Water Resources Research

RESEARCH ARTICLE
10.1029/2019WR026594

Key Points:
- GRACE-based estimates of ET are used to evaluate modeled, remote sensing, and empirical ET products at the basin scale in the contiguous United States
- Long-term mean modeled and observation-based ET is persistently smaller in magnitude compared to GRACE-based ET
- Interannual variability of GRACE-based ET is greater than in modeled and remote sensing-based products

Correspondence to:
M. A. Pascolini-Campbell,
madeleine.a.pascolini-campbell@jpl.nasa.gov

Citation:
Pascolini-Campbell, M. A., Reager, J. T., & Fisher, J. B. (2020). GRACE-based mass conservation as a validation target for basin-scale evapotranspiration in the contiguous United States. *Water Resources Research*, *56*, e2019WR026594. https://doi.org/10.1029/2019WR026594

Received 21 OCT 2019
Accepted 3 FEB 2020
Accepted article online 10 FEB 2020

GRACE-based Mass Conservation as a Validation Target for Basin-Scale Evapotranspiration in the Contiguous United States

Madeleine A. Pascolini-Campbell[1], John T. Reager[2], and Joshua B. Fisher[2]

[1]NASA Postdoctoral Program Fellow, NASA Jet Propulsion Laboratory, Caltech Institute of Technology, Pasadena, CA, USA, [2]NASA Jet Propulsion Laboratory, Caltech Institute of Technology, Pasadena, CA, USA

Abstract Here, we evaluate basin-scale evapotranspiration (ET) estimates for eleven major river basins in the contiguous United States against a water balance approach with Gravity Recovery and Climate Experiment (GRACE) satellite observations. The relatively precise measurements of large-scale changes in water mass from GRACE are used to estimate the storage rate term in the terrestrial water budget and consequently provide an estimate, with propagated uncertainty, of basin-aggregated ET from mass conservation. We apply GRACE-based ET to two modeling systems (NLDAS-2 and GLDAS-2.1) comprised of five land surface models and three remote sensing-based products (MOD16, PT-JPL, and FLUXCOM) for 2003 to 2014. Both the land surface model-based and remote sensing-based ET are persistently lower than GRACE-based ET in all eleven basins tested. We also find that interannual variability is greater for GRACE-ET than the model and remote sensing products, and this is attributed to precipitation variability.

1. Introduction

Evapotranspiration (ET) is defined as the loss of water to the atmosphere through vaporization from the land surface and transpiration of plants and constitutes a fundamental flux in global and regional hydrological and energy cycles. Correct knowledge of ET is critical for understanding water availability and demand (McCabe & Wood, 2006), local hydrometeorology (Koster et al., 2003), and for closure of the surface energy budget (Trenberth et al., 2009). The hydrological cycle is expected to intensify in a warming climate (Huntington, 2006), but as yet, observed global trends in ET remain contested (Dong & Dai, 2017; Fisher et al., 2017). The river basin is also often the scale of interest in water resources management, and therefore, accurate knowledge of large-scale ET fluxes is needed.

ET is inherently difficult to measure, and as a result, large discrepancies exist across observational estimates because of the differences in methodological approach and assumptions, issues in scaling, and the associated and resulting uncertainties (Liu et al., 2016; Senay et al., 2011; Wang & Dickinson, 2012). Similarly, there is often much difference across numerical modeling approaches and the associated future projections of ET responses to climate change, due to physical parameterizations and model structural representations (Dai & Zhao, 2017; Milly & Dunne, 2010; Wang et al., 2015).

Because of the difficulty in capturing and quantifying evapotranspired water vapor, the most direct in situ approach involves a measurement of soil mass loss, such as that estimated by a weighing lysimeter (Wang & Dickinson, 2012). Proxy approaches can be applied that use changes in relative land surface temperature or atmospheric turbulence, but these approaches are affected by the errors resulting from the complex nonlinearities between atmospheric or land-surface proxies and actual ET. For example, eddy covariance methods have been used to estimate ET from in situ stations, but suffer from uncertainty due to gap infilling from missing data, and due to upscaling beyond the approximate and variable 1-km representative area (Baldocchi, 2003). Eddy-covariance techniques tend to be the standard for the validation of ET at local scales (Baldocchi et al., 2001). However, these networks are sparse in time and space and present challenges when used in larger basin-scale evaluations.

ET can also be indirectly estimated from a combination of radiative, atmospheric, and surface data using various algorithms (Penman-Montieth (Penman, 1948), Priestly-Taylor (Priestly & Taylor, 1972), and Thornthwaite (Thornthwaite, 1948)) (see Table 1). These algorithms require various inputs including temperature,

10.1029/2019WR026594

Table 1
Summary of Algorithms and Variables

Model/Variable	Algorithm	Units	Reference
Penman-Montieth	$\frac{\Delta(R_n-G)+\rho_a c_p \frac{e_s-e_a}{r_a}}{\Delta+\gamma(1+\frac{r_s}{r_a})}$	mm day^{-1}	Penman (1948)
Thornthwaite	$16(\frac{10T}{I})^a$	mm day^{-1}	Thornthwaite (1948)
Priestley and Taylor	$\alpha(\frac{\Delta(R_n-G)}{\Delta+\gamma})$	mm day^{-1}	Priestley and Taylor (1972)
Net radiation	R_n	MJ m^{-2} day^{-1}	
Soil heat flux	G	MJ m^{-2} day^{-1}	
Vapor pressure deficit	$(e_s - e_a)$	kPa	
Mean air density	r_a	kg m^{-3}	
Specific heat air	c_p	MJ kg^{-1} °C^{-1}	
Slope saturation vapor pressure	Δ	kPa °C^{-1}	
Pyschrometric constant	γ	kPa °C^{-1}	
Bulk surface and aerodynamic resistances	r_s, r_a	s m^{-1}	
Temperature	T	°C	
Heat index	I	°C	
Constants	α, a		

water vapor pressure, net radiation, and information on the land cover and vegetation type to calculate ET (Fisher et al., 2011). Remote sensing-based products (from remote sensing and FLUXCOM) and land surface models (LSMs) use (respectively) in situ measurements, satellite observations, and variables in algorithms to produce global gridded ET variables (e.g., Fisher et al., 2008; Mu et al., 2007; Zhang et al., 2010). An advantage of these ET products is the large spatial and temporal scales they produce. However, their accuracy is limited by the quality of the input observations and is also sensitive to the choice of algorithm used (Badgley et al., 2015; Fisher et al., 2017; Gao et al., 2010; Polhamus et al., 2013). Many of the parameters required by the ET algorithms, for example, stomatal resistance, aerodynamic resistance (in Penman-Montieth) may also be challenging to retrieve globally (Fisher et al., 2017). Furthermore, at the basin scale, bulk estimates of ET from gridded observation, remote sensing, and LSMs are difficult to validate due to the sparse network of in situ monitoring sites as mentioned above. To overcome this, previous studies have assessed bulk ET from models and observations using independent ET estimates and water balance closures with runoff and precipitation data only (Long et al., 2014; Miralles et al., 2016); however, these approaches are also limited by relying on single data sets especially in the choice of precipitation (Gibson et al., 2019).

Since 2002, the Gravity Recovery And Climate Experiment (GRACE) (Tapley et al., 2004) has provided accurate global measurements of terrestrial water storage anomalies (TWSA) improving our knowledge of the dynamic terrestrial hydrological cycle by revealing formerly hidden information of changes in aggregated water storage, including water in groundwater, soils, surface water, vegetation water, and snow (e.g., Rodell et al., 2018). One constraint of the GRACE observations is that the spatial resolution of the measurements is defined by the spacecraft orbit, and GRACE measurements are generally characterized by reduced measurement error with increasing signal magnitude and increasing spatial resolution. Data products are produced with a nominal monthly temporal resolution and come with a formal error product.

Using a water balance approach, GRACE TWSA can be combined with precipitation and runoff data to calculate basin-scale ET (Güntner, 2008; Ramillien et al., 2006; Rodell et al., 2011, 2004). GRACE also provides a novel opportunity to estimate basin-scale ET as TWSA captures all water fluxes, both from natural climate (precipitation, soil moisture, ice and snow mass, and runoff), as well as from interference in the hydrological cycle by human activity (irrigation, reservoir creation, and groundwater extraction) (Castle et al., 2016; Rodell et al., 2009, 2011).

Basin-scale GRACE-ET (ET_{GRACE}) estimates are constrained by the accuracy and availability of runoff and precipitation data. In the contiguous United States, runoff data is readily provided by the United States Geological Survey (USGS) and is one of the more accurately measured components of the hydrological cycle

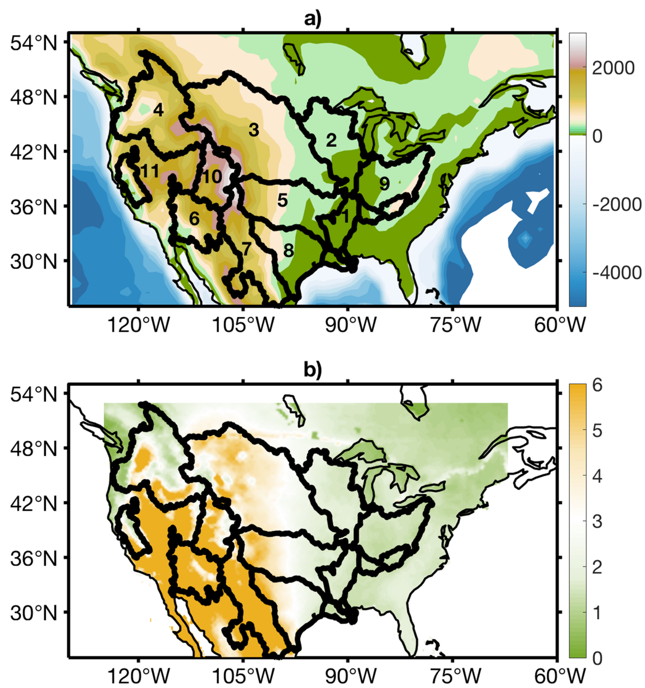

Figure 1. (a) Map showing river basins used in this study: (1) Mississippi, (2) Upper Mississippi, (3) Missouri, (4) Columbia, (5) Arkansas-White-Red, (6) Colorado, (7) Bravo, (8) Texas Gulf Coast, (9) Ohio, (10) Upper Colorado, and (11) Sacramento-San Joaquin. Shading indicates elevation (meters). (b) Map showing the eleven basins and aridity index (PET/P) calculated using NARR PET and CPC-based precipitation from NLDAS-2 forcing.

(Fekete & Vörösmarty, 2007). Precipitation error is introduced by different measurement techniques (satellite, rain gage, and reanalysis) (Gehne et al., 2016) and, for large basins in the United States, is typically the greatest uncertainty in the water balance (Gao et al., 2010). Because GRACE does well at larger spatial scales, for large river basins, uncertainty introduced by TWSA in the terrestrial water budget is relatively low (Ramillien et al., 2006). In light of the uncertainty associated with precipitation, previous studies have used ensembles of precipitation to calculate water balance ET for river basins in the United States and globally (Liu et al., 2016; Swann & Koven, 2017).

In this study, we calculate ET_{GRACE} for eleven major U.S. basins for 2003 to 2016 and evaluate five LSM-based and three remote sensing-based products against ET_{GRACE}. We use USGS runoff and seven different precipitation data sets to calculate ET_{GRACE}, building upon previous studies (Liu et al., 2016; Swann & Koven, 2017). Although this method is limited by the accuracy of the input data sets (Swann & Koven, 2017), it produces estimates of ET which are constrained by mass conservation and can therefore be used to evaluate basin-scale ET fluxes which is not currently possible through the sparse Fluxnet network. Given that errors may accumulate in LSM and remote sensing products over larger scales, GRACE-based ET therefore has the potential to provide a check on basin-scale ET products. Accurate knowledge of basin-scale ET is of use for water resource management, in which the basin is the policy-relevant scale, and may also inform the development of LSM and remote sensing ET products.

2. Methods

2.1. River Basins

We focus our analysis on eleven river basins in the United States as follows: Mississippi, Upper Mississippi, Missouri, Ohio, Arkansas-White-Red, Columbia, Colorado, Upper Colorado, Bravo, Texas Gulf Coast, and Sacramento-San Joaquin (Figure 1). We choose to calculate ET for both the entire Mississippi and its encompassing basins in order to examine the sensitivity of ET accuracy to scale. The large scale of the basins used in this study is intended to increase the reliability of our ET_{GRACE} estimate which is limited by the coarse resolution of GRACE TWS mascons (Ramillien et al., 2006).

10.1029/2019WR026594

Table 2
Summary of USGS Discharge Data

River Basin	Gage Location	Gage Number	Temporal Availability
Mississippi	Vicksburg, Mississippi	07289000	2008–2018
Upper Mississippi	Thebes, Illinois	07022000	1989–2019
Missouri	Hermann, Missouri	06934500	1987–2017
Columbia	Dalles, Oregon	14105700	1989–2019
	Birchbank, British Columbia	12323000	1982–2016
Arkansas-White-Red	Spring Bank, Arkansas	07344370	1997–2019
	Ft. Smith, Arkansas	07249455	1997–2019
Colorado	Above Imperial Dam, Arizona-California	09429490	1976–2019
Bravo	Rio Grande near Brownsville, Texas	08475000	2003–2011
Texas Gulf Coast	Brazos River near Rosharon, Texas	08116650	1967–2018
	Red River at Spring Bank, Arkansas	07344370	1997–2018
	Sabine River near Ruliff, Texas	08030500	1924–2019
	Trinity Canal near Dayton, Texas	08067070	1981–2019
Ohio	Old Shawneetown, Illinois-Kentucky	03381700	2002–2019
Upper Colorado	Lee's Ferry, Arizona	09380000	1921–2019
Sacramento-San Joaquin	Verona, California	11425500	1929–2019
	San Joaquin near Vernalis, California	11303500	1923–2018

2.2. Mass Conservation ET estimate (ET_{GRACE})

For each of the eleven basins, we use a water balance approach to calculate ET for 2003 to 2016 (or for the years within this period available in USGS runoff data). We calculate ET as

$$ET_{GRACE} = P - Q - \frac{dS}{dt}, \tag{1}$$

where P is precipitation, Q is runoff at the basin outlet, and $\frac{dS}{dt}$ is the change in total water storage (at monthly resolution), which is calculated from GRACE TWS (Ramillien et al., 2006).

For the GRACE TWS and uncertainty grids, we use the most recent GRACE mascon solution RL06 (available online at: https://podaac.jpl.nasa.gov/dataset/). We choose to use only the RL06 mascon solution in this study, as it represents an improvement from previous spherical harmonic solutions by reducing leakage error (Watkins et al., 2015; Wiese et al., 2016). We use linear interpolation for missing months.

Although the GRACE TWSA data are provided on a monthly time step, they are not collected uniformly, and therefore, signals from neighboring months introduce errors that accumulate in the differencing process and introduce high frequency artifacts (Landerer et al., 2010). To address this issue, we use a centered finite difference approach to obtain $\frac{dS}{dt}$, which is effectively a smoothing operation (Landerer et al., 2010). To calculate $\frac{dS}{dt}$, we find the change in water storage for 1 month by differencing the preceding and following months and dividing by 2 ΔT (where ΔT is 1 month). This method has been used in a number of studies (Landerer et al., 2010; Long et al., 2014; Swann & Koven, 2017).

To examine the sensitivity of using the centered finite difference approach, we compare the time series of ET for each basin using this method as well as a backward difference method (not shown). We do not find any significant differences in the mean ET for any basins when using the different methods to calculate ET (using a Student's t test and $\alpha = 0.05$). We also compare the seasonal cycles generated from either method and do not find any differences in the timing of peak ET.

Q is from USGS gaged-monthly discharge at the outlet of each of the basins (Table 2) (available online at https://waterdata.usgs.gov/). We select the stream gage closest to the outlet of the river basin for which data is available over the time period of interest (2003–2016). When more than one outlet station is used for Q (for example, the Texas Gulf Coast), the different time series of Q are added together, and this new series of total Q is used to calculate ET in (1).

10.1029/2019WR026594

Table 3
Models and Data Sets

Data	Temporal Availability	Temporal Resolution	Variable
GLDAS Version 2.1 Noah	2000–present	Monthly	ET (LSM-based)
GLDAS Version 2.1 CLSM	2000–2014	Daily	ET (LSM-based)
NLDAS Version 2 Noah	1979–present	Monthly	ET (LSM-based)
NLDAS Version 2 Mosaic	1979–present	Monthly	ET (LSM-based)
NLDAS Version 2 VIC	1979–present	Monthly	ET (LSM-based)
MOD16	2000–2014	Monthly	ET (r.s. derived)
AVHRR PT-JPL	2002–2017	Monthly	ET (r.s. derived)
FLUXCOM	2001–2014	Monthly	Latent Heat (r.s. and obs. derived)
PRISM	1895–present	Monthly	Precip (Gage)
GPCC	1891–2016	Monthly	Precip (Gage)
CRU TSv4.02	1901–present	Monthly	Precip (Gage)
CPC	1979–2018	Monthly	Precip (Gage)
GHCN CLIMGRID	1895–present	Monthly	Precip (Gage)
NLDAS-2 Forcing	1979–present	Monthly	Precip (Gage)
MERRA-2	1980–2019	Monthly	Precip (Reanalysis)
NARR	1979–2018	PET (Reanalysis)	

For precipitation, we use seven different data sets in order to constrain calculated ET in light of the errors associated with precipitation (as in Swann & Koven, 2017). These include Parameter-elevation Relationships on Independent Slopes Model (PRISM) data set available at 1/8° spatial resolution (Daly et al., 2008), Global Precipitation Climatology Centre (GPCC) station based gridded data at 0.5° resolution (Schneider et al., 2016), Climatic Research Unit (CRU) gridded precipitation data version TS v4.02 based on observations and gridded at 0.5° resolution (Harris et al., 2014), Climate Prediction Center (CPC) Unified Gauge-Based Analysis of Daily Precipitation over the continental US available at 0.25° resolution (Chen et al., 2008; Xie et al., 2007), Global Historical Climatology Network (GHCN) gridded daily precipitation available at $1/8^o$ resolution (CLIMGRID) (Vose et al., 2014), 1/8° resolution precipitation from the North American Land Data Assimilation System (NLDAS-2) forcing which uses CPC-based precipitation adjusted with PRISM topography (Daly et al., 1994; Higgins et al., 2000), and finally, MERRA-2 Reanalysis precipitation at 0.5^o by 0.625° (Gelaro et al., 2017) (Table 3). All the precipitation data sets used are at a monthly temporal resolution.

The different components of equation (1) (precipitation [P], outlet discharge [Q], and $\frac{dS}{dt}$ for each basin) are plotted in Figures 2 and 3 over the period of availability. The magnitude of P and $\frac{dS}{dt}$ tend to dominate Q in the calculation of ET. Several basins (Mississippi and Bravo) do not have USGS records of Q extending through to 2016 at the outlet gages.

2.3. LSM ET Products

Model-derived ET is taken from the Global Land Data Assimilation System (GLDAS) Version 2.1 for the Noah and Catchment Land Surface Model (CLSM) models (available online at https://disc.gsfc.nasa.gov/) (Rodell et al., 2004) (see Table 3). These are available from 2000 to present at a monthly and daily temporal resolution, and 0.25° globally. We also use output from NLDAS-2 Noah, Mosaic, and VIC models which have a resolution of 1/8° over North America from 1979 to present (Xia et al., 2012). In the study, we create multi-model averages of ET for the GLDAS-2.1 (Noah and CLSM) and NLDAS-2 (Noah, Mosaic, and VIC) outputs. Each of the LSMs uses a Penman-Montieth-based formulation for PET with different parameterizations to obtain actual ET (Kumar et al., 2018). The parameterizations are based on scaling PET using variables related to vegetation, land surface, and water availability.

2.4. Remote Sensing-Based ET Products

We use three different ET products derived from observations and remote sensing. First, we use MODIS Global Evapotranspiration Project (MOD16) which calculates ET using an algorithm based on Penman-Montieth (Mu et al., 2011). MOD16 is available from 2000 to 2014 at a spatial resolution of

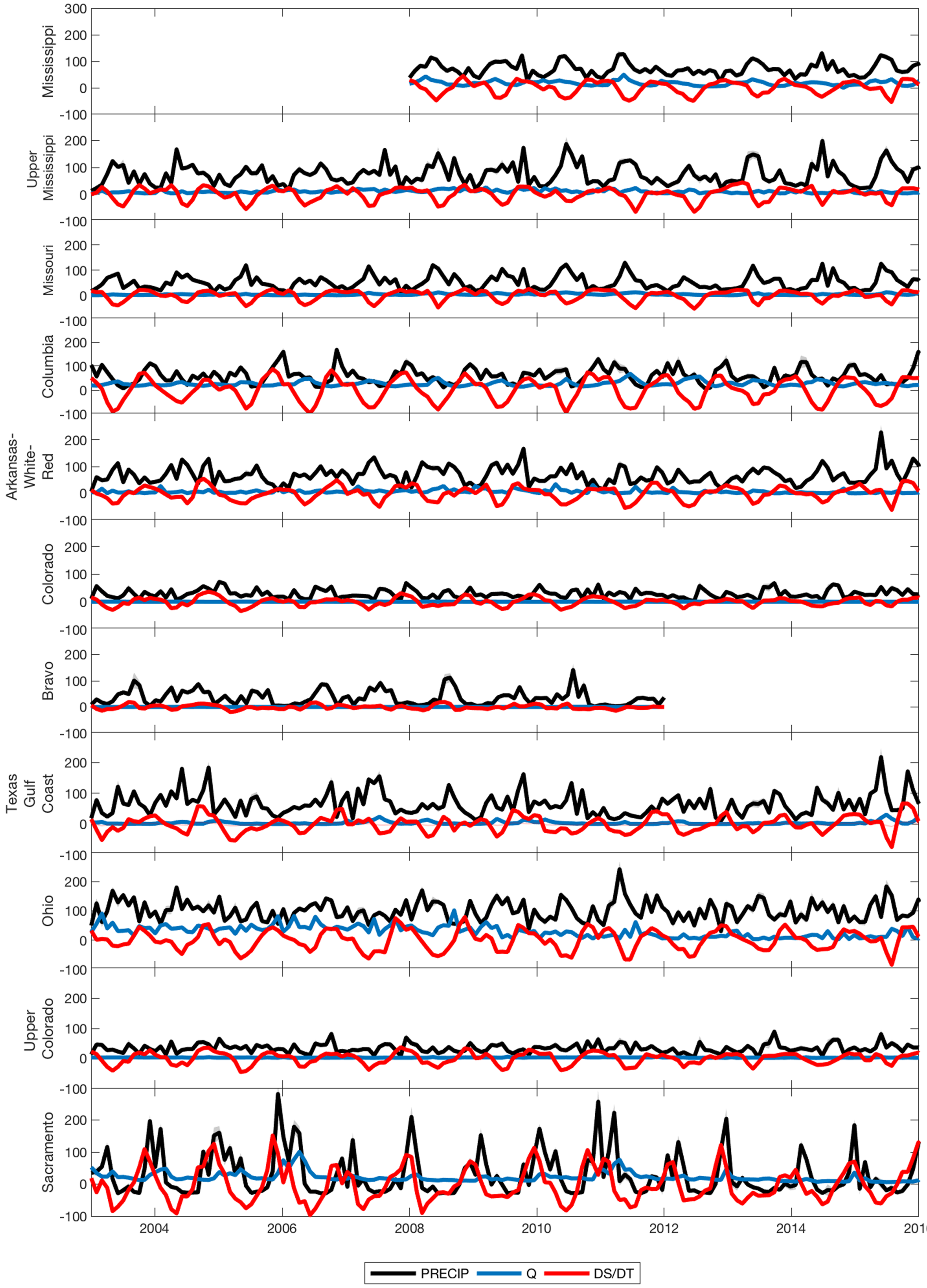

Figure 2. For each basin: seven precipitation products mean (black line), discharge at outlet (blue line), and GRACE $\frac{dS}{dt}$ (red line) for 2003-2016. y-axis units are in mm month^{-1}.

10.1029/2019WR026594

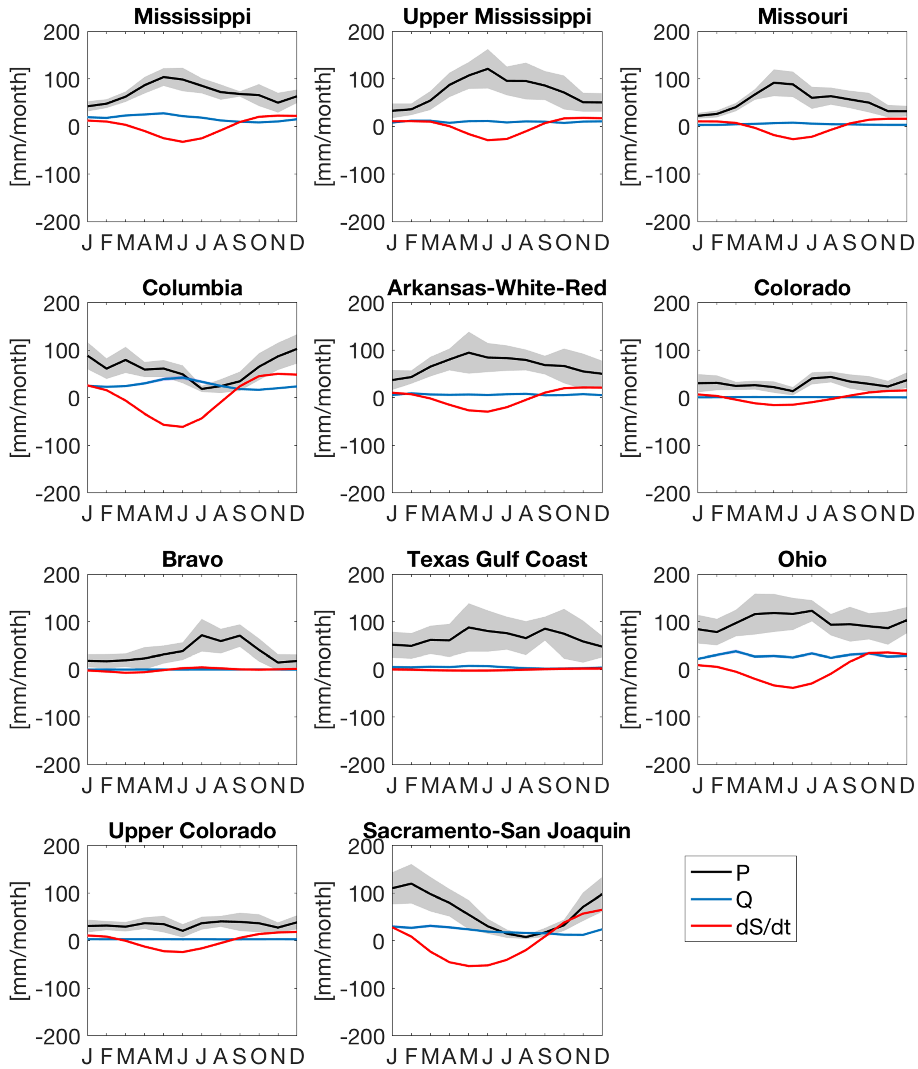

Figure 3. Seasonal cycles of precipitation (plotted as the mean of the seven data sets) (black line), runoff (blue line), and $\frac{dS}{dt}$ (red line) for 2003–2016. The red shading indicates the standard deviation in monthly precipitation over the period of 2003 to 2016.

0.05° (available online at http://files.ntsg.umt.edu/data/NTSG_Products/MOD16/). The algorithm uses land cover, leaf area index (LAI), fraction of photosynthetically active radiation (FPAR), and albedo from MODIS, as well as meteorological reanalysis data from MERRA GMAO (including air pressure, temperature, humidity, and radiation) to calculate ET.

Second, we use ET calculated using the updated Priestly-Taylor Jet Propulsion Laboratory algorithm (PT-JPL); we use PT-JPL data at 36-km resolution (Fisher et al., 2008). The PT-JPL algorithm uses net radiation, normalized difference vegetation index (NDVI), soil adjusted vegetation index, maximum air temperature, and water vapor pressure to reduce potential ET (PET) (see Table 1; Priestley & Taylor, 1972) to actual ET. The algorithm also partitions total ET into three sources: canopy transpiration, soil evaporation, and interception evaporation (Fisher et al., 2011). Input variables to the calculation are taken from remote sensing and observations (Fisher et al., 2008).

We also use ET from FLUXCOM which integrates satellite observations, meteorological measurements, Fluxnet sites, and different machine learning algorithms to estimate carbon and energy fluxes (Jung et al., 2019). FLUXCOM is available from 2001 to 2014 at a resolution of 0.5° (available online at http://www.fluxcom.org). The FLUXCOM data involve two different products: one is based entirely on remote sensing data (RS), and one involves both remote sensing data and meteorological observations (RS + METEO) (Jung

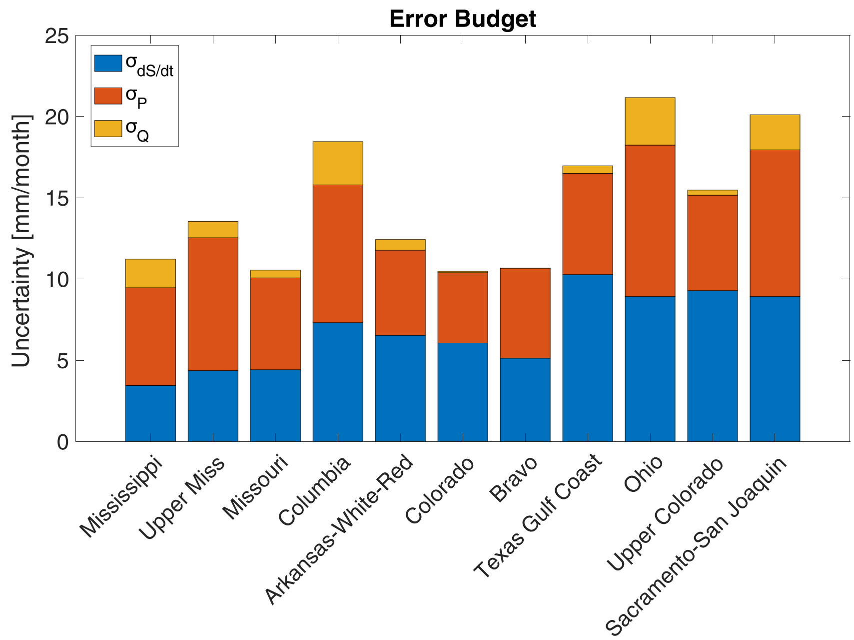

Figure 4. The error for ET_{GRACE} is decomposed as $\sigma^2_{ET} = \sigma^2_{dS/dt} + \sigma^2_P + \sigma^2_Q$, and $\sigma_{dS/dt}$, σ_P, and σ_Q are plotted for each of the eleven basins.

et al., 2019). In this study, we average the different data sets to obtain a single FLUXCOM series for each of the basins.

Given the different temporal availabilities of the products, we focus on the period 2003 to 2014 in order to obtain the maximum number of complete overlapping years of data.

2.5. Uncertainty Treatment for ET Estimates

For our calculated ET_{GRACE}, the uncertainty comprises three different components: (1) the time-varying GRACE $\frac{dS}{dt}$ monthly uncertainty which involves propagating the error of TWSA by adding the error variance of the two months of TWSA used in the calculation (as in Castle et al., 2016), (2) the uncertainty in discharge, (3) uncertainty in precipitation (Rodell et al., 2004).

For the TWSA error, we use the formal uncertainty grids provided with the GRACE mascon solutions. Although the grids are provided at 0.5° × 0.5° resolution, the real resolution of the mascon is much coarser (3° × 3°). We therefore convert the grids into mascon averages, which reduces the grid to 4,551 unique estimates globally. For each basin, we calculate an area weighted root sum of squares, where if, for example, the river basin is only partially covering a mascon, the weighted value of the uncertainty for the covered area is used in the basin uncertainty calculation (personal communication, David Wiese).

For USGS runoff, errors have been found to range from 2% (ideal conditions) to 20% (very poor conditions) with most measurement errors between 3% and 6% (Sauer & Meyer, 1992). In this study, we use 10% for Q as higher estimate for error. Uncertainties ranging from 5% to 15% have been used in other studies for USGS hydrological measurements (Castle et al., 2014, 2016; Senay et al., 2011; Thomas et al., 2016).

For error in precipitation, we use the standard deviation of the time series of the seven different precipitation data sets. The error of TWSA is also spatially averaged according to the area of the drainage basin in question, reducing the overall error at larger scales associated with this quantity. Overall uncertainty for ET_{GRACE} (σ^2_{ET}) is then calculated as

$$\sigma^2_{ET} = \sigma^2_{ds/dt} + \sigma^2_P + \sigma^2_Q. \tag{2}$$

We plot the error budget for ET_{GRACE} in terms of its individual components $\sigma_{dS/dt}$, σ_P, and σ_Q as described in equation (2) above (Figure 4). We also calculate the individual errors as a percentage of the total error (Table 4). For seven of the eleven basins, σ_P is the largest source of error accounting for up to 60.3% of the total error in the Upper Mississippi basin. We also note that the error due to $\sigma_{dS/dt}$ is larger in the smaller basins (Texas Gulf Coast, Ohio, Upper Colorado, and Sacramento-San Joaquin). This is consistent with other

Table 4
Error Budget: Components of Error as Percentage of Total ET_{GRACE} error (%)

	dS/dt	P	Q
Mississippi	30.8	53.6	15.7
Upper Mississippi	32.2	60.3	7.5
Missouri	41.9	53.6	4.5
Columbia	39.7	45.9	14.4
Arkansas-White-Red	52.6	42.2	5.2
Colorado	57.9	41.3	0.8
Bravo	48.1	51.8	0.1
Texas Gulf Coast	60.5	36.7	2.7
Ohio	42.2	44.0	13.8
Upper Colorado	60.0	38.0	2.0
Sacramento-San Joaquin	44.4	44.9	10.8

studies that have shown that uncertainty increases for smaller river basins due to the coarse resolution of the GRACE mascons (Wiese et al., 2016).

For the LSM-based ET, no formal uncertainty products are presented. We therefore define an uncertainty as the standard deviation among the independent model estimates. That includes a spread across five models, three of those with independent forcing and run at a different resolution from the other two. These discrepancies in model structure, model forcing, and model resolution across this suite of models should be a fair representation of all potential model error sources in the estimation of ET. It is beyond the scope of this study to identify the individual error sources contributing to LSM-based ET estimates in individual models, but we take the spread across these estimates as the defined representation of several types of potential uncertainties.

For the remote sensing-based ET products, no formal uncertainty products are presented. Therefore, uncertainty is estimated here as the standard deviation across the products. The three products used constitute the state of the art in modern publicly available ET products, and differences between them represent differences in fundamental assumptions, errors in measurement techniques, and errors in algorithm and approach. We therefore take the spread across the popular ET products as a representative uncertainty of potential error sources in remote sensing-based ET.

3. Results

3.1. Seasonal Cycle Comparison

We calculate the long-term mean seasonal cycle ET_{GRACE} for each of the eleven river basins over 2003 to 2014 and compare it with the five model and three remote sensing-based products (Figure 5). The uncertainty associated with ET_{GRACE} is indicated by the shading, and in general, the uncertainty in our estimate is dominated by error propagation of TWSA into $\frac{dS}{dt}$. Wetter basins in the eastern United States have greater values of spring–summer ET, reaching approximately 100 mm month^{-1} for the Mississippi, 125 mm month^{-1} for the Upper Mississippi, and 75 mm month^{-1} for the Columbia. More arid basins have lower values, with approximately 60 mm month^{-1} for the Bravo, 75 mm month^{-1} for the Sacramento-San Joaquin, and 50 mm month^{-1} for the Upper Colorado.

In general, the timing and seasonality of ET agree for the different products compared with ET_{GRACE}, except for MOD16 ET, which stands out as below the other estimates. Notable differences include the two-peaked ET for the Colorado, Upper Colorado, Bravo, and Texas Gulf Coast rivers in ET_{GRACE} which does not appear in the modeling and remote sensing-based products. The basins containing two-peaked seasonal cycles of ET also are characterized by two-peaked seasonality in precipitation (which is the largest contribution in the calculation of ET and of particular importance in arid basins (Ukkola & Prentice, 2013) (Figure 3). Ohio ET_{GRACE} indicates a small ET peak in winter, which is also consistent with the winter peak in precipitation which occurs in this basin (Figure 3).

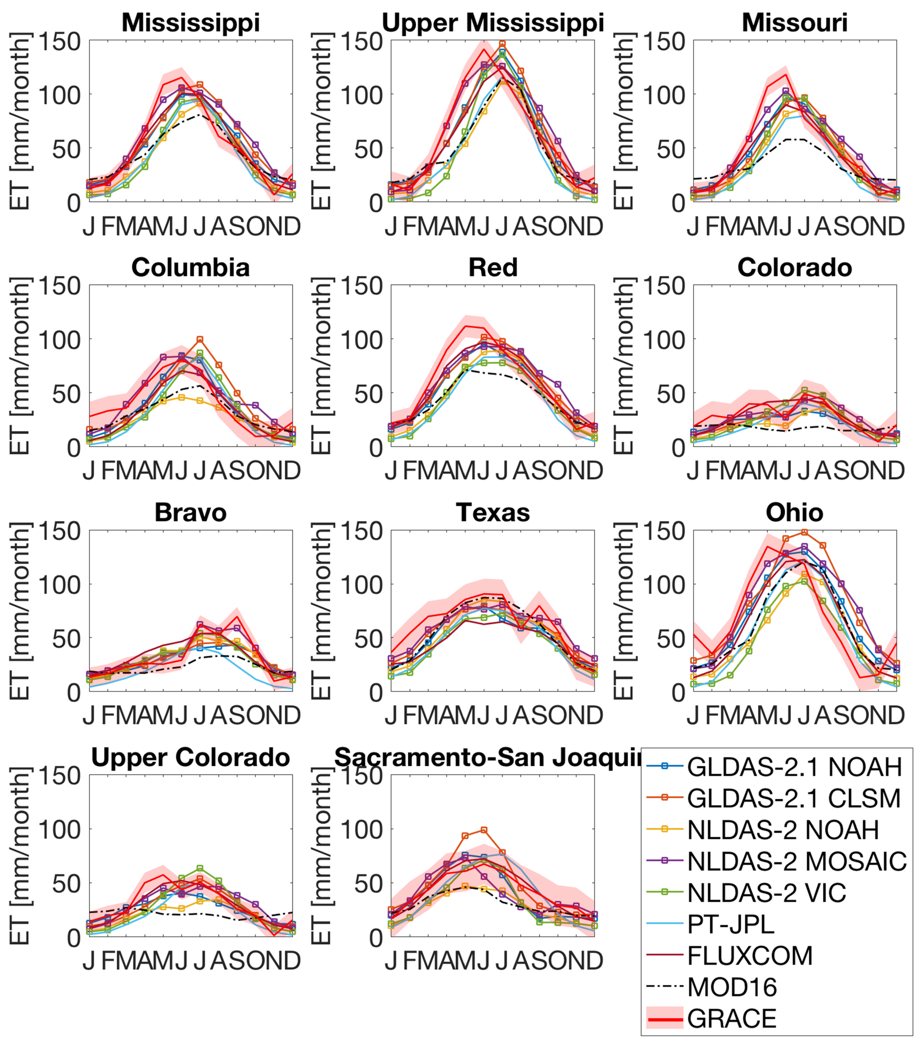

Figure 5. Seasonal cycle of ET_{GRACE} (red line and shading), LSM-based ET, and remote sensing-based ET annually averaged over 2003–2014.

3.2. Evaluation of ET products against ET_{GRACE}

We use observed climate variables to calculate the water budget closure to address how well LSM-based and remote sensing-based ET products can close the water budget on long time scales (over 2003 to 2014) (Figure 6). For each ET product, we calculate

Water budget closure:

$$Water\ budget\ closure = P - E - Q - \frac{dS}{dt}, \tag{3}$$

where P, E, Q, and $\frac{dS}{dt}$ are taken as the long-term average values corresponding to each basin. For P, we use an average of the seven precipitation products used to calculate ET_{GRACE}, Q is from USGS, and $\frac{dS}{dt}$ from GRACE. Provided there are no major changes in storage $\frac{dS}{dt}$ (for example, from groundwater extraction and glacier retreat), it is expected that in the long term, the water budget closure should be equal to zero.

The average water budget residual for ET_{GRACE}, which is calculated using the water balance (2), is by definition zero, implying that it accounts for all fluxes of water in the long term. We also plot the standard deviation in the water budget closure for each of the seven members of the ET_{GRACE} ensemble against mean P (indicated by error bars on Figure 6). This demonstrates variability in the water balance closure due to precipitation. For the rest of the analysis, we assess model and remote sensing-based ET products against mean ET_{GRACE}, which does close the water balance.

10.1029/2019WR026594

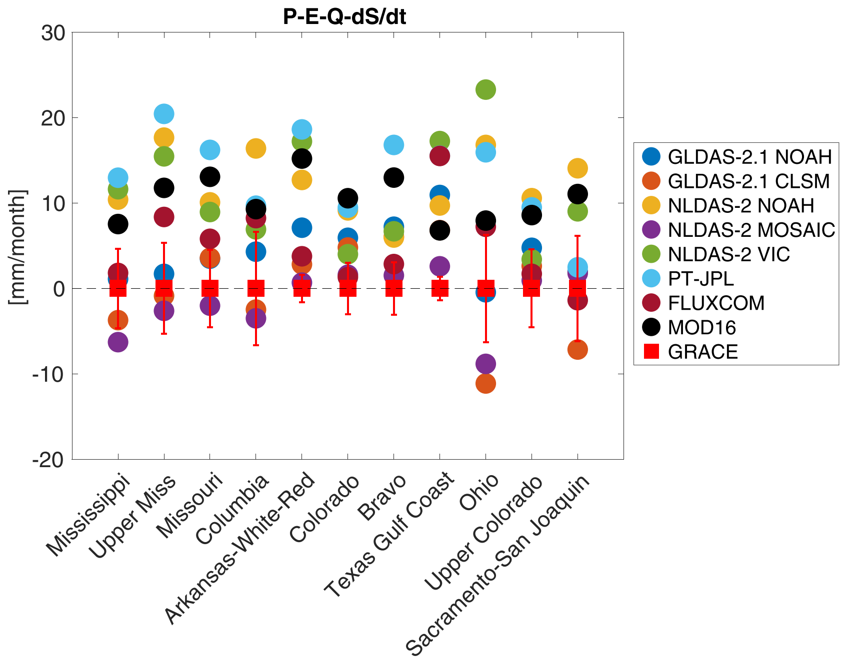

Figure 6. Water budget closure for the LSMs and remote sensing-based products using observed precipitation (mean of the seven data sets), runoff and $\frac{dS}{dt}$ over 2003–2014 for each of the eleven basins used in this study. For ET_{GRACE}, we plot the mean closure (indicated by red points) and the standard deviation of the water budget closure for the individual members of the ET_{GRACE} ensemble (red error bars).

Note that for the LSM-based ET products, this is not an evaluation of successful mass budget closure within the LSMs. This is an independent assessment using well-constrained discharge and terrestrial water storage observations of whether the LSM-based ET estimate leads to mass budget closure in the real world. Because many models do impose a water mass balance, and other LSM-based variables (e.g. storage and runoff) may be erroneous, we can expect a priori that successful closure may be unlikely.

Figure 6 demonstrates the water budget non-closure when using different ET estimates (also see Table 5). For all basins (except GLDAS2.1-CLSM and NLDAS-2 Mosaic), ET estimates generally result in an overall net positive water balance, indicating that ET is smaller in magnitude for models and remote sensing-based products compared with climatological water balance closure. We compare the long-term mean of the precipitation forcing used for the NLDAS-2 and GLDAS-2.1 models against the mean of the seven precipitation data sets and find that the model precipitation forcing is biased low in all basins compared to the seven data sets used in this study to calculate ET_{GRACE} (not shown). This provides a potential explanation for the greater value of ET_{GRACE} compared to many of the LSM-based ET estimates. Reasons for the greater values

Table 5
Comparison of LSM and Remote Sensing-Based ET with ET_{GRACE}

	Mi	UM	MS	CU	Ar	Co	Br	Tx	Oh	UC	Sc
GLDAS-2.1 Noah	−	−	−	−	−	−	−	−	+	−	−
GLDAS-2.1 CLSM	+	+	−	+	−	−	−	−	+	−	+
NLDAS-2 Noah	−	−	−	−	−	−	−	−	−	−	−
NLDAS-2 Mosaic	+	+	+	+	−	−	−	−	+	−	−
NLDAS-2 VIC	−	−	−	−	−	−	−	−	−	−	−
PT-JPL	−	−	−	−	−	−	−	−	−	−	−
FLUXCOM	−	−	−	−	−	−	−	−	−	−	+
MOD16	−	−	−	−	−	−	−	−	−	−	−

Note. + (−) indicates that the long-term mean ET is greater (smaller) than ET_{GRACE}.

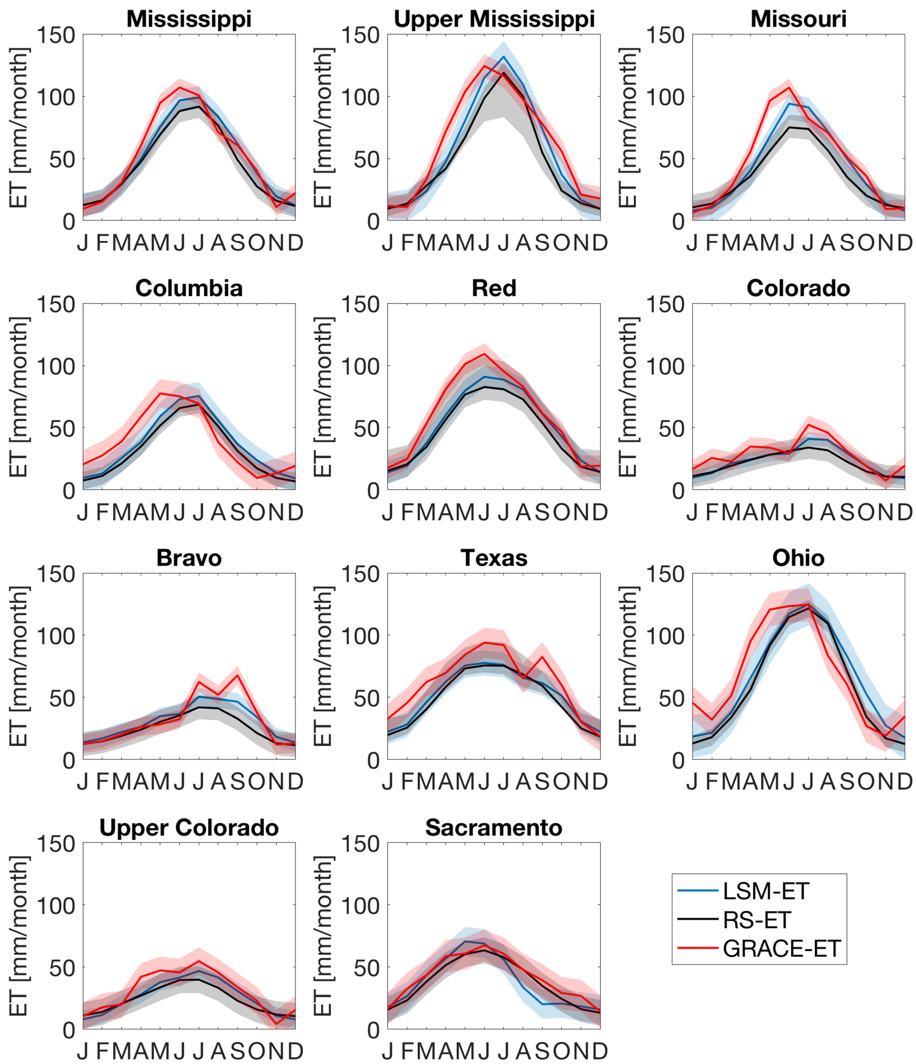

Figure 7. Seasonal cycle of ET including the uncertainty for ET_{GRACE}, LSM-based ET, and remote sensing-based ET annually averaged over 2003–2014. Shading indicates the uncertainty.

of NLDAS-2 Mosaic in several basins could be attributed to other differences in model forcing not explored in this study.

Next, we plot the mean seasonal cycle of ET_{GRACE}, LSM-based and remote sensing-based ET along with the monthly uncertainty (described in section 2.5) (Figure 7). ET_{GRACE} tends to exceed mean LSM-based and remote sensing-based ET during the spring and summer in multiple basins (see also Table 5). The range of potential ET values (shown by the shading), however, is large, accounting for 5% (Mississippi) up to 18% (Colorado) of total monthly ET during certain months.

We also investigate the relationship between the relative uncertainty (mean uncertainty divided by mean ET) for each of the ET products against basin size, aridity, and forest cover (not shown) as these variables have been found to be related to ET uncertainty (Long et al., 2014; Velpuri et al., 2013). Relative error was found to only be significantly related to basin aridity and forest cover for remote sensing-based ET. We did not find any significant relationships for LSM-based and ET_{GRACE} with any of the variables.

We then calculate the long-term mean bias against ET_{GRACE} of LSM-based and remote sensing-based ET as

$$\epsilon_{bias} = \overline{ET_{product}} - \overline{ET_{GRACE}}. \tag{4}$$

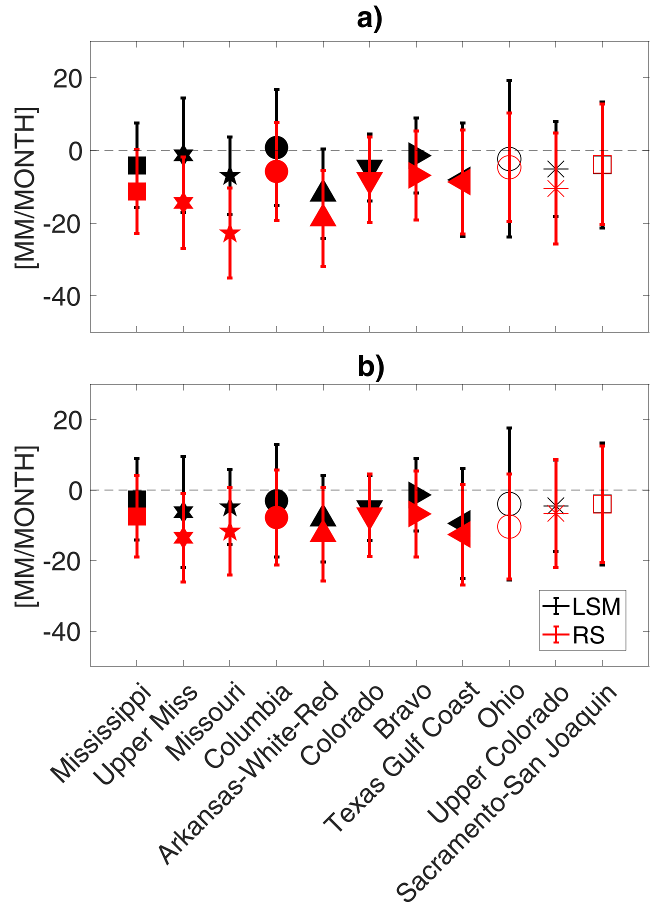

Figure 8. Top: (a) Monthly bias of model (black shapes) and remote sensing-based derived (red shapes) ET for mean spring–summer (May to August). Bottom: (b) Long-term mean annual ET bias for each basin. Error bars indicate the uncertainty for the bias (which is calculated as the sum of the variance of the error for the ET estimate (model or remote sensing-based) and ET_{GRACE}). Biases are calculated for mean ET over 2003 to 2014.

Bias is calculated for long-term annual (Figure 8, bottom panels) and spring-summer ET (Figure 8, top panels). Uncertainties in the bias estimates are calculated as the sum of the variances of the LSM-based and remotely sensed products with the error variance of ET_{GRACE} and is plotted as error bars.

Overall, we find that all basins are negatively biased when compared with ET_{GRACE} for the annual mean. For spring to summer (May to August), we find that LSM-based and remote sensing-based ET is also lower for all basins, except for the Columbia LSM-based ET which is slightly greater than ET_{GRACE} (however, this bias is not significant) (Figure 8, top panel).

We note that the uncertainty in the bias estimates (as indicated by the error bars) is large and increases for the spring–summer with uncertainty of up to 20 mm month^{-1} (Columbia, Ohio, and Sacramento-San Joaquin). These large uncertainties indicate the biases could change sign based on the particular model or remote sensing-based estimate used. The only estimates that are significantly negatively biased for May to August (with a negative sign over the full range of uncertainty) are remote sensing-based ET during the spring and summer for the Upper Mississippi, Missouri, and Arkansas-White-Red and for LSM-based Arkansas-White Red. For the annual mean, the only significantly negative biased products are for remote sensing-based Upper Mississippi and Texas Gulf Coast.

3.3. Interannual Variability

To investigate how closely LSM-based and remote sensing-based ET agree with ET_{GRACE} on interannual time scales, we removed the seasonal cycle from the time series of each (with each of the time series smoothed at 5 months) (Figure 9). The monthly interannual time series exhibit a large amount of variability in ET_{GRACE} compared to the LSM and remote sensing-based estimates despite smoothing.

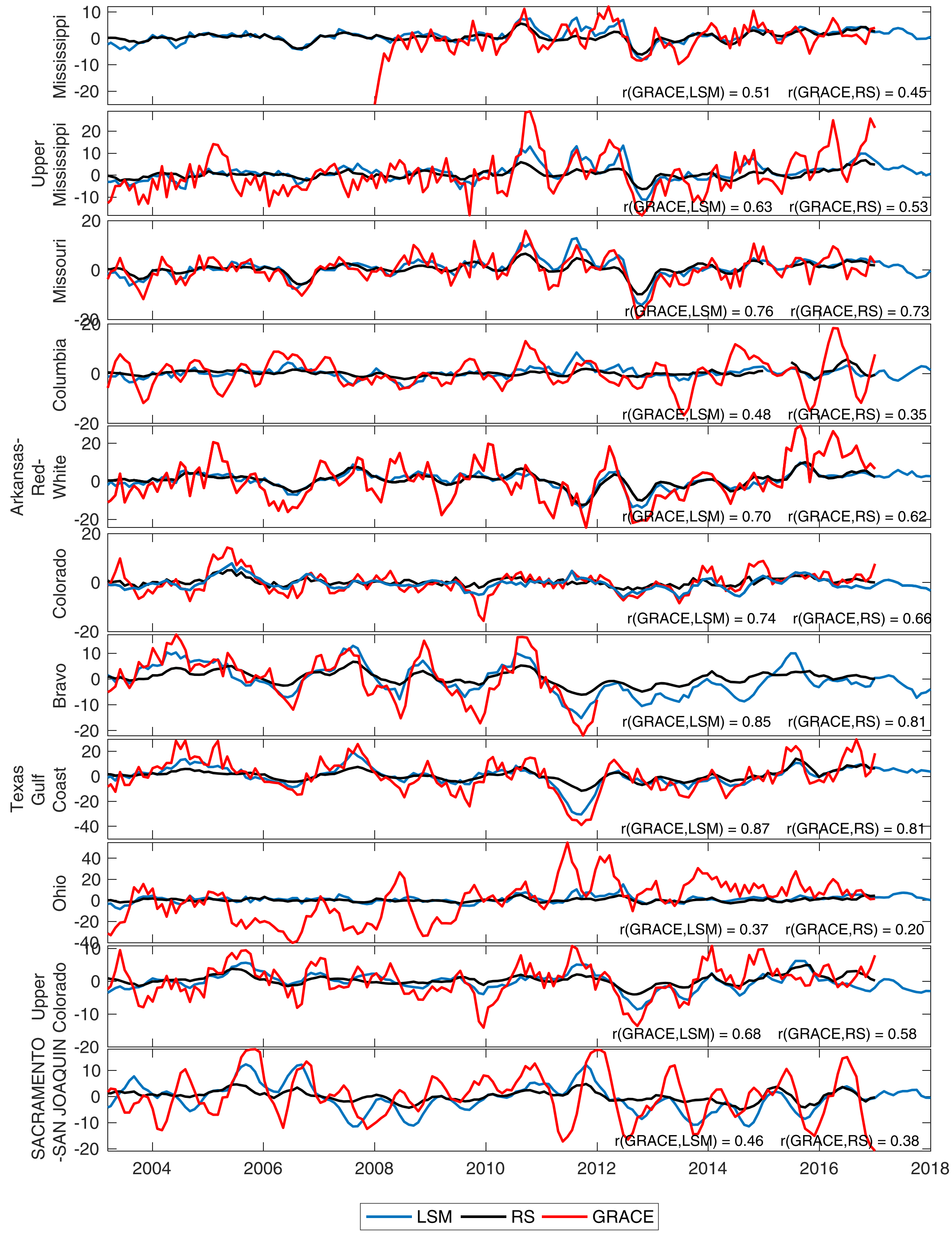

Figure 9. Time series of monthly ET_{GRACE}, LSM, and remote sensing-based ET with seasonal cycle removed. All time series have been smoothed by 5 months. In each figure, we also include the r value for the correlation between ET_{GRACE} and LSM, and ET_{GRACE} and RS.

We correlate the ET_{GRACE} series with the LSM-based and remote sensing-based ET and find that in all basins, the r value was greater for the correlations with the LSM-based products. The basins that have the greatest agreement with ET_{GRACE} in interannual variability ($r > 0.80$) are the Bravo ($r = 0.85$ [LSM], $r = 0.82$ [RS]) and the Texas Gulf Coast basin ($r = 0.85$ [LSM], $r = 0.80$ [RS]). The basins with the least agreement ($r < 0.50$) are the Columbia basin ($r = 0.46$ [LSM], $r = 0.35$ [RS]), Ohio basin ($r = 0.37$ [LSM], $r = 0.20$ [RS]) and Sacramento-San Joaquin ($r = 0.46$ [LSM], $r = 0.38$ [RS]).

We also create multiple linear regressions for interannual ET_{GRACE} using the interannual variability in precipitation, Q, and $\frac{dS}{dt}$ (not shown). We find that in all basins, precipitation explains most of the variability in interannual ET_{GRACE} ranging from 48% (Columbia) to over 83% (Bravo), with most basins above 65% (Mississippi, Upper Mississippi, Missouri, Arkansas-White-Red, Colorado, Bravo, Texas, and Upper Colorado).

We note that Ohio basin seasonal cycle removed ET_{GRACE} is increasing by 2.52 mm year^{-1} over the period 2003 to 2014 (positive trend significant using a Mann-Kendall test at $\alpha = 0.05$). A positive trend also exists for seasonal cycle removed LSM-based (0.29 mm year^{-1}, significant using a Mann-Kendall test at $\alpha = 0.05$) and remote sensing-based ET (0.06 mm year^{-1}, not significant using a Mann-Kendall test at $\alpha = 0.05$). The positive trend in ET_{GRACE} is primarily attributed to a negative trend in outlet Q (-2.88 mm year^{-1}) over this time period (negative trend significant using a Mann-Kendall test at $\alpha = 0.05$). Trends in precipitation and $\frac{dS}{dt}$ are not found to be significant.

4. Discussion

We find that ET_{GRACE} is greater in magnitude than the aggregated model and remote sensing-based ET both annually and during the summer months of peak ET. In all cases, we find that aggregated LSM-based ET is persistently greater than remote sensing-based ET, consistent with the literature (Gao et al., 2010). This is likely due to the MOD16 product which has lower magnitude ET for the majority of basins. Knowledge of these biases could also inform future analyses of ET at basin scales. We also find that the precipitation forcing data used to drive the NLDAS-2 and GLDAS-2.1 models is biased low compared to the mean of the seven precipitation data sets we use to calculate ET_{GRACE} for all basins. This suggests a potential explanation for the greater magnitude of ET_{GRACE} compared to LSM-based estimates. Previous studies attributed the higher ET_{GRACE} values to irrigation and ground water extraction, which are not represented in LSMs or remote sensing-based ET estimates (Castle et al., 2016; Pan et al., 2017). Our results are generally consistent conceptually; however, we also find that the uncertainty associated with the negative biases is very large and therefore on the basis of our findings, cannot claim definitively that ET_{GRACE} is significantly larger in magnitude, nor attribute a human influence.

In addition to providing a mass constraint on basin ET, ET_{GRACE} estimates also have the potential to capture hydrologic features not present in the LSM and remote sensing-based ET products. ET_{GRACE} calculated for the Texas Gulf Coast, Colorado, Upper Colorado, and Bravo river basins demonstrates two-peaked patterns, which were also found to be present in the seasonal cycles of precipitation for these basins. Each of these basins has a winter peak in precipitation followed by a summer peak which is related to the North American summer monsoon (Adams & Comrie, 1997). Given the dominance of precipitation in the calculation of the water balance, and particularly in arid basins (Ukkola & Prentice, 2013), we suggest that these represent actual features of the seasonal cycle. In addition, we find that the Ohio river basin has a small ET_{GRACE} peak in the winter months (December–January) which also appears in the seasonal cycle of precipitation and is not present in the LSM or remote sensing-based estimates.

Error in ET_{GRACE} is from $\frac{dS}{dt}$, precipitation, and runoff, and analysis of the error budget for each basin indicates that precipitation uncertainty is the greatest component of the error budget for larger basins such as the Mississippi, Upper Mississippi, Missouri, and Columbia, in agreement with previous studies (Gao et al., 2010). Precipitation error is known to increase in regions with complex terrain (such as the mountainous western United States) and sparse rain gage measurements (Gibson et al., 2019). Despite these errors, by using an ensemble of precipitation products, we present a range of ET values from our water balance calculation which combine both gage based and satellite precipitation measurements. For the smaller river basins, the uncertainty in $\frac{dS}{dt}$ increases due to the coarse resolution of the GRACE mascons (Wiese et al., 2016), representing a caveat in using GRACE for basins which are much smaller than the GRACE mascon resolution. We also note that although uncertainty from ET_{GRACE} is similar in magnitude to that in LSMs and remote sensing-based products used in this study, given that ET_{GRACE} is based on observed variables and conserves mass, the actual value of ET lies within the range of uncertainty of ET_{GRACE}.

On interannual time scales, ET_{GRACE} generally has greater variability than the LSM and remote sensing-based products. This is attributed to the interannual variability of precipitation, which explains most of the variability in ET_{GRACE} (explaining 65% or more of ET variability for eight of the eleven basins in this study), and this is also in agreement with previous studies (Liu et al., 2016; Swann & Koven, 2017; Ukkola & Prentice, 2013). This finding suggests using caution in the interpretation of interannual variability from LSM and remote sensing-based ET products, which may not adequately represent climate variability.

5. Conclusion

We calculate ET using GRACE TWSA, USGS runoff and seven different precipitation data sets in a mass conservation approach. We then use ET_{GRACE} and the associated uncertainties as a best estimate of the total evaporative water flux at the basin scale and over long time periods for eleven basins in the United States, extending the work of previous authors (Liu et al., 2016; Long et al., 2014; Swann & Koven, 2017). Although limited by the input data, this method provides a mass conservation constraint on basin-scale ET fluxes, enabling ET_{GRACE} to evaluate LSM and remote sensing products, which is not otherwise possible due to the difficulties associated with upscaling sparse eddy covariance measurements (Baldocchi, 2003).

As we move to an era of increasing societal concern about water resources sustainability under the demands of increasing population and changing climate, it is important to be able to provide scientific analysis at the scales that matter for water policy. As hydrologists, we know that the fundamental spatial basis for hydrological fluxes and stores is at the scale of the basin. At these scales, a river basin aggregates precipitation into a river network, through which human activities such as water storage, diversion, and irrigation are made possible. These concepts are fundamental to the modern science of hydrology and water resources management.

However, no existing ET product is really built to operate at the basin scale. ET estimated through proxy methods such as remote sensing or the upscaling of in situ measurements contains biases that get multiplied and magnified in summation (Baldocchi, 2003). The unique value of the GRACE observations is that they exhibit higher accuracy with increasing spatial domain; as the basin gets bigger, GRACE works better (Wiese et al., 2016). Therefore, GRACE provides a check on the other existing upscaled ET approaches that can inform developers of accumulated errors. Also, GRACE provides the best estimate of information at the scales most relevant for water policy: for decision making over large domains and long timescales that affect our ability to cope with increasing demands and variable supplies.

To conclude, the water balance method is useful in the evaluation of basin-scale ET fluxes as it provides a mass conservation constraint. Given the large number of ET products currently available (Fisher et al., 2011; Miralles et al., 2016), and the contrasting nature of current ET trends reported in the literature (Jung et al., 2010; Zhang et al., 2015), evaluating basin-scale ET estimates is of importance for water resources, and may help inform the development of LSM and remote sensing-based products.

Acknowledgments
The research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. California Institute of Technology. Government sponsorship acknowledged. MPC was supported by an appointment to the NASA Postdoctoral Program at the Jet Propulsion Laboratory, administered by Universities Space Research Association under contract with NASA. JBF was supported in part by NASA ECOSTRESS and NASA SUSMAP.

All data in this work is publicly available and described in the text. The author thanks Martin Jung for providing access to the FLUXCOM data set. The author thanks A. J. Purdy for help acquiring the PT-JPL data. The author also thanks David N. Wiese for the help with processing the uncertainty of the GRACE mascon data. The authors thank the anonymous reviewers for the helpful comments on the manuscript.

References

Adams, D. K., & Comrie, A. C. (1997). The North American Monsoon. *Bulletin of the American Meteorological Society*, *78*(10), 2197–2213. https://doi.org/10.1175/1520-0477(1997)078<2197:TNAM>2.0.CO;2

Badgley, G., Fisher, J. B., Jiménez, C., Tu, K. P., & Vinukollu, R. (2015). On uncertainty in global terrestrial evapotranspiration estimates from choice of input forcing datasets. *Journal of Hydrometeorology*, *16*(4), 1449–1455.

Baldocchi, D. (2003). Assessing the eddy covariance technique for evaluating carbon dioxide exchange rates of ecosystems: Past, present and future. *Global Change Biology*, *9*(4), 479–492.

Baldocchi, D., Falge, E., Gu, L., Olson, R., Hollinger, D., Running, S., 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. *Bulletin of the American Meteorological Society*, *82*(11), 2415–2434. https://doi.org/10.1175/1520-0477(2001)082<2415:FANTTS>2.3.CO;2

Castle, S. L., Reager, J. T., Thomas, B. F., Purdy, A. J., Lo, M.-H., Famiglietti, J. S., & Tang, Q. (2016). Remote detection of water management impacts on evapotranspiration in the Colorado River Basin. *Geophysical Research Letters*, *43*, 5089–5097. https://doi.org/10.1002/2016GL068675

Castle, S. L., Thomas, B. F., Reager, J. T., Rodell, M., Swenson, S. C., & Famiglietti, J. S. (2014). Groundwater depletion during drought threatens future water security of the Colorado River Basin. *Geophysical Research Letters*, *41*, 5904–5911. https://doi.org/10.1002/2014GL061055

Chen, M., Shi, W., Xie, P., Silva, V. B., Kousky, V. E., Wayne Higgins, R., & Janowiak, J. E. (2008). Assessing objective techniques for gauge-based analyses of global daily precipitation. *Journal of Geophysical Research*, *113*, D04110. https://doi.org/10.1029/2007JD009132

Dai, A., & Zhao, T. (2017). Uncertainties in historical changes and future projections of drought. Part I: Estimates of historical drought changes. *Climatic Change*, *144*(3), 519–533. https://doi.org/10.1007/s10584-016-1705-2

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

Daly, C., Neilson, R. P., & Phillips, D. L. (1994). A statistical-topographic model for mapping climatological precipitation over mountainous terrain. *Journal of Applied Meteorology*, *33*(2), 140–158.

Dong, B., & Dai, A. (2017). The uncertainties and causes of the recent changes in global evapotranspiration from 1982 to 2010. *Climate Dynamics*, *49*(1-2), 279–296.

Fekete, BalázsM., & Vörösmarty, C. J. (2007). The current status of global river discharge monitoring and potential new technologies complementing traditional discharge measurements. *IAHS publ*, *309*, 129–136.

Fisher, J. B., Melton, F., Middleton, E., Hain, C., Anderson, M., Allen, R., et al. (2017). The future of evapotranspiration: Global requirements for ecosystem functioning, carbon and climate feedbacks, agricultural management, and water resources. *Water Resources Research*, *53*, 2618–2626. https://doi.org/10.1002/2016WR020175

Fisher, J. B., Tu, K. P., & Baldocchi, D. D. (2008). Global estimates of the land–atmosphere water flux based on monthly AVHRR and ISLSCP-II data, validated at 16 FLUXNET sites. *Remote Sensing of Environment*, *112*(3), 901–919. https://doi.org/10.1016/j.rse.2007.06.025

Fisher, J. B., Whittaker, R. J., & Malhi, Y. (2011). ET come home: Potential evapotranspiration in geographical ecology. *Global Ecology and Biogeography*, *20*(1), 1–18.

Gao, H., Tang, Q., Ferguson, C. R., Wood, E. F., & Lettenmaier, D. P. (2010). Estimating the water budget of major US river basins via remote sensing. *International Journal of Remote Sensing*, *31*(14), 3955–3978.

Gehne, M., Hamill, T. M., Kiladis, G. N., & Trenberth, K. E. (2016). Comparison of global precipitation estimates across a range of temporal and spatial scales. *Journal of Climate*, *29*(21), 7773–7795.

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., et al. (2017). The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2). *Journal of Climate*, *30*(14), 5419–5454. https://doi.org/10.1175/JCLI-D-16-0758.1

Gibson, P. B., Waliser, D. E., Lee, H., Tian, B., & Massoud, E. (2019). Climate model evaluation in the presence of observational uncertainty: Precipitation indices over the contiguous United States. *Journal of Hydrometeorology*, *20*, 1339–1357.

Güntner, A. (2008). Improvement of global hydrological models using GRACE data. *Surveys in geophysics*, *29*(4-5), 375–397.

Harris, I., Jones, P. D., Osborn, T. J., & Lister, D. H. (2014). Updated high-resolution grids of monthly climatic observations—The CRU TS3.10 Dataset. *International Journal of Climatology*, *34*(3), 623–642. https://doi.org/10.1002/joc.3711

Higgins, R., Shi, W., Yarosh, E., & Joyce, R. (2000). *Improved United States precipitation quality control system and analysis NCEP/Climate Prediction Center ATLAS No. 7*. NOAA, pp. 40.

Huntington, T. G. (2006). Evidence for intensification of the global water cycle: Review and synthesis. *Journal of Hydrology*, *319*(1), 83–95. https://doi.org/10.1016/j.jhydrol.2005.07.003

Jung, M., Koirala, S., Weber, U., Ichii, K., Gans, F., Camps-Valls, G., et al. (2019). The FLUXCOM ensemble of global land-atmosphere energy fluxes. *Scientific Data*, *6*(1), 74.

Jung, M., Reichstein, M., Ciais, P., Seneviratne, S. I., Sheffield, J., Goulden, M. L., et al. (2010). Recent decline in the global land evapotranspiration trend due to limited moisture supply. *Nature*, *467*(7318), 951.

Koster, R. D., Suarez, M. J., Higgins, R. W., & Van den Dool, H. M. (2003). Observational evidence that soil moisture variations affect precipitation. *Geophysical Research Letters*, *30*(5), 1241. https://doi.org/10.1029/2002GL016571

Kumar, S., Holmes, T., Mocko, D., Wang, S., & Peters-Lidard, C. (2018). Attribution of flux partitioning variations between land surface models over the continental US. *Remote Sensing*, *10*(5), 751.

Landerer, F. W., Dickey, J. O., & Güntner, A. (2010). Terrestrial water budget of the Eurasian pan-Arctic from GRACE satellite measurements during 2 003–2009. *Journal of Geophysical Research*, *115*, D23115. https://doi.org/10.1029/2010JD014584

Landerer, F. W., & Swenson, S. (2012). Accuracy of scaled GRACE terrestrial water storage estimates. *Water Resources Research*, *48*.

Liu, W., Wang, L., Zhou, J., Li, Y., Sun, F., Fu, G., et al. (2016). A worldwide evaluation of basin-scale evapotranspiration estimates against the water balance method. *Journal of Hydrology*, *538*, 82–95.

Long, D., Longuevergne, L., & Scanlon, B. R. (2014). Uncertainty in evapotranspiration from land surface modeling, remote sensing, and GRACE satellites. *Water Resources Research*, *50*, 1131–1151. https://doi.org/10.1002/2013WR014581

McCabe, M. F., & Wood, E. F. (2006). Scale influences on the remote estimation of evapotranspiration using multiple satellite sensors. *Remote Sensing of Environment*, *105*(4), 271–285.

Milly, P. C. D., & Dunne, K. A. (2010). On the hydrologic adjustment of climate-model projections: The potential pitfall of potential evapotranspiration. *Earth Interactions*, *15*(1), 1–14. https://doi.org/10.1175/2010EI363.1

Miralles, D. G., Jiménez, C., Jung, M., Michel, D., Ershadi, A., McCabe, M., et al. (2016). The WACMOS-ET project-Part 2: Evaluation of global terrestrial evaporation data sets. *Hydrology and Earth System Sciences*, *20*(2), 823–842.

Mu, Q., Heinsch, F. A., Zhao, M., & Running, S. W. (2007). Development of a global evapotranspiration algorithm based on MODIS and global meteorology data. *Remote sensing of Environment*, *111*(4), 519–536.

Mu, Q., Zhao, M., & Running, S. W. (2011). Improvements to a MODIS global terrestrial evapotranspiration algorithm. *Remote Sensing of Environment*, *115*(8), 1781–1800.

Pan, Y., Zhang, C., Gong, H., Yeh, P. J.-F., Shen, Y., Guo, Y., et al. (2017). Detection of human-induced evapotranspiration using GRACE satellite observations in the Haihe River basin of China. *Geophysical Research Letters*, *44*, 190–199. https://doi.org/10.1002/2016GL071287

Penman, H. L. (1948). Natural evaporation from open water, bare soil and grass. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, *193*(1032), 120–145.

Polhamus, A., Fisher, J. B., & Tu, K. P. (2013). What controls the error structure in evapotranspiration models? *Agricultural and forest meteorology*, *169*, 12–24.

Priestley, C. H. B., & Taylor, R. (1972). On the assessment of surface heat flux and evaporation using large-scale parameters. *Monthly Weather Review*, *100*(2), 81–92.

Ramillien, G., Frappart, F., Güntner, A., Ngo-Duc, T., Cazenave, A., & Laval, K. (2006). Time variations of the regional evapotranspiration rate from Gravity Recovery and Climate Experiment (GRACE) satellite gravimetry. *Water Resources Research*, *42*, W10403. https://doi.org/10.1029/2005WR004331

Rodell, M., Famiglietti, J. S., Chen, J., Seneviratne, S. I., Viterbo, P., Holl, S., & Wilson, C. R. (2004). Basin scale estimates of evapotranspiration using GRACE and other observations. *Geophysical Research Letters*, *31*, L20504. https://doi.org/10.1029/2004GL020873

Rodell, M., Famiglietti, J. S., Wiese, D. N., Reager, J. T., Beaudoing, H. K., Landerer, F. W., & Lo, M. H. (2018). Emerging trends in global freshwater availability. *Nature*, *557*(7707), 651–659. https://doi.org/10.1038/s41586-018-0123-1

Rodell, M., Houser, P., Jambor, U., Gottschalck, J., & Mitchell (2004). The Global Land Data Assimilation System. *Bulletin of the American Meteorological Society*, *85*(3), 381–394.

Rodell, M., McWilliams, E. B., Famiglietti, J. S., Beaudoing, H. K., & Nigro, J. (2011). Estimating evapotranspiration using an observation based terrestrial water budget. *Hydrological Processes*, *25*(26), 4082–4092. https://doi.org/10.1002/hyp.8369

Rodell, M., Velicogna, I., & Famiglietti, J. S. (2009). Satellite-based estimates of groundwater depletion in India. *Nature*, *460*(7258), 999.

Sauer, V. B., & Meyer, R. (1992). Determination of error in individual discharge measurements (*Tech. Rep.*): US Geological Survey; Books and Open-File Reports Section [distributor].

Schneider, U., Becker, A., Finger, P., Meyer-Christoffer, A., Rudolf, B., & Ziese, M. (2016). GPCC full data reanalysis version 7.0: Monthly land-surface precipitation from rain gauges built on GTS based and historic data. https://doi.org/10.5065/D6000072

Senay, G. B., Leake, S., Nagler, P. L., Artan, G., Dickinson, J., Cordova, J. T., & Glenn, E. P. (2011). Estimating basin scale evapotranspiration (ET) by water balance and remote sensing methods. *Hydrological Processes*, *25*(26), 4037–4049. https://doi.org/10.1002/hyp.8379

Swann, A. L., & Koven, C. D. (2017). A direct estimate of the seasonal cycle of evapotranspiration over the Amazon basin. *Journal of Hydrometeorology*, *18*(8), 2173–2185.

Tapley, B. D., Bettadpur, S., Ries, J. C., Thompson, P. F., & Watkins, M. M. (2004). GRACE measurements of mass variability in the Earth system. *Science*, *305*(5683), 503–505.

Thomas, B. F., Landerer, F. W., Wiese, D. N., & Famiglietti, J. S. (2016). A comparison of watershed storage trends over the eastern and upper Midwestern regions of the United States, 2003–2015. *Water Resources Research*, *52*, 6335–6347. https://doi.org/10.1002/2016WR018617

Thornthwaite, C. W. (1948). An approach toward a rational classification of climate. *Geographical Review*, *38*, 55–94.

Trenberth, K. E., Fasullo, J. T., & Kiehl, J. (2009). Earth's global energy budget. *Bulletin of the American Meteorological Society*, *90*(3), 311–324.

Ukkola, A., & Prentice, I. (2013). A worldwide analysis of trends in water-balance evapotranspiration. *Hydrology and Earth System Sciences*, *17*(10), 4177–4187.

Velpuri, N. M., Senay, G. B., Singh, R. K., Bohms, S., & Verdin, J. P. (2013). A comprehensive evaluation of two MODIS evapotranspiration products over the conterminous United States: Using point and gridded FLUXNET and water balance ET. *Remote Sensing of Environment*, *139*, 35–49.

Vose, R. S., Applequist, S., Squires, M., Durre, I., Menne, M. J., Williams, C. N., et al. (2014). Improved historical temperature and precipitation time series for U.S. climate divisions. *Journal of Applied Meteorology and Climatology*, *53*(5), 1232–1251. https://doi.org/10.1175/JAMC-D-13-0248.1

Wang, K., & Dickinson, R. (2012). A review of global terrestrial evapotranspiration: Observation, modeling, climatology, and climatic variability. *Reviews of Geophysics*, *50*, RG2005. https://doi.org/10.1029/2011RG000373

Wang, W., Xing, W., & Shao, Q. (2015). How large are uncertainties in future projection of reference evapotranspiration through different approaches?. *Journal of Hydrology*, *524*, 696–700. https://doi.org/10.1016/j.jhydrol.2015.03.033

Watkins, M. M., Wiese, D. N., Yuan, D.-N., Boening, C., & Landerer, F. W. (2015). Improved methods for observing Earth's time variable mass distribution with GRACE using spherical cap mascons. *Journal of Geophysical Research*, *120*, 2648–2671. https://doi.org/10.1002/2014JB011547

Wiese, D. N., Landerer, F. W., & Watkins, M. M. (2016). Quantifying and reducing leakage errors in the JPL RL05M GRACE mascon solution. *Water Resources Research*, *52*, 7490–7502. https://doi.org/10.1002/2016WR019344

Xia, Y., Mitchell, K., Ek, M., Cosgrove, B., Sheffield, J., Luo, L., et al. (2012). 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. *Journal of Geophysical Research*, *117*, D03110. https://doi.org/10.1029/2011JD016051

Xie, P., Chen, M., Yang, S., Yatagai, A., Hayasaka, T., Fukushima, Y., & Liu, C. (2007). A gauge-based analysis of daily precipitation over East Asia. *Journal of Hydrometeorology*, *8*(3), 607–626.

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

Zhang, K., Kimball, J. S., Nemani, R. R., Running, S. W., Hong, Y., Gourley, J. J., & Yu, Z. (2015). Vegetation greening and climate change promote multidecadal rises of global land evapotranspiration. *Scientific Reports*, *5*, 15956. https://doi.org/10.1038/srep15956