higher uncertainty in RS ET is likely because RS ET products are not constrained by the water balance. In addition, errors in relatively coarse meteorological forcing (e.g., $1^\circ \times 1.25^\circ$ for the Global Modeling and Assimilation Office (GMAO) data set in ET_{MODIS}) may be amplified at relatively longer time scales (i.e., monthly) and river basin scales, though ET_{MODIS} and ET_{AVHRR} incorporate remotely sensed LAI or NDVI of high spatial resolution (~ 1 km). ET output from LSMs has the lowest uncertainties but also has relatively coarse spatial resolution (~14 km). GRACE-inferred ET has the coarsest resolution and the highest uncertainty of all ET products; therefore, use of GRACE TWSC to infer river basin-scale ET may be problematic and warrants further study.

4.3. Comparison of GRACE and Water Budget Estimates of Total Water Storage Changes

In the following sections, discussion is based on CSR TWSA in the three regions (Figure 5, left), and results from both CSR and GRGS centers are presented in Tables 5 and 6. GRACE-derived TWSC from the simple

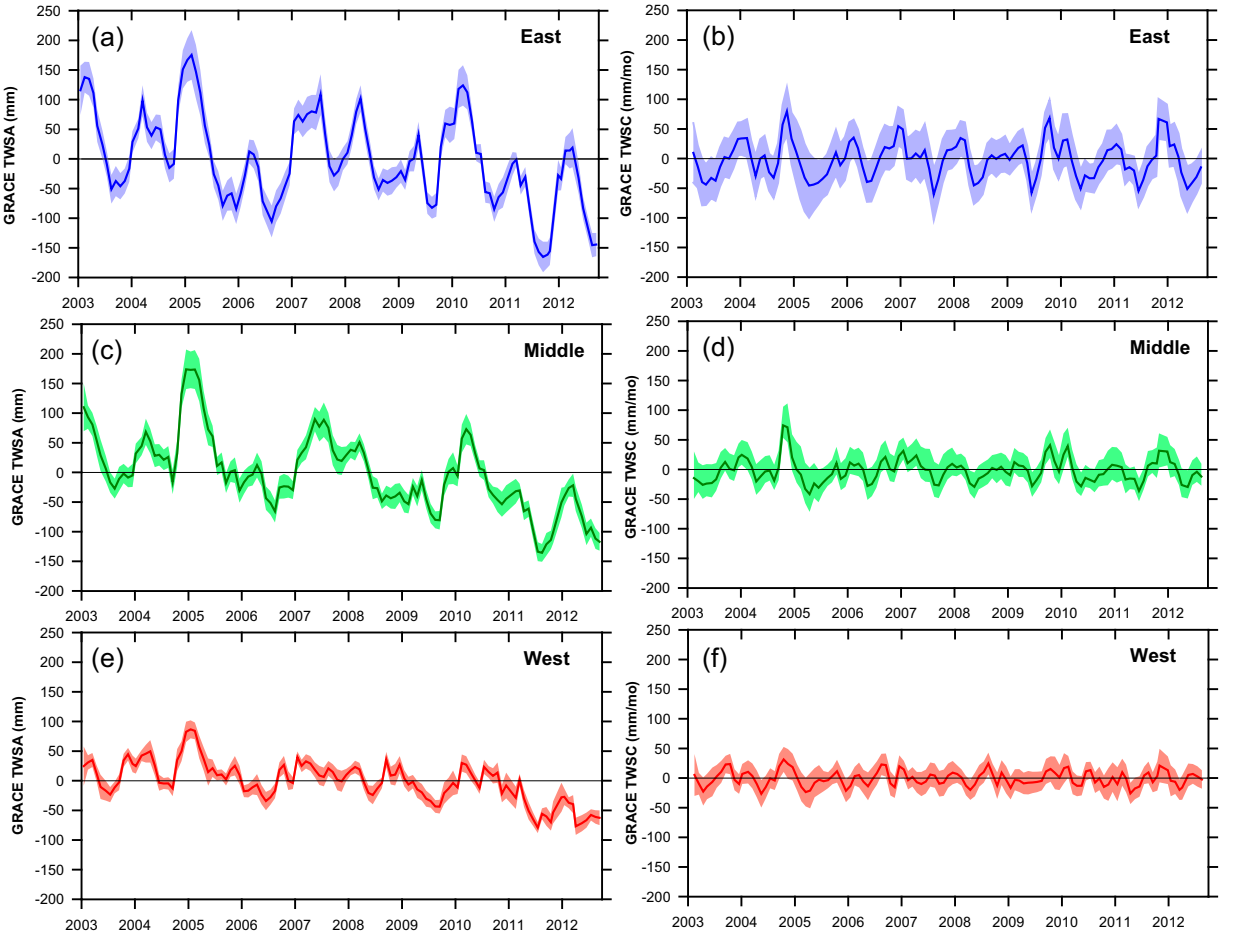

Figure 5. Time series of CSR TWSA and TWSC from the double difference derivative method for the east, middle, and west regions. Backgrounds indicate uncertainties in each time series.

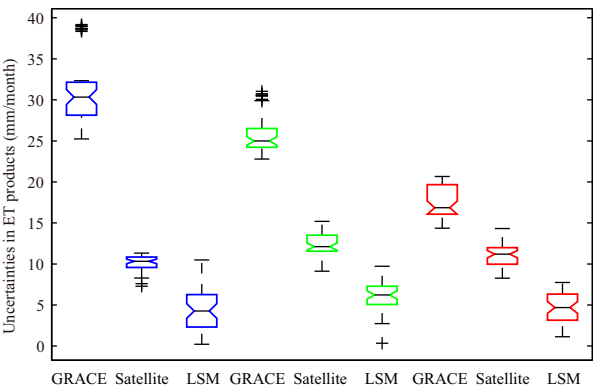

Figure 6. Uncertainties in GRACE-inferred ET estimates, LSM ET, and RS ET products. Blue boxes denote uncertainties for the east region, green boxes denote uncertainties for the middle region, and red boxes denote uncertainties for the west regions. On each box, the central mark is the median (q_2), the edges of the box are the 25th (q_1) and 75th (q_3) percentiles, and the whiskers extend to the most extreme data points ($q_3 + 1.5 \times$ interquartile range ($q_3 - q_1$) and $q_1 - 1.5 \times$ interquartile range ($q_3 - q_1$)) not considered outliers, and outliers are plotted individually. If the notches of two boxes do not overlap, this indicates a statistically significant difference between the medians.

derivative method, the double difference derivative method, and the smoothed differentiation filter were compared with water budget estimates of TWSC for the three river basins, showing that the double difference derivative method results in the highest correspondence between GRACE-derived TWSC and water budget estimates of TWSC. In section 4.3, only the GRACE-derived TWSC from the double difference derivative method (Figure 5, right) is discussed.

4.3.1. Comparison Among Regions

Correspondence between GRACE-derived TWSC and *P-R*-ET generally decreases from E to W, with

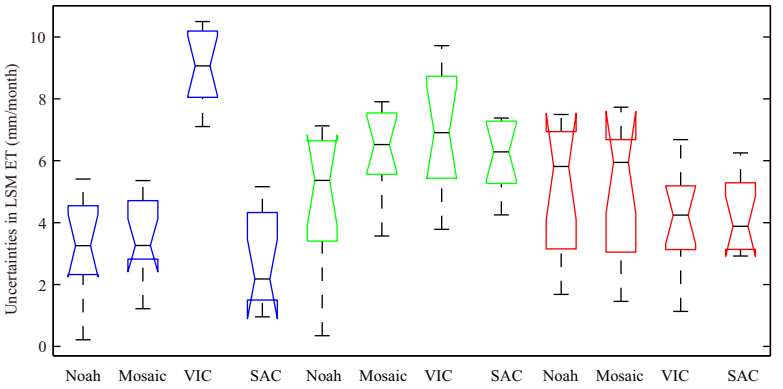

Figure 7. Uncertainties in different ET products from LSMs. Blue boxes denote uncertainties for the east region, green boxes denote uncertainties for the middle region, and red boxes denote uncertainties for the west region. On each box, the central mark is the median (q_2), the edges of the box are the 25th (q_1) and 75th (q_3) percentiles, and the whiskers extend to the most extreme data points ($q_3 + 1.5 \times$ interquartile range ($q_3 - q_1$) and $q_1 - 1.5 \times$ interquartile range ($q_3 - q_1$)) not considered outliers, and outliers are plotted individually. If the notches of two boxes do not overlap, this indicates a statistically significant difference between the medians.

the mean coefficient of determination (R^2) of 0.59 (E), 0.34 (M), and 0.32 (W) (Table 5 and Figures 10, S5, and S6). Relatively lower R^2 in the W river basins may be related to larger uncertainties in ET and *P*, and lower signal-to-noise ratio of GRACE-derived TWSC (STNR for W: 2.9) than those in wetter climate (STNR for E: 4.1). For instance, mean CV among the six ET products over the study period is slightly higher in the W (0.23) than in the M (0.19) or E (0.19), suggesting that variations among the ET products increase with aridity. Increasing discrepancy among ET products indicates increasing uncertainties in ET estimates. Furthermore, the density of NOAA meteorological stations is much lower in the W ($1.17/10^4$ km^2) than in the M ($3.42/10^4$ km^2) or E ($3.85/10^4$ km^2). Relatively sparse precipitation in west Texas may therefore be responsible for larger uncertainties in water budget estimates of TWSC. Higher correlation between *P* and ET in the W than the M and E (Table 4) indicates that ET in the W tends to be soil moisture-limited. More uncertainties in *P* would result in higher uncertainties in ET and therefore in the water budget estimates of TWSC in the W.

Root-mean-square-difference (RMSD) for all ET products is generally higher in the E (~28 mm/mo) and lower in the W (~18 mm/mo) (Table 5), which may be related to the wetter climate and more extreme precipitation events (also see the higher ends of Figure 10). Under these extreme conditions, RMSD of TWSC between GRACE and water budget estimates could increase, thereby resulting in generally larger RMSD in the E. In addition, more anthropogenic influences (e.g., reservoir regulation and water diversions) would also increase uncertainties in outflows and consequently water budget calculations in the E. The lack of accounting for subsurface flow in equation (1) could also cause the relatively larger RMSD in the wet climate. Unlike the E, the M and W have relatively lower streamflow and are less regulated (see Figure S7 in supporting information), which dampens the magnitudes of both RMSD and bias in water budget estimates of TWSC compared with GRACE-based counterparts.

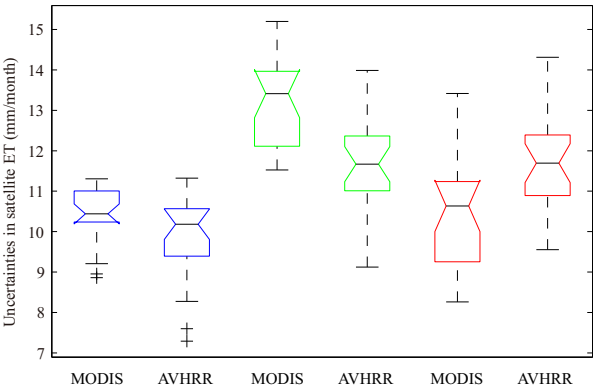

Figure 8. Uncertainties in different RS ET products. Blue boxes denote uncertainties for the east region, green boxes denote uncertainties for the middle region, and red boxes denote uncertainties for the west region. On each box, the central mark is the median (q_2), the edges of the box are the 25th (q_1) and 75th (q_3) percentiles, and the whiskers extend to the most extreme data points ($q_3 + 1.5 \times$ interquartile range ($q_3 - q_1$) and $q_1 - 1.5 \times$ interquartile range ($q_3 - q_1$)) not considered outliers, and outliers are plotted individually. If the notches of two boxes do not overlap, this indicates a statistically significant difference between the medians.

4.3.2. Comparison Between the Flux-Based and Storage-Based Methods

GRACE-derived TWSA is more highly correlated with water

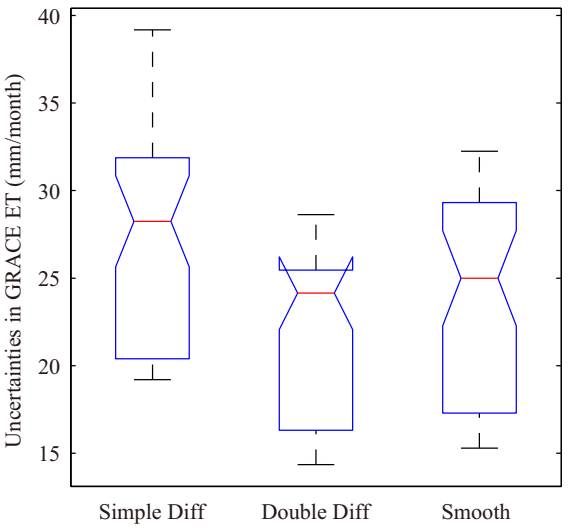

Figure 9. Uncertainties in GRACE-inferred ET estimates in all regions using different methods of calculating TWSC (i.e., simple derivative, double difference derivative, and smoothed differentiation filter). On each box, the central mark is the median (q_2), the edges of the box are the 25th (q_1) and 75th (q_3) percentiles, and the whiskers extend to the most extreme data points ($q_3 + 1.5 \times$ interquartile range ($q_3 - q_1$) and $q_1 - 1.5 \times$ interquartile range ($q_3 - q_1$)) not considered outliers, and outliers are plotted individually. If the notches of two boxes do not overlap, this indicates a statistically significant difference between the medians.

budget estimates of TWSA using the storage-based method ($n = 117$, R^2: 0.69 for E, 0.70 for M, and 0.49 for W) than the flux-based method (R^2: 0.59 for E, 0.34 for M, and 0.32 for W) (Tables 5 and 6; Figures 11, S8, and S9). Improvements in consistency between GRACE-derived TWSA and water budget estimates of TWSA are likely due to two factors: (1) high frequency signals in GRACE-derived TWSC are amplified in the flux-based method; and (2) systematic biases in *P*, *R*, ET, and GRACE-derived TWSA were reduced by removing the linear trends in these components in the storage-based method. RMSD between TWSA from the storage-based method and TWSA from GRACE decreases from the E (~47 mm), M (~38 mm), and W (~34 mm), similar to the trend for TWSC (E: ~28 mm/mo; M: ~26 mm/mo; W: ~18 mm/mo). Note that high R^2 does not necessarily mean low RMSD, e.g., two highly correlated variables distant from the 1:1 line could show both high R^2 and RMSD.

4.3.3. Comparison Among Different Evapotranspiration Products

The RMSD between water budget estimates of TWSC using LSM ET and GRACE-derived TWSC is generally lower than TWSC using RS-based products for all three regions (Figure 12a). For instance, the RMSD between water budget estimates of TWSC using LSM ET and GRACE-derived TWSC is ~27 mm/mo, but the RMSD between water budget estimates of TWSC using RS ET (ET_{MODIS} and ET_{AVHRR}) and GRACE-derived TWSC is ~33 and ~31 mm/mo in the E, respectively (Table 5). The RMSD between water budget estimates

Table 5. Statistics of Discrepancies Between GRACE-Derived TWSC (Double Difference Derivative) and *P*-*R*-ET at the Monthly Scale[a]

		RMSD (mm/month)		Bias (mm/month)		R^2		R_p (Pearson)		R_k (Kendall)	
Region	ET Products	CSR	GRGS	CSR	GRGS	CSR	GRGS	CSR	GRGS	CSR	GRGS
East	ET_{Noah}	27.45	29.66	10.73	10.14	0.64	0.56	0.80	0.75	0.60	0.54
	ET_{Mosaic}	24.81	27.24	0.34	−0.18	0.53	0.43	0.73	0.66	0.52	0.44
	ET_{VIC}	27.31	30.77	6.97	6.31	0.67	0.56	0.82	0.75	0.62	0.53
	ET_{SAC}	27.61	30.43	12.89	12.41	0.64	0.52	0.80	0.72	0.59	0.51
	ET_{MODIS}	32.50	34.52	18.18	17.74	0.53	0.43	0.73	0.66	0.53	0.46
	ET_{AVHRR}	30.54	32.40	13.78	13.32	0.55	0.47	0.74	0.69	0.55	0.49
	Average	*28.37*	*30.84*	*10.48*	*9.96*	*0.59*	*0.50*	*0.77*	*0.70*	*0.57*	*0.49*
Middle	ET_{Noah}	24.04	24.75	4.89	4.52	0.38	0.33	0.62	0.57	0.40	0.38
	ET_{Mosaic}	21.54	21.79	−1.01	−1.32	0.35	0.31	0.59	0.55	0.37	0.34
	ET_{VIC}	24.75	26.21	0.05	−0.37	0.43	0.35	0.66	0.60	0.43	0.39
	ET_{SAC}	23.28	23.89	5.05	4.78	0.38	0.33	0.62	0.57	0.39	0.37
	ET_{MODIS}	32.01	31.53	16.51	16.27	0.21	0.20	0.45	0.45	0.25	0.27
	ET_{AVHRR}	28.91	28.69	11.16	10.90	0.27	0.26	0.52	0.51	0.30	0.31
	Average	*25.76*	*26.14*	*6.11*	*5.80*	*0.34*	*0.30*	*0.58*	*0.54*	*0.36*	*0.34*
West	ET_{Noah}	17.31	18.29	8.29	8.57	0.32	0.34	0.57	0.58	0.39	0.42
	ET_{Mosaic}	13.22	15.49	3.73	3.97	0.32	0.30	0.56	0.55	0.39	0.38
	ET_{VIC}	16.86	18.52	6.11	6.34	0.40	0.33	0.63	0.58	0.42	0.40
	ET_{SAC}	17.41	18.24	7.74	8.02	0.39	0.38	0.62	0.62	0.44	0.44
	ET_{MODIS}	22.95	22.42	12.50	12.77	0.23	0.33	0.48	0.57	0.30	0.40
	ET_{AVHRR}	21.63	20.93	8.64	8.96	0.24	0.33	0.49	0.58	0.32	0.41
	Average	*18.23*	*18.98*	*7.84*	*8.11*	*0.32*	*0.34*	*0.56*	*0.58*	*0.38*	*0.41*

[a]Statistical significance for the Pearson correlation coefficient is less than 0.001.

Table 6. Statistics of Discrepancies Between GRACE-Derived TWSA and Integrated *P-R*-ET Derived From the Storage Method at the Monthly Scale[a]

		RMSD (mm)		R^2		R^p (Pearson)		R^k (Kendall)	
Region	ET Products	CSR	GRGS	CSR	GRGS	CSR	GRGS	CSR	GRGS
East	ET_{Noah}	37.97	45.81	0.80	0.72	0.89	0.85	0.72	0.64
	ET_{Mosaic}	48.84	55.25	0.57	0.45	0.76	0.67	0.56	0.49
	ET_{VIC}	44.50	54.39	0.77	0.66	0.88	0.81	0.70	0.62
	ET_{SAC}	49.21	56.57	0.66	0.56	0.81	0.75	0.62	0.54
	ET_{MODIS}	50.87	55.91	0.65	0.59	0.81	0.77	0.60	0.55
	ET_{AVHRR}	50.11	55.17	0.67	0.62	0.82	0.78	0.61	0.57
	Average	*46.91*	*53.85*	*0.69*	*0.60*	*0.83*	*0.77*	*0.63*	*0.57*
Middle	ET_{Noah}	28.91	35.61	0.81	0.72	0.90	0.85	0.71	0.63
	ET_{Mosaic}	28.55	32.46	0.69	0.59	0.83	0.77	0.60	0.54
	ET_{VIC}	28.80	38.77	0.86	0.71	0.93	0.84	0.74	0.63
	ET_{SAC}	30.17	35.06	0.77	0.70	0.88	0.83	0.66	0.62
	ET_{MODIS}	57.35	56.25	0.49	0.53	0.70	0.73	0.49	0.54
	ET_{AVHRR}	52.33	55.18	0.57	0.53	0.75	0.73	0.51	0.52
	Average	*37.68*	*42.22*	*0.70*	*0.63*	*0.83*	*0.79*	*0.62*	*0.58*
West	ET_{Noah}	29.95	26.11	0.54	0.64	0.73	0.80	0.53	0.59
	ET_{Mosaic}	22.92	26.16	0.40	0.44	0.64	0.67	0.45	0.49
	ET_{VIC}	31.04	27.73	0.63	0.67	0.79	0.82	0.58	0.61
	ET_{SAC}	31.69	27.87	0.60	0.66	0.78	0.81	0.56	0.60
	ET_{MODIS}	41.99	35.07	0.40	0.58	0.63	0.76	0.43	0.57
	ET_{AVHRR}	48.88	43.47	0.34	0.46	0.58	0.68	0.38	0.49
	Average	*34.41*	*31.07*	*0.49*	*0.57*	*0.69*	*0.76*	*0.49*	*0.56*

[a]Statistical significance for the Pearson correlation coefficient is less than 0.001. Biases here are zero (not shown) because of the comparisons of detrended TWSA from the storage-based method and detrended TWSA observed by GRACE.

of TWSC using LSM ET and GRACE-derived TWSC is ~23 mm/mo for the M, which is much smaller than *P-R*-ET_{MODIS} of ~32 mm/mo or *P-R*-ET_{AVHRR} of ~29 mm/mo. For the W, the RMSD for LSMs ET is ~16 mm/mo, which is also lower than that for the RS-based ET products of ~22 mm/mo. Similar trends in R^2 were also

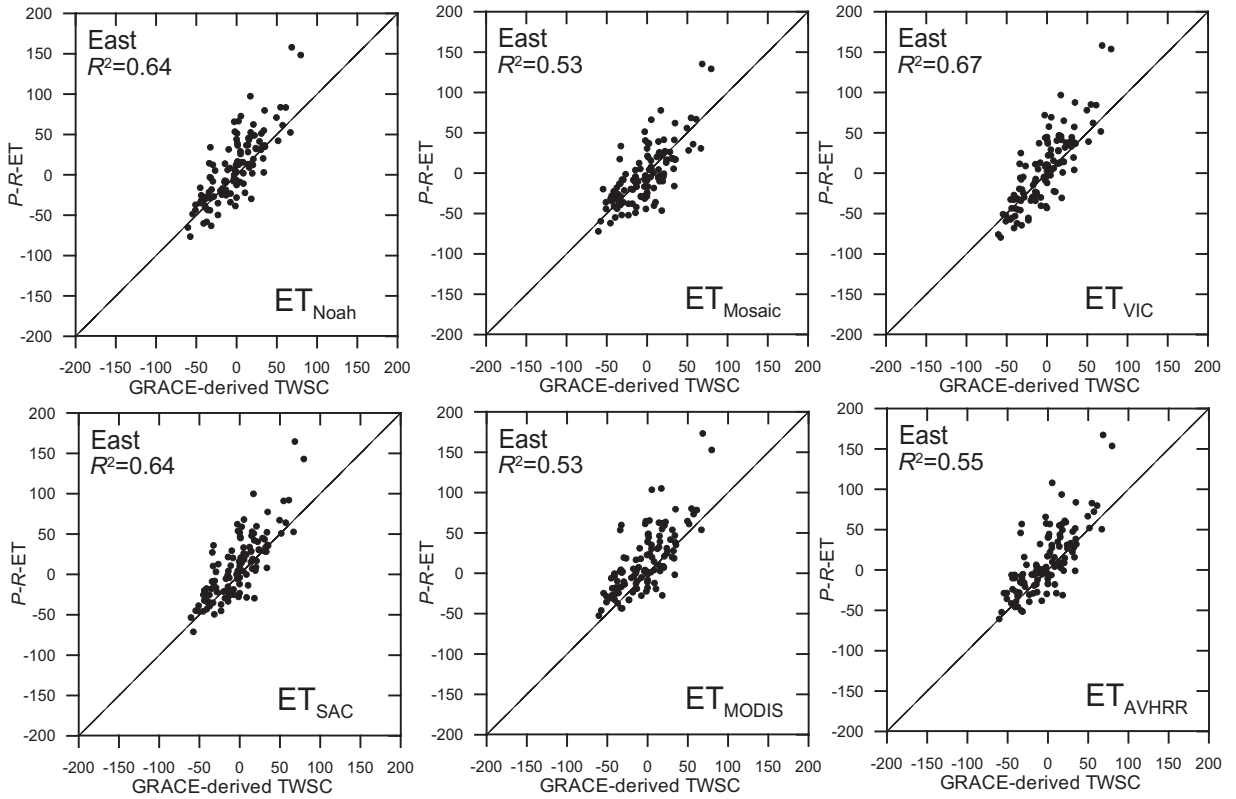

Figure 10. Scatterplots of GRACE-based TWS change from the double difference derivative method and water budget estimates of TWS change using ET from Noah, Mosaic, VIC, and SAC land surface models in NLDAS-2, and from MODIS-based and AVHRR-based sensor for the east river basins.

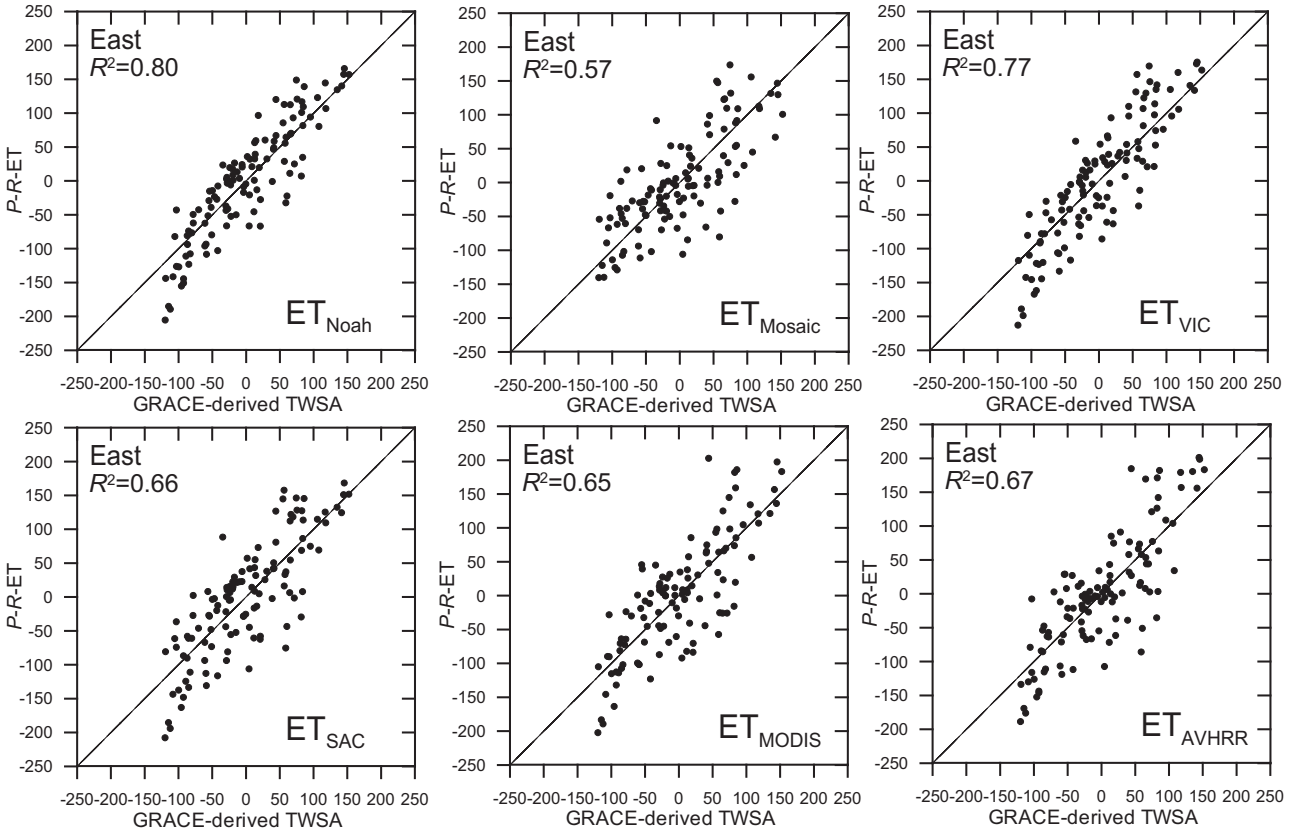

Figure 11. Scatterplots of GRACE TWSA and TWSA from the storage method using ET estimates from Noah, Mosaic, VIC, and SAC in NLDAS-2 and MODIS-based and AVHRR-based ET for the east region.

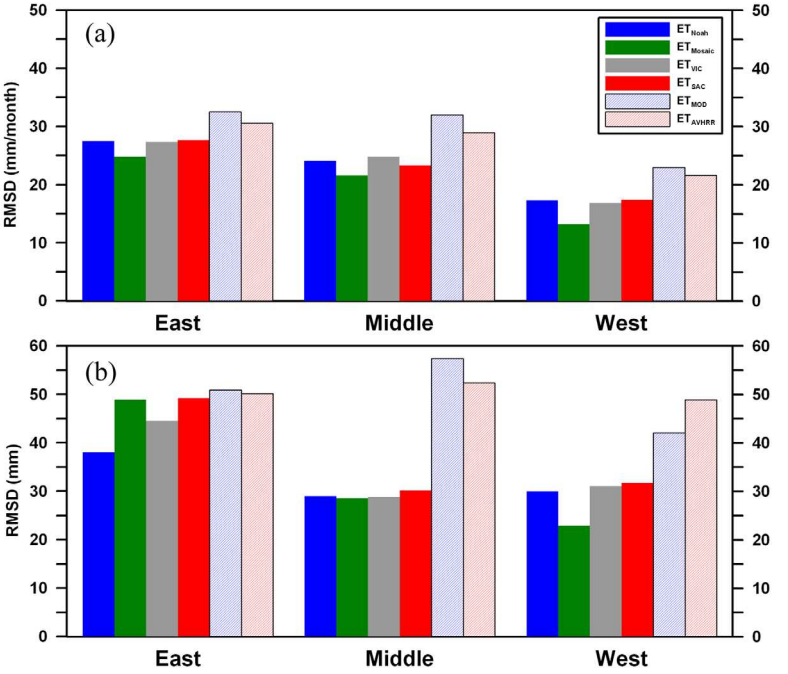

Figure 12. RMSD between (a) TWSC or (b) TWSA using water budgets and different ET products and GRACE-derived TWSC from the double difference derivative method or TWSA for the east, middle, and west regions in the South Central United States.

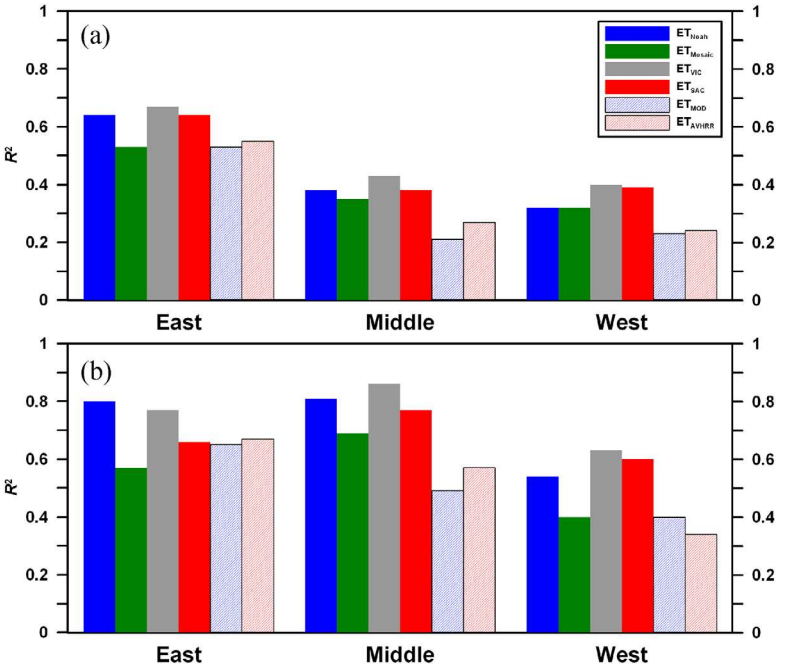

Figure 13. Coefficients of determination between (a) TWSC or (b) TWSA using water budgets and different ET products and GRACE-derived TWSC from the double difference derivative method or TWSA for the east, middle, and west regions in the South Central United States.

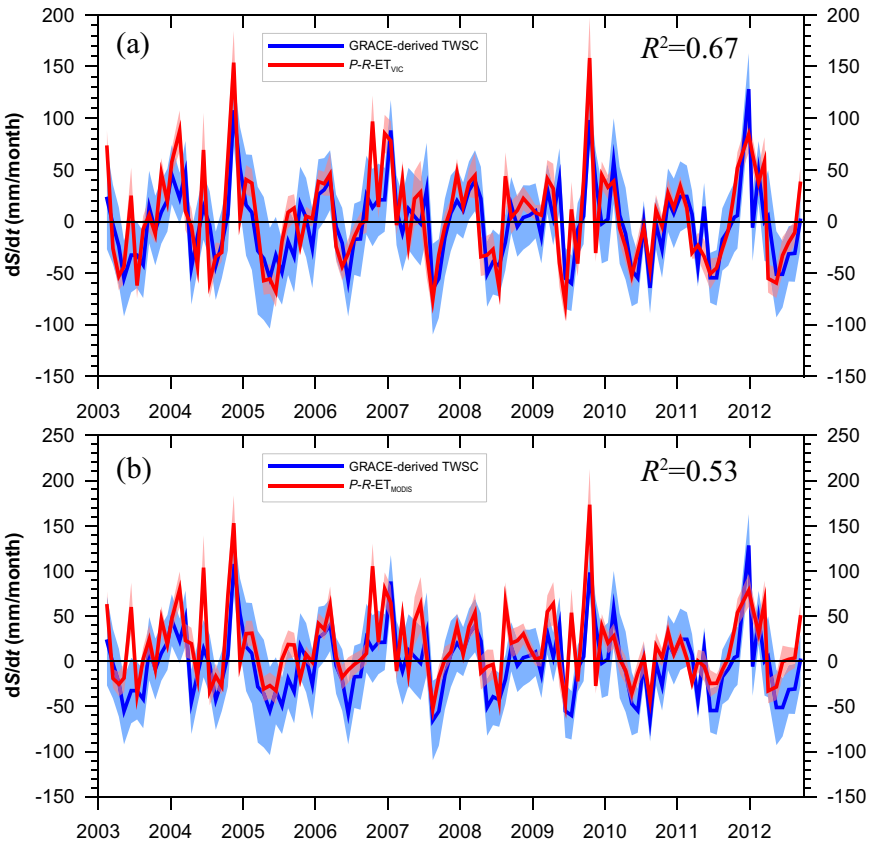

Figure 14. Time series of GRACE-derived TWSC from the double difference derivative method and TWSC using water budgets and (a) ET_{VIC} and (b) ET_{MODIS} for the east region. Blue shading areas represent uncertainties in GRACE-derived TWSC, and red shading areas represent uncertainties in TWSC using water budgets.

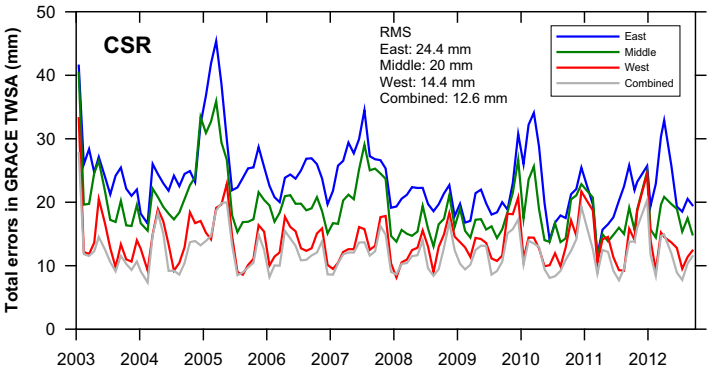

Figure 15. Time series of total error in GRACE-derived TWSA for the east, middle, west, and combined regions.

found, i.e., higher RMSD often corresponds to lower R^2 for the same region. For the E, R^2 between water budget estimates of TWSC using LSM ET except ET_{Mosaic} and GRACE-derived TWSC is on the order of ~0.6; however, R^2 for TWSC using MODIS or AVHRR is on the order of ~0.5 (Figure 13a).

Similar contrast in RMSD and R^2 between LSM ET and RS ET was also found in the comparison between GRACE-derived TWSA and water budget estimates of TWSA from the storage-based method. LSM ET, except ET_{Mosaic}, displays lower RMSD and higher R^2 than RS ET, especially for the M and W (Figures 12b and 13b). Furthermore, ET_{VIC}, ET_{Noah}, and ET_{SAC} show higher R^2 in all regions than ET_{Mosaic} (Tables 5 and 6; Figure 13). ET_{AVHRR} has a higher R^2 than ET_{MODIS} in the E and M (Tables 5 and 6; Figure 13).

4.3.4. Uncertainties in Water Budget Closure

Uncertainties in water budget estimates of TWSC were estimated from uncertainties in each term in equation (9). Here relative uncertainty in PRISM *P* was assumed to be 15% [*Jeton et al.*, 2005] and was prespecified for *R* as 5% [*Rodell et al.*, 2004a]. Intercomparison of ET_{Noah}, ET_{Mosaic}, ET_{VIC}, and ET_{SAC} from NLDAS was performed over four quadrants of the United States [*Mitchell et al.*, 2004; *Xia et al.*, 2012b]. There is, however, far less information regarding relative uncertainties in LSM ET. Uncertainties in monthly ET_{MODIS} and ET_{AVHRR} were reported as ~25% [*Mu et al.*, 2011] and ~28%, respectively [*Zhang et al.*, 2010]. Here uncertainty of ET time series was estimated from the standard deviations of the six ET products for each month. Finally, uncertainties in GRACE-derived TWSC and those in water budgets were compared. Comparisons between GRACE-derived TWSC and water budget estimates of TWSC using ET_{VIC} (highest R^2) or ET_{MODIS} (lowest R^2) for the E are given in Figure 14.

RMS of uncertainties in GRACE-derived TWSC for the E of 34 mm/mo, M of 28 mm/mo, and W of 20 mm/mo regions are two times those for water budget estimates of TWSC for E of ~16 mm/mo, M of ~14 mm/mo, and W of ~10 mm/mo. The water budget estimates of TWSC using any ET product (examples for ET_{VIC} and ET_{MODIS}) are generally within the uncertainties (shaded areas) of GRACE-derived TWSC for the E (Figure 14). For M and W, the consistency between GRACE-derived and water balance estimates of TWSC is lower as reflected by lower R^2 and discussed in section 4.3.1. In addition, uncertainties in ET quantified from standard deviations of different ET products for each month may be underestimated because of correlation among different LSM ET or RS ET products due to similarity in forcing data and algorithms. Therefore, uncertainties in water budget estimates of TWSC may be larger than inferred from the analysis in this study.

To investigate the influence of basin size on uncertainty in GRACE-derived TWSA, time series of uncertainties in GRACE TWSA for the three study regions, and the entire study region were compared (Figure 15). It is apparent that as the size of region increases, RMS error of uncertainties in GRACE TWSA decreases (E: ~24 mm; M: ~20 mm; W: ~14 mm; combined river basin: ~13 mm). However, the consistency (R^2) between GRACE-derived TWSC and water budget estimate of TWSC for the entire study region did not greatly improve, with all statistical metrics similar to those for the M region. This is likely because uncertainties in *P*, *R*, and ET increase with the size of the study region.

5. Conclusions

This study shows that uncertainties in ET are lowest (~5 mm/mo) in LSM products, moderate in RS products (~10–15 mm/mo), and highest in GRACE-inferred estimates (~20–30 mm/mo) using the three cornered hat

method. There is a tradeoff between spatial resolution and uncertainty, with lower uncertainty for LSM ET and coarser resolution (~14 km) relative to higher uncertainty and finer resolution (~1–8 km) for RS ET. RS ET estimation is not constrained by a water balance, showing generally lower magnitudes most of the study period but higher magnitudes during drought relative to LSM ET. ET_{Noah} has the lowest uncertainty in LSM ET products, and ET_{MODIS} has higher uncertainty than ET_{AVHRR} in RS ET.

RS ET products have lower magnitudes and are less responsive to precipitation than LSM ET. Mosaic generated the highest annual ET magnitudes due to the highest ET estimates during cold seasons in all LSMs. In general, the water budget estimates of TWSC or TWSA appear more consistent with GRACE-derived estimates in humid river basins (R^2 ~0.6 or 0.7) than in semihumid river basins (R^2 ~0.3 or 0.7) and arid/semiarid basins (R^2 ~0.3 or 0.5). TWSC or TWSA derived from the water balance and LSM ET are generally closer to GRACE-derived TWSC or TWSA than those derived from the water balance and RS ET. TWSC or TWSA using the water balance and ET_{Noah} and ET_{VIC} shows a generally higher correlation, and TWSC or TWSA using ET_{MODIS} shows the lowest correlation with the GRACE-derived counterpart. Water budget estimates of TWSC or TWSA using ET_{Mosaic} show the lowest correlation with GRACE-derived counterparts in all LSMs.

Improvements are continually being made in each water balance product. For example, NLDAS-2 represents significant improvements over NLDAS-1 as noted previously. Further improvements could be made by developing a global hyperresolution (e.g., 1 km) model as suggested by *Wood et al.* [2011]. However, "grand challenges" of such an effort include quantification of influences of fine-scale topography and vegetation, improved representation of land-atmospheric interactions and resulting spatial information on soil moisture and ET, etc. Satellite-based approaches for ET modeling lack a water-balance constraint and therefore lead to higher uncertainties than LSM ET, especially during drought. Overestimation in RS ET during drought would likely result in overestimation of soil moisture depletion and therefore overestimation of the time required for drought recovery. Similarly, underestimation of RS ET over wet periods would lead to larger potential for floods. Challenges remain to more effectively incorporate remotely sensed variables into reliable ET estimation and land surface modeling.

GRACE data processing is continually being improved. The CSR RL05 solution used in this study has been shown to result in ~40% lower uncertainty relative to the previous version, CSR RL04 [*Long et al.*, 2013]. Results in this study show that uncertainty in GRACE-inferred ET is higher than LSM or RS ET. Spatial resolution of GRACE TWSA is projected to increase to ~50,000 km^2 and temporal resolutions to weekly or biweekly through a more advanced satellite gravimetry system and novel orbital configurations for the follow-on GRACE satellites (GRACE-FO) to be launched in 2017 [*Famiglietti and Rodell*, 2013]. Advancements in GRACE satellites should maximize its value in ET modeling and assessing water budget closure for water resources management. Comprehensive comparisons of the various ET products and quantification of their uncertainties should be valuable for selection of appropriate ET products for water management and for guiding future research in ET estimation.

Acknowledgments

This study was financially supported by the NASA Project NNX09AN10G. Additional support was provided by the Jackson School of Geosciences at The University of Texas at Austin and the BP project (No. CW1964449). We thank the Environmental Modeling Center, National Centers for Environment and Prediction (NCEP), NOAA, for providing NLDAS-2 data sets. We are grateful to three reviewers and associate editor for their insightful and constructive comments. The manuscript is improved as a result. Publication authorized by the Director, Bureau of Economic Geology, Jackson School of Geosciences, The University of Texas at Austin.

References

Anderson, M. C., J. M. Norman, J. R. Mecikalski, J. A. Otkin, and W. P. Kustas (2007a), A climatological study of evapotranspiration and moisture stress across the continental United States based on thermal remote sensing. 1: Model formulation, *J. Geophys. Res.*, *112*, D10117, doi:10.1029/2006JD007506.

Anderson, M. C., J. M. Norman, J. R. Mecikalski, J. A. Otkin, and W. P. Kustas (2007b), A climatological study of evapotranspiration and moisture stress across the continental United States based on thermal remote sensing. 2: Surface moisture climatology, *J. Geophys. Res.*, *112*, D11112, doi:10.1029/2006JD007507.

Bastiaanssen, W. G. M., M. Menenti, R. A. Feddes, and A. A. M. Holtslag (1998), A remote sensing surface energy balance algorithm for land (SEBAL). 1: Formulation, *J. Hydrol.*, *213*(1–4), 198–212.

Bastiaanssen, W. G. M., E. J. M. Noordman, H. Pelgrum, G. Davids, B. P. Thoreson, and R. G. Allen (2005), SEBAL model with remotely sensed data to improve water-resources management under actual field conditions, *J. Irrig. Drain. Eng.*, *131*(1), 85–93.

Bettadpur, S. (2007), *Level-2 Gravity Field Product User Handbook*, GRACE Project, pp. 327–734, Cent. for Space Res., Univ. of Texas, Austin.

Bruinsma, S., J.-M. Lemoine, R. Biancale, and N. Valès (2010), CNES/GRGS 10-day gravity field models (release 2) and their evaluation, *Adv. Space Res.*, *45*(4), 587–601.

Burnash, R. J. C., R. L. Ferral, and R. A. McGuire (1973), A generalized streamflow simulation system: Conceptual models for digital computer, report, 204 pp., Joint Fed. State River Forecast Cent., Sacramento, Calif.

Chen, F., K. Mitchell, J. Schaake, Y. K. Xue, H. L. Pan, V. Koren, Q. Y. Duan, M. Ek, and A. Betts (1996), Modeling of land surface evaporation by four schemes and comparison with FIFE observations, *J. Geophys. Res.*, *101*(D3), 7251–7268, doi:10.1029/95JD02165.

Dai, Y. J., et al. (2003), The common land model, *Bull. Am. Meteorol. Soc.*, *84*(8), 1013–1023.

Daly, C., M. Halbleib, J. I. Smith, W. P. Gibson, M. K. Doggett, G. H. Taylor, J. Curtis, and P. P. Pasteris (2008), Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States, *Int. J. Climatol.*, *28*(15), 2031–2064.

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.
Famiglietti, J. S., and M. Rodell (2013), Water in the balance, *Science*, *340*(6138), 1300–1301.
Fisher, J. B., K. P. Tu, and D. D. Baldocchi (2008), Global estimates of the land-atmosphere water flux based on monthly AVHRR and ISLSCP-II data, validated at 16 FLUXNET sites, *Remote Sens. Environ.*, *112*(3), 901–919.
Gao, H. L., Q. H. Tang, C. R. Ferguson, E. F. Wood, and D. P. Lettenmaier (2010), Estimating the water budget of major US river basins via remote sensing, *Int. J. Remote Sens.*, *31*(14), 3955–3978.
Gao, Y. C., and D. Long (2008), Intercomparison of remote sensing-based models for estimation of evapotranspiration and accuracy assessment based on SWAT, *Hydrol. Processes*, *22*(25), 4850–4869.
Grubbs, F. E. (1948), On estimating precision of measuring instruments and product variability, *J. Am. Stat. Assoc.*, *43*(242), 243–264.
Higgins, R. W., W. Shi, E. Yarosh, and R. Joyce (2000), Improved United States precipitation quality control system and analysis, *NCEP/Climate Prediction Center ATLAS No. 7*, 40 pp., Camp Springs, Md.
Huffman, G. J., R. F. Adler, D. T. Bolvin, G. J. Gu, E. J. Nelkin, K. P. Bowman, Y. Hong, E. F. Stocker, and D. B. Wolff (2007), The TRMM multisatellite precipitation analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales, *J. Hydrometeorol.*, *8*(1), 38–55.
Jeton, A. E., S. A. Watkins, T. J. Lopes, and J. Huntington (2005), Evaluation of precipitation estimates from PRISM for the 1961–90 and 1971–2000 data sets, Nevada Report, U.S. Geol. Surv., *Sci. Invest. Rep. 2005-5291*, Carson City, Nev.
Jiang, L., and S. Islam (2001), Estimation of surface evaporation map over southern Great Plains using remote sensing data, *Water Resour. Res.*, *37*(2), 329–340.
Koster, R. D., and M. J. Suarez (1994), The components of a SVAT scheme and their effects on a GCMs hydrological cycle, *Adv. Water Resour.*, *17*(1–2), 61–78.
Koster, R. D., and M. J. Suarez (1996), Energy and water balance calculations in the Mosaic LSM, *NASA Tech. Memo. NASA TM-104606*, vol. *9*, p. 60, Goddard Space Flight Cent., Greenbelt.
Liang, X., D. P. Lettenmaier, E. F. Wood, and S. J. Burges (1994), A simple hydrologically based model of land-surface water and energy fluxes for general-circulation models, *J. Geophys. Res.*, *99*(D7), 14,415–14,428.
Long, D., and V. P. Singh (2010), Integration of the GG model with SEBAL to produce time series of evapotranspiration of high spatial resolution at watershed scales, *J. Geophys. Res.*, *115*, D21128, doi:10.1029/2010JD014092.
Long, D., and V. P. Singh (2012a), A Two-source Trapezoid Model for Evapotranspiration (TTME) from satellite imagery, *Remote Sens. Environ.*, *121*, 370–388.
Long, D., and V. P. Singh (2012b), A modified surface energy balance algorithm for land (M-SEBAL) based on a trapezoidal framework, *Water Resour. Res.*, *48*, W02528, doi:10.1029/2010JD014092.
Long, D., and V. P. Singh (2013), Assessing the impact of end-member selection on the accuracy of satellite-based spatial variability models for actual evapotranspiration estimation, *Water Resour. Res.*, *49*, 2601–2618, doi:10.1002/wrcr.20208.
Long, D., B. R. Scanlon, L. Longuevergne, A.-Y. Sun, D. N. Fernando, and H. Save (2013), GRACE satellite monitoring of large depletion in water storage in response to the 2011 drought in Texas, *Geophys. Res. Lett.*, *40*, 3395–3401, doi:10.1002/grl.50655.
Longuevergne, L., B. R. Scanlon, and C. R. Wilson (2010), GRACE Hydrological estimates for small basins: Evaluating processing approaches on the High Plains Aquifer, USA, *Water Resour. Res.*, *46*, W11517, doi:10.1029/2009WR008564.
Luo, J. W., J. Bai, P. He, and K. Ying (2004), Axial strain calculation using a low-pass digital differentiator in ultrasound elastography, *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, *51*(9), 1119–1127.
Marx, A., H. Kunstmann, D. Schuttemeyer, and A. F. Moene (2008), Uncertainty analysis for satellite derived sensible heat fluxes and scintillometer measurements over Savannah environment and comparison to mesoscale meteorological simulation results, *Agric. For. Meteorol.*, *148*(4), 656–667.
McCabe, M. F., and E. F. Wood (2006), Scale influences on the remote estimation of evapotranspiration using multiple satellite sensors, *Remote Sens. Environ.*, *105*(4), 271–285.
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.
Mu, Q. Z., F. A. Heinsch, M. Zhao, and S. W. Running (2007), Development of a global evapotranspiration algorithm based on MODIS and global meteorology data, *Remote Sens. Environ.*, *111*(4), 519–536.
Mu, Q. Z., M. S. Zhao, and S. W. Running (2011), Improvements to a MODIS global terrestrial evapotranspiration algorithm, *Remote Sens. Environ.*, *115*(8), 1781–1800.
Nishida, K., R. R. Nemani, S. W. Running, and J. M. Glassy (2003), An operational remote sensing algorithm of land surface evaporation, *J. Geophys. Res.*, *108*(D9), 4270, doi:10.1029/2002JD002062.
Pan, M., A. K. Sahoo, T. J. Troy, R. K. Vinukollu, J. Sheffield, and E. F. Wood (2011), Multisource estimation of long-term terrestrial water budget for major global river basins, *J. Clim.*, *25*(9), 3191–3206.
Premoli, A., and P. Tavella (1993), A revisited three-cornered hat method for estimating frequency standard instability, *IEEE Trans Instrum. Meas.*, *42*(1), 7–13.
Ramillien, G., F. Frappart, A. Guntner, T. Ngo-Duc, A. Cazenave, and K. Laval (2006), Time variations of the regional evapotranspiration rate from Gravity Recovery and Climate Experiment (GRACE) satellite gravimetry, *Water Resour. Res.*, *42*, W10403, doi:10.1029/2005WR004331.
Rodell, M., I. Velicogna, and J. S. Famiglietti (2009), Satellite-based estimates of groundwater depletion in India, *Nature*, *460*(7258), 999–1002.
Rodell, M., J. S. Famiglietti, J. L. Chen, S. I. Seneviratne, P. Viterbo, S. Holl, and C. R. Wilson (2004a), Basin scale estimates of evapotranspiration using GRACE and other observations, *Geophys. Res. Lett.*, *31*, L20504, doi:10.1029/2004GL020873.
Rodell, M., et al. (2004b), The Global Land Data Assimilation System, *Bull. Am. Meteorol. Soc.*, *85*(3), 381–394.
Sahoo, A. K., M. Pan, T. J. Troy, R. K. Vinukollu, J. Sheffield, and E. F. Wood (2011), Reconciling the global terrestrial water budget using satellite remote sensing, *Remote Sens. Environ.*, *115*(8), 1850–1865.
Sheffield, J., C. R. Ferguson, T. J. Troy, E. F. Wood, and M. F. McCabe (2009), Closing the terrestrial water budget from satellite remote sensing, *Geophys. Res. Lett.*, *36*, L07403, doi:10.1029/2009GL037338.
Su, Z. (2002), The Surface Energy Balance System (SEBS) for estimation of turbulent heat fluxes, *Hydrol. Earth Syst. Sci.*, *6*(1), 85–99.
Tapley, B. D., S. Bettadpur, J. C. Ries, P. F. Thompson, and M. M. Watkins (2004), GRACE measurements of mass variability in the Earth system, *Science*, *305*(5683), 503–505.

Tavella, P., and A. Premoli (1991), Characterization of frequency standard instability by estimation of their covariance matrix, paper presented at the 23rd Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting, U.S. Naval Observatory, Pasadena, Calif., 3–5 Dec.

Tavella, P., and A. Premoli (1994), Estimating the instabilities of N-Clocks by measuring differences of their readings, *Metrologia*, *30*(5), 479–486.

Twine, T. E., W. P. Kustas, J. M. Norman, D. R. Cook, P. R. Houser, T. P. Meyers, J. H. Prueger, P. J. Starks, and M. L. Wesely (2000), Correcting eddy-covariance flux underestimates over a grassland, *Agric. For. Meteorol.*, *103*(3), 279–300.

Vinukollu, R. K., E. F. Wood, C. R. Ferguson, and J. B. Fisher (2011), Global estimates of evapotranspiration for climate studies using multi-sensor remote sensing data: Evaluation of three process-based approaches, *Remote Sens. Environ.*, *115*(3), 801–823.

Wang, K. C., and R. E. Dickinson (2012), A review of global terrestrial evapotranspiration: Observation, modeling, climatology, and climatic variability, *Rev. Geophys.*, *50*, RG2005, doi:10.1029/2011RG000373.

Wei, H. L., Y. L. Xia, K. E. Mitchell, and M. B. Ek (2013), Improvement of the Noah land surface model for warm season processes: Evaluation of water and energy flux simulation, *Hydrol. Processes*, *27*(2), 297–303.

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.

Wurbs, R. A. (2006), Methods for developing naturalized monthly flows at gaged and ungaged sites, *J. Hydrol. Eng.*, *11*(1), 55–64.

Xia, Y. L., 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. L., 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, Y. T., and S. H. Shang (2013), A hybrid dual source scheme and trapezoid framework based evapotranspiration model (HTEM) using satellite images: Algorithm and model test, *J. Geophys. Res.*, *118*, 2284–2300, doi:10.1002/jgrd.50259.

Yang, Y. T., D. Long, and S. H. Shang (2013), Remote estimation of terrestrial evapotranspiration without using meteorological data, *Geophys. Res. Lett.*, *40*, 3026–3030, doi:10.1002/grl.50450.

Yeh, P. J. F., S. C. Swenson, J. S. Famiglietti, and M. Rodell (2006), Remote sensing of groundwater storage changes in Illinois using the Gravity Recovery and Climate Experiment (GRACE), *Water Resour. Res.*, *42*, W12203, doi:10.1029/2006WR005374.

Zhang, K., J. S. Kimball, R. R. Nemani, and S. W. Running (2010), A continuous satellite-derived global record of land surface evapotranspiration from 1983 to 2006, *Water Resour. Res.*, *46*, W09522, doi:10.1029/2009WR008800.