AGU PUBLICATIONS

Water Resources Research

RESEARCH ARTICLE
10.1002/2015WR017311

Water balance-based actual evapotranspiration reconstruction from ground and satellite observations over the conterminous United States

Zhanming Wan[1,2], Ke Zhang[1,3,4], Xianwu Xue[1,2], Zhen Hong[1,2], Yang Hong[1,2,5], and Jonathan J. Gourley[6]

[1]Hydrometeorology & Remote Sensing (HyDROS) Laboratory, School of Civil Engineering and Environmental Sciences, University of Oklahoma, Norman, Oklahoma, USA, [2]Advanced Radar Research Center and Center for Spatial Analysis, National Weather Center, Norman, Oklahoma, USA, [3]Cooperative Institute for Mesoscale Meteorological Studies, University of Oklahoma, Norman, Oklahoma, USA, [4]State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Hohai University, Nanjing, Jiangsu, China, [5]Department of Hydraulic Engineering, Tsinghua University, Beijing, China, [6]NOAA/National Severe Storms Laboratory, Norman, Oklahoma, USA

Abstract The objective of this study is to produce an observationally based monthly evapotranspiration (ET) product using the simple water balance equation across the conterminous United States (CONUS). We adopted the best quality ground and satellite-based observations of the water budget components, i.e., precipitation, runoff, and water storage change, while ET is computed as the residual. Precipitation data are provided by the bias-corrected PRISM observation-based precipitation data set, while runoff comes from observed monthly streamflow values at 592 USGS stream gauging stations that have been screened by strict quality controls. We developed a land surface model-based downscaling approach to disaggregate the monthly GRACE equivalent water thickness data to daily, 0.125° values. The derived ET computed as the residual from the water balance equation is evaluated against three sets of existing ET products. The similar spatial patterns and small differences between the reconstructed ET in this study and the other three products show the reliability of the observationally based approach. The new ET product and the disaggregated GRACE data provide a unique, important hydro-meteorological data set that can be used to evaluate the other ET products as a benchmark data set, assess recent hydrological and climatological changes, and terrestrial water and energy cycle dynamics across the CONUS. These products will also be valuable for studies and applications in drought assessment, water resources management, and climate change evaluation.

Key Points:
- Applied land surface models to disaggregate GRACE data
- Developed water balance-based approach to estimate ET
- Produced a high quality observational-based ET record

Correspondence to:
K. Zhang,
kezhang@ou.edu;
Y. Hong,
yanghong@ou.edu

Citation:
Wan, Z., K. Zhang, X. Xue, Z. Hong, Y. Hong, and J. J. Gourley (2015), Water balance-based actual evapotranspiration reconstruction from ground and satellite observations over the conterminous United States, *Water Resour. Res.*, *51*, 6485–6499, doi:10.1002/2015WR017311.

Received 31 MAR 2015
Accepted 18 JUL 2015
Accepted article online 22 JUL 2015
Published online 21 AUG 2015

1. Introduction

As one of the major components of the global hydrologic cycle, evapotranspiration (ET) is a complicated process and composed of evaporation from land surface and water bodies, and transpiration from vegetation to the atmosphere [*Allen et al.*, 1998]. Evaporation and transpiration processes occur simultaneously and are difficult to separate [*Anderson et al.*, 2007; *Liu et al.*, 2011; *Mallick et al.*, 2014]. Accurately estimating actual ET is of great importance because it is a crucial variable in water resources management, agriculture, and ecology [*Khan et al.*, 2010], and an important process in the fields of hydrology, meteorology and atmospheric sciences [*Chauhan and Shrivastava*, 2009].

Several approaches have been developed to estimate actual ET, including meteorology-driven diagnostic models such as the Penman-Monteith (PM) method [*Monteith*, 1965], satellite data-driven PM approaches [*Cleugh et al.*, 2007; *Mu et al.*, 2007; *Zhang et al.*, 2008, 2009, 2010], satellite data-driven Priestly-Taylor empirical approach [*Fisher et al.*, 2008], energy balance methods [*Bastiaanssen et al.*, 1998; *Su*, 2002; *Wang and Bras*, 2009, 2011], vegetation index-ET empirical relationship methods [*Gillies et al.*, 1997; *Nishida et al.*, 2003; *Tang et al.*, 2009], and data-driven statistical methods [*Jung et al.*, 2010]. The water balance approach is another way to determine ET by quantifying it as the residual in the water balance equation. This method is simple and sound in theory, and warrants accurate estimate of ET as long as the other water components can be accurately measured. Additionally, unlike the other approaches, it does not require additional meteorological inputs except precipitation. One good example for measuring/estimating ET using the water balance

approach is the lysimeter. The water balance method has been used to estimate ET in previous studies [*Long et al*., 2014; *Ramillien et al*., 2006; *Zeng et al*., 2012; *Zhang et al*., 2010], but this approach is usually applied to one or multiple basins to derive the areal-mean ET of these basins that serve as an ET validation data set.

The recent ET estimates by model simulations and satellite-driven algorithms are usually evaluated against point FLUXNET eddy covariance measurements [*Mu et al*., 2007, *Velpuri et al*., 2013; *Zhang et al*., 2009] and simulations from land surface models [*Jung et al*., 2010; *Schwalm et al*., 2013]. Few of these studies use basin-wide ET estimates from water balance computations as benchmark values to evaluate the remotely sensed ET estimates [*Zeng et al*., 2012; *Zhang et al*., 2010]. The water balance-based ET is rarely available, covers few regions, and has coarse spatial resolution due to the limited data availability and continuity.

To produce a subbasin-wide ET product with continuous temporal coverage and downscaled gridded water storage change data with a relatively finer spatial resolution (0.125°), we utilized the trustworthy ground and satellite-observed hydrological data provided by USGS, NASA, and USDA to estimate monthly actual ET and monthly 0.125° water storage change data from April 2002 to September 2013 across the conterminous United States (CONUS). The method developed in this study computes actual ET as the residual in the simple water balance equation. The objective of this study is to produce an observationally based monthly evapotranspiration (ET) product using the simple water balance equation across the CONUS. This data set can be used to evaluate the other ET products as a benchmark data set, assess recent hydrological and climatological changes across the CONUS. These products will be also valuable for studies and applications in drought assessment, water resources management, and climate change evaluation.

2. Data and Methodology

2.1. Study Area and Data

The spatial domain of this study is the CONUS, ranging from 25°N to 50°N and from 124.75°W to 67°W (Figure 1). The data used in this study include observations of precipitation, runoff, and water storage change from ground and satellite data, and river network and topographical data from a remote sensing-derived digital elevation model (DEM). The river network data have a spatial resolution of 0.125° and were derived from an upscaled global data set from the combined HydroSHEDS and HYDRO1K datasets [*Wu et al*., 2012].

The precipitation data are from the PRISM (Parameter-elevation Regressions on Independent Slopes Model) daily precipitation data set produced by the PRISM group at Oregon State University (http://www.prism.oregonstate.edu). The PRISM daily precipitation product is a 4 km gridded estimate of precipitation for the CONUS based on observations from a wide range of monitoring networks with sophisticated quality control, and bias and topography corrections [*Daly et al*., 2008]. The PRISM interpolation method calculates climate–elevation regression for each grid cell, and stations entering the regression are assigned weights based primarily on the physiographic similarity of the station to the grid cell. Factors considered are location, elevation, coastal proximity, topographic facet orientation, vertical atmospheric layer, topographic position, and orographic enhancement caused by the underlying terrain [*Daly et al*., 2008]. The PRISM data set is the source of USDA's official climatological data. In this study, all analyses were conducted on a geographical grid with a resolution of 0.125°. Therefore, the PRISM precipitation was first aggregated from 4 km to 0.125° and then summed from daily values to monthly values.

Monthly mean streamflow observations from all USGS stream gauging stations, which have continuous discharge data between April 2002 and September 2013, were chosen to derive the monthly runoff depth at the subbasin level. Some of these stations were further screened out if differences between their drainage areas as provided by USGS metadata and the areas derived from the 0.125° DEM-based flow accumulation are larger than 20%. If multiple streamflow measurement stations fall in the same 0.125° grid cell, only the station with the largest drainage area was kept for further analysis. The drainage area of each station must contain at least two 0.125° grid cells. After the strict screening process, streamflow data from 592 USGS stations were chosen for further analysis (Figure 1).

Monthly equivalent water thickness (EWT) of water storage is provided by the Gravity Recovery and Climate Experiment (GRACE) satellite-derived data set. GRACE is a twin-satellite mission launched in March 2002 to observe the variation of Earth's gravity field anomalies. GRACE satellites provide information on changes in the gravity fields, which are controlled primarily by variations in water distribution and are used to derive

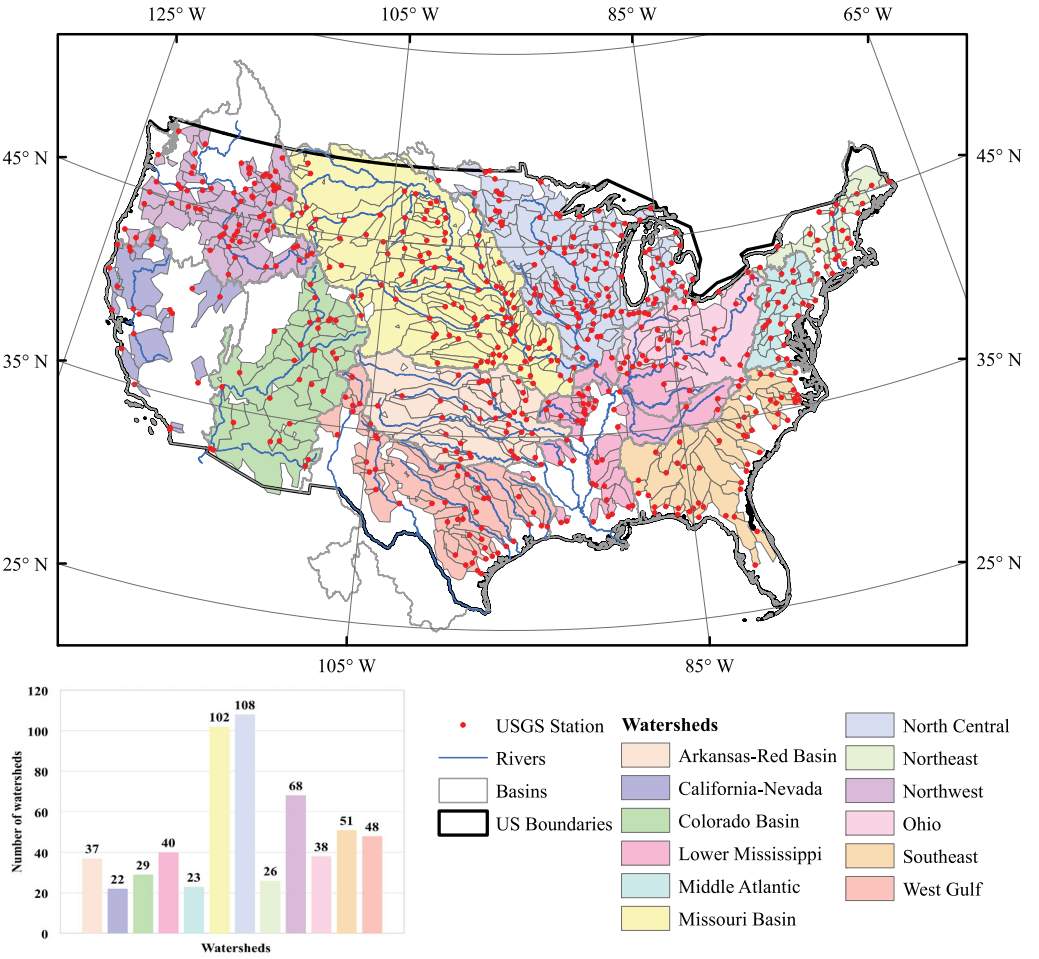

Figure 1. Locations of 592 USGS stream gauging stations used in this study and spatial distributions of their corresponding subbasins over the CONUS; the blank areas are regions without sufficient good-quality observational data.

terrestrial water storage change at a spatial resolution of $\geq$200,000 km^2 [*Tapley et al.*, 2004b]. The latest GRACE land grid data Release-05 (RL05) released in February 2014 is used in this study. The RL05 is a level-3 GRACE product containing the EWT product in centimeters with a spatial resolution of 1° × 1° [*Chambers*, 2006]. This gridded data set was converted from sets of spherical harmonic coefficients of the standard GRACE product describing the monthly variations in Earth's gravity filed after applying a series of GRACE filters [*Swenson and Wahr*, 2006; *Wahr et al.*, 1998, 2006]. Gridded scaling factors are also applied to the gridded GRACE EWT to minimize the leakage error due to resampling and postprocessing, i.e., the filtering and smoothing processes [*Landerer and Swenson*, 2012]. Although GRACE provides an opportunity to better constrain the water budget equation, it has relatively coarse spatial resolution and suffers periodic data gaps due to battery management issues and during certain orbit periods (http://grace.jpl.nasa.gov/data/gracemonthlymassgridsoverview/). To achieve a continuous terrestrial water storage change data with a spatial resolution of 0.125°, we developed a downscaling approach in which the GRACE data were used to constrain the water storage thickness simulated by four land surface models (LSMs) from North American Land Data Assimilation System project phase 2 (NLDAS-2) and correct the bias in the modeled water storage thicknesses. Details of this downscaling method are described in section 2.2.

2.2. Methodology

In this study, we derived monthly areal-mean actual ET on a subbasin level using the water balance equation by assuming no net groundwater flow across the boundary of a river basin of interest:

$$ET = P - R - \Delta S + \varepsilon, \quad (1)$$

where P (mm) is the monthly precipitation; R (mm) is the monthly runoff depth; ΔS (mm) is the monthly terrestrial water storage change, i.e., change in the monthly EWT; and ε is an error term. Because the water budget terms (P, R and ΔS) are derived from ground and satellite observations, there are some measurement and processing errors in these data sets [*Daly et al*., 2008; *Landerer and Swenson*, 2012; *Swenson and Wahr*, 2006; *Tapley et al*., 2004a]. However, quantifying the error for each of the data sets for each subbasin is impractical, and we assume that the errors are random and small in magnitude relative to the values of the water balance variables. Therefore, the derived monthly ET values inherit these errors given that they are computed as the residual. The sources and detailed processing of the three water budget terms used to compute the ET are described in section 2.1 and the remaining part of this section.

2.2.1. Calculation of Subbasin Runoff Depth

Since many of these USGS streamflow measurement stations are nested within the same parent watersheds (Figure 1), we first derived the topological relationships among these stations within the same parent basins from the river network data (i.e., flow direction, flow accumulation area). The drainage areas of all neighboring upstream stations from a given station were subtracted out from the drainage area of this station so that each station was attributed to unique contributing areas, i.e., a subbasin associated to a specific station does not contain or overlap with other subbasins. For example, there are 102 stations in the Missouri river basin; therefore, the application of the above procedure produces 102 subbasins that do not overlap with each other (Figure 1).

Missing values exist in some of the 592 USGS stations for different reasons, but these data gaps must be less than 20% of the total record, else they are removed. Linear interpolation is not a good solution when the data gap encompasses 2 or more months. This is because linear interpolation can artificially smooth the fluctuation of monthly discharge values. Instead, we applied an alternative method in which the multiyear mean value of a missing month ($\overline{Q_m}$), the discharge of its nearest month (Q_n), and the multiyear mean value of the nearest month ($\overline{Q_n}$) are used to fill the missing value of the missing month (Q_m):

$$Q_m = \frac{\overline{Q_m} \times Q_n}{\overline{Q_n}}. \quad (2)$$

In essence, we assume that the ratio of monthly discharge in a missing month to its multiyear mean is equal to the ratio of monthly discharge in its nearest month to the multiyear mean discharge of the nearest month.

The monthly runoff depth of subbasin i is then computed by the following equation:

$$R_i = \frac{\left(Q_i - \sum_{n=1}^{N} Q_n\right) \times T}{A_i} \times 1000, \quad (3)$$

where R_i is the monthly runoff depth of subbasin i (mm); Q_i is the monthly discharge at station i ($m^3\ s^{-1}$); Q_n is the monthly discharge of neighbor upstream station n of station i ($m^3\ s^{-1}$); N is the total number of neighbor upstream stations for station i; A_i is the contributing land area of subbasin i (m^2); and T is time (s) in a month.

2.2.2. Downscaling of GRACE Equivalent Water Thickness Data

As we discussed previously, the GRACE data have periodic gaps and a coarse spatial resolution. To utilize the GRACE data to derive continuous, finer resolution time series of water storage change, we developed a model-based approach to downscale the GRACE data. First, hourly 0.125° simulations of the Variable Infiltration Capacity (VIC), Noah Land Surface Model (Noah), Mosaic, and Sacramento Soil Moisture Accounting (SAC) models from North American Land Data Assimilation System project phase 2 (NLDAS-2) were used in this study to estimate daily water thickness of soil water storage across the CONUS. The four models form the land surface model (LSM) ensemble executed over the CONUS in NLDAS-2 [*Xia et al*., 2012b]. The VIC model is a semidistributed grid-based land surface hydrological model, which solves for full water and energy balances [*Liang et al*., 1994, 1996]. The Noah model is a community LSM, which simulates soil moisture (both liquid and frozen), soil temperature, skin temperature, snowpack depth, snowpack water equivalent (and hence snowpack density), canopy water content, and the energy flux and water flux terms of the surface energy balance and surface water balance [*Chen et al*., 1996; *Ek et al*., 2003; *Koren et al*., 1999; *Mitchell et al*., 2004]. The Mosaic model was developed for use in NASA's global climate model and simulates

energy and energy balance, and soil moisture and temperature [*Koster et al*., 2000]. Originally formulated as a lumped conceptual hydrological model, SAC has since been converted into a distributed version and has adopted some components of the surface-vegetation-atmosphere transfer scheme developed within the coupled climate modeling community [*Koren et al*., 2007]. In the NLDAS-2 project, the VIC model is equipped with three soil layers with a fixed 10 cm top layer and two other layers with spatially varying thicknesses, while the Noah model has spatially uniform four soil layers with fixed thicknesses of 10, 30, 60, and 100 cm [*Xia et al*., 2012b]. Mosaic has three soil layers with thicknesses of 10, 30, and 160 cm, while SAC has five soil layers to cover a 2 m soil profile [*Xia et al*., 2012b].

To downscale the GRACE data, we first aggregated four sets of hourly LSM data separately to produce four sets of daily equivalent water thickness (EWT: mm) data on a 0.125° grid. The EWT is the integral of water above and inside the soil column within each grid cell, including surface water and soil water computed in the four LSMs. However, like many other LSMs, the four NLDAS-2 LSMs do not simulate groundwater fluxes [*Xia et al*., 2014; *Xia et al*., 2012a, 2012b]; thus, the models do not account for changes in groundwater fluxes such as water depletion and recharge. However, these changes can be captured by the GRACE data over large spatial extent. We then normalized the daily EWT by its mean value from January 2004 to December 2009 grid cell by grid cell to produce normalized EWT (S_i) by following the same normalization procedure used in the GRACE data (http://grace.jpl.nasa.gov/data/gracemonthlymassgridsoverview/). Considering that the footprint of GRACE signals is ~200,000 km^2 (about 4° by 4°) [*Longuevergne et al*., 2010] and the GRACE data are believed to have large uncertainty for resolutions < its footprint [*Long et al*., 2014; *Longuevergne et al*., 2010], we aggregated the 1° GRACE data to 4° and then downscaled the 4° data to 0.125° using the following method. The 0.125° normalized EWTs from the LSMs were aggregated to 4.0° to match with the 4°GRACE grid using area as a weighting factor as:

$$S_M = \frac{\sum (S_i \times a_i)}{\sum a_i} = \frac{\sum (S_i \times a_i)}{A}, \tag{4}$$

where S_M (mm) is the 4.0° LSM normalized EWT; a_i (m^2) is the area of the 0.125° grid cell i; A (m^2) is the total area of the 4.0° grid cell containing the 0.125° grid cell i. The difference between the 4.0° LSM normalized EWT and 4.0° GRACE normalized EWT (S_G) represents the bias (B) of the modeled EWT if we treat the GRACE data as "truth":

$$B = S_M - S_G. \tag{5}$$

The total water volume offset ($B \times A$) between the model and GRACE data were further distributed to the 0.125° grid using water volume as weight:

$$b_i = \frac{B \times A \times \frac{S_{oi} \times a_i}{\sum (S_{oi} \times a_i)}}{a_i} = \frac{B \times A \times S_{oi}}{\sum (S_{oi} \times a_i)}, \tag{6}$$

where b_i is the bias of the 0.125° model EWT; and S_{0i} is the prenormalized 0.125° model EWT. Since the GRACE data are a monthly composite product and different number of daily measurements are used for different months to calculate monthly values, the bias b_i is treated as the bias in the middle of a month. Then linear interpolation is applied to produce daily bias values for each grid cell. Finally, once the bias b_i is subtracted from S_i, we can obtain the 0.125° bias-corrected daily EWT (S_i'):

$$S_i' = S_i - b_i. \tag{7}$$

This downscaling method preserves the accuracy of the GRACE data and provides that the summation of the 0.125° bias-corrected EWT over any 4° GRACE grid cell is equal to the original 4° GRACE value at the same grid cell. Moreover, this downscaling method produces a finer resolution, continuous daily EWT series. The monthly water storage change (ΔS_m) in month m at grid cell i is derived as the difference between bias-corrected daily EWT value on the last day of a given month and on the last day of its previous month as:

$$\Delta S_m = S_i'(d_m) - S_i'(d_{m-1}), \tag{8}$$

where d_m and d_{m-1} are the Julian days of months m and $m-1$, respectively.

Since we downscaled the GRACE data using the outputs from four LSMs, we correspondingly produced four sets of 0.125° ΔS_m and monthly actual ET. The four sets of data form an ensemble. We used the ensemble mean as the final product. Hereafter, the reconstructed ET and downscaled ΔS_m denotes their ensemble means except as otherwise noted. To quantify the uncertainty in the reconstructed ET due to difference in the model outputs, we applied the commonly used ensemble standard deviation (SD), i.e., ensemble spread, as a metric:

$$SD=\sqrt{\frac{1}{M-1}\sum_{m=1}^{M}\left(ET_m-\overline{ET}\right)^2}, \tag{9}$$

where

$$\overline{ET}=\frac{1}{M}\sum_{m=1}^{M}ET_m, \tag{10}$$

and M (=4) is the number of ensemble members. Considering that there is only one set of precipitation and runoff data, the ensemble spread of ΔS is essentially the same as that of ET according to equation (1).

2.2.3. Evaluation of the Water Balance-Based ET

To evaluate the reconstructed ET values using the subbasin water balance approach, we compared the ET estimates with three data sets of ET estimations with reported good quality. One ET data set is produced by a remote sensing driven process-based algorithm [*Zhang et al*., 2010], the second data set is a data-driven, upscaled eddy-covariance flux measurements from the global FLUXNET work using a sophisticated machine learning method [*Jung et al*., 2010], and the third data set is the MOD16A2 global ET product [*Mu et al*., 2011]. All of the three ET data sets are widely assessed and used in the atmospheric and earth sciences community [*Cai et al*., 2011; *Wang and Alimohammadi*, 2012; *Zeng et al*., 2012], and are treated as benchmark ET products in some studies [*Schwalm et al*., 2013; *Zeng et al*., 2012].

Three statistical variables were used to measure the similarity between the three products, including mean difference (MD), root mean square difference (RMSD) and the coefficient of determination (R^2). The mean difference is defined as the average difference between the estimates to be evaluated (y_i) and the estimates to be compared against (x_i):

$$\mathrm{MD}=\frac{\sum_{i=1}^{n}(y_i-x_i)}{n} \tag{11}$$

where n is the sample size. RMSD measures the closeness between two ET products and is defined as:

$$\mathrm{RMSD}=\sqrt{\frac{\sum_{i=1}^{n}(y_i-x_i)^2}{n}} \tag{12}$$

The R^2 coefficient is used to evaluate the covariance between the two estimates of ET.

3. Results

3.1. Downscaled Equivalent Water Thickness and its Spatiotemporal Patterns

Figure 2 shows the normalized regional mean EWT values over the CONUS and its twelve hydrologic regions from April 2002 to September 2013 using the original monthly GRACE data, the original model daily EWT, and the downscaled daily EWT. Although there are some discrepancies between the GRACE data and the original ensemble mean of EWT from the NLDAS-2 LSMs, the mean of model results shows a generally good agreement with the GRACE data in terms of the seasonality and interannual variability (Figure 2). It is clear that the downscaled daily EWT matches the original GRACE data quite well with the added benefit of improved resolution using the model-based downscaling technique (Figure 2). The EWT series shows a clear, consistent seasonality with peak values falling between February and April when snow storage reaches maximum values and with minimum values around September when air temperatures are high accompanied by low seasonal precipitation in most hydrological units of the CONUS (Figure 2). It also shows large interannual variability; the difference between the highest water storage and the lowest water storage

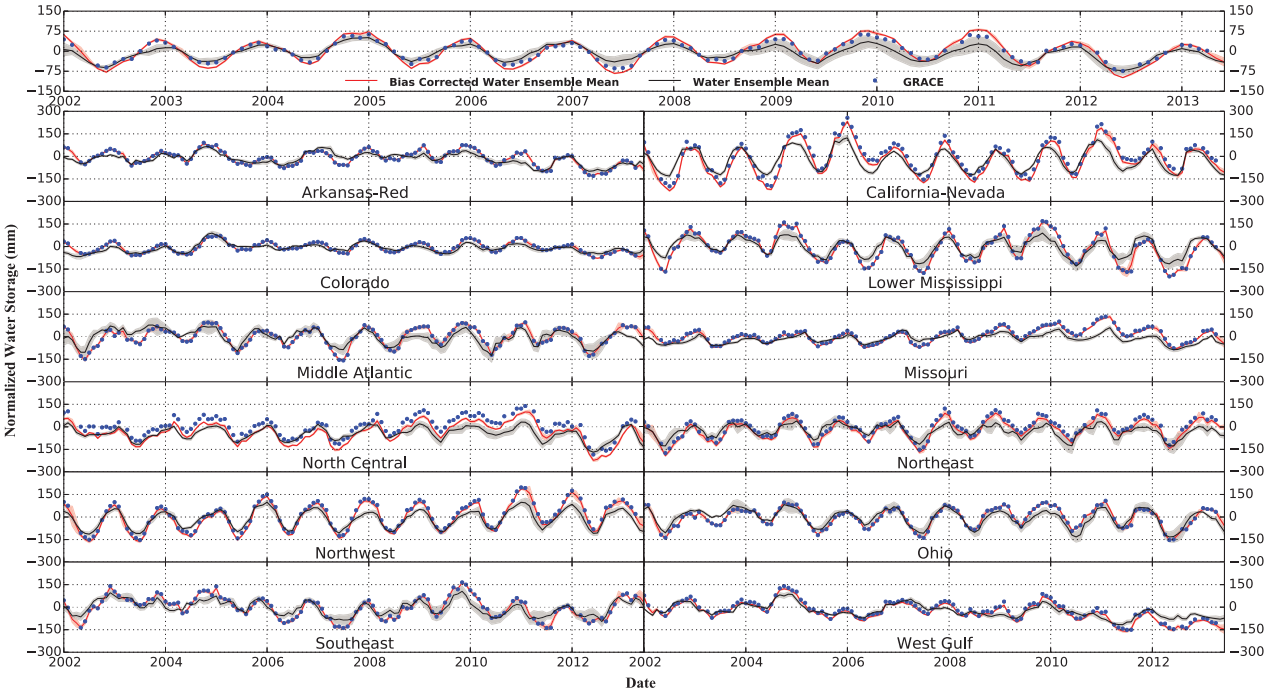

Figure 2. Time series of monthly terrestrial water storage change over CONUS and its twelve hydrologic regions from the original and land surface model-based downscaled GRACE data from 2002 to 2013; the downscaled data are the ensemble mean, while the grey area denotes the ensemble spread.

during the 12 years is about 180 mm, which is equivalent to 1,055 km^3 of liquid water. The min-max spreads of the original model water storage and the downscaled GRACE data (grey areas in Figure 2) are generally narrow with relatively large spreads in few months in a couple of hydrological regions, e.g., the Northwest and Southeast regions (Figure 2), indicating that the difference between the NLDAS-2 LSMs water storage data are subtle in these large regions.

The spatial pattern of the 11 year (April 2002 to March 2013) mean water storage change shows that most of the CONUS had very small water storage changes (Figure 3a), indicating most of these areas are in a water storage balanced state. However, some areas in the southern CONUS (e.g., eastern Texas and western Louisiana) and central Minnesota show negative multiyear water storage change, implying that these areas have lost water in the past 12 years. The loss of water storage in these areas is largely attributed to groundwater depletion and recent drought episode [*Freshwater Society*, 2013; *Long et al*., 2013]. In contrast, part of Florida shows a small gain of water storage during the past 11 years (Figure 3a).

3.2. Spatial Patterns of Water Budget Terms in CONUS

Spatial patterns of ground-observed, 11 year mean annual precipitation and annual runoff depth are shown in Figure 3. The mean annual precipitation displays a clear spatial gradient in which annual precipitation gradually decreases from the Southeast US to the Midwest and to the Rocky Mountains, and then increases from the Rocky Mountains to West Coast (Figure 3b). The spatial pattern of runoff depth is very similar to that of precipitation with a correlation coefficient of 0.84 ($P < 0.001$); the west and east coasts of the US and the Southeast have the highest annual runoff, while the Rocky Mountains and the Great Plains have the lowest annual runoff (Figure 3b). The similarity between the spatial patterns of precipitation and runoff indicates that precipitation is the major controlling factor of runoff.

3.3. Evaluation of Water Balance-Based ET Reconstruction and its Spatial Pattern

Multiyear average annual ET from the ensemble mean of water balance-based reconstructions (ET_{Recon}; Figure 4a) is compared with the remote sensing-based estimate [*Zhang et al*., 2010] (ET_{Zhang}; Figure 4b), the data-driven upscaled estimate [*Jung et al*., 2010] (ET_{Jung}; Figure 4c), and the MOD16 ET product (ET_{Mu}; Figure

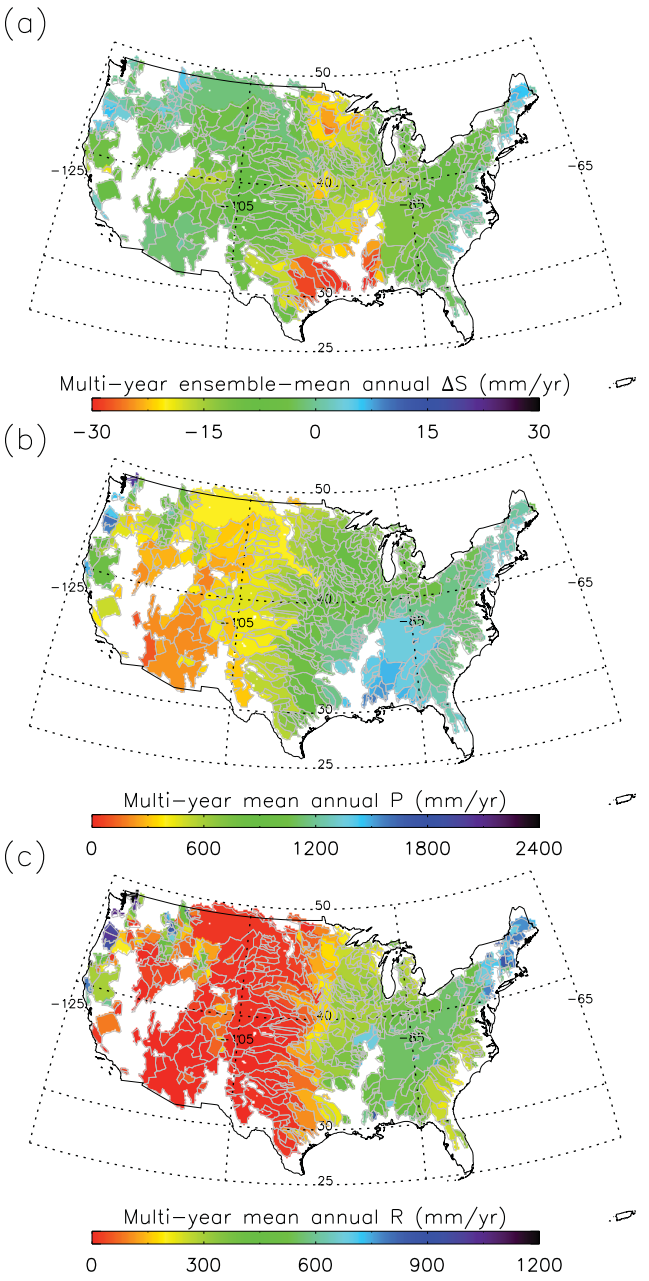

Figure 3. Spatial patterns of ground and satellite observed multiyear (from April 2002 to March 2013) mean annual (a) ensemble-mean terrestrial water storage change (ΔS), (b) precipitation (P), and (c) runoff depth (R).

4d). ET estimates from all four methods show similar spatial patterns. ET is the highest in the Southeast and decreases westward and northward, and reaches its minimum in the interior of the Intermountain West such as the deserts in Nevada. ET increases again from the Intermountain West to the West coast (Figure 4). Although ET_{Recon} has a similar pattern as those of precipitation and runoff, the correlation coefficient of ET_{Recon} and precipitation is 0.72 ($P < 0.001$); i.e., weaker than that of runoff and precipitation. This is because ET is not only largely controlled by precipitation but also impacted by other factors such as land-cover type, radiation, humidity, wind speed, temperature, etc.

The uncertainty in the reconstructed ET resulted from the difference in the four LSMs outputs is generally small (Figure 5): the mean ensemble spread of the reconstructed ET is less than 9 mm/month for 79% of the study region, and the largest ensemble spread is less than 30 mm/month. The regions with relatively large ET ensemble spread are mainly located in the coastal areas and part of the Midwest, while the other regions have generally small uncertainty spread, indicating that the four LSMs have generally compatible spatial patterns of water storage (Figure 5).

The four sets of ET estimates across the 592 CONUS subbasins show very similar spatial gradients (Figure 4), although some differences can be noticed. For example, the ET_{Recon} in this study generally has higher values than the other three products in the Southwest (Figure 4). The intercomparison between the four ET estimates show high correlations indicated by the high R^2 values ($\geq$0.74). The mean difference between these ET estimates for the 592 basins ranges from 6.8 to 96.5 mm yr^{-1} (Figure 6). The RMSD between the four ET estimates varies between 64.4 and 146.3 mm yr^{-1} (Figure 6). It is notable that the ET_{Recon} values show higher similarity and correlation with ET_{Zhang} and ET_{Jung} relative to ET_{Mu} (Figures 6a–6c). In addition, the ET_{Zhang} and ET_{Jung} values are very close to each other and show similar quality. Although the two prior estimates were produced by different approaches, they used similar climatologies and remote-sensing data [*Jung et al.*, 2010; *Zhang et al.*, 2010]. This may explain why these two products have very similar results across the CONUS. The generally close spatial patterns and small differences between the four ET estimates from different approaches indicate the high accuracy and robustness of these ET estimates. In other words, the water balance-based ET reconstruction conducted in this study is valid.

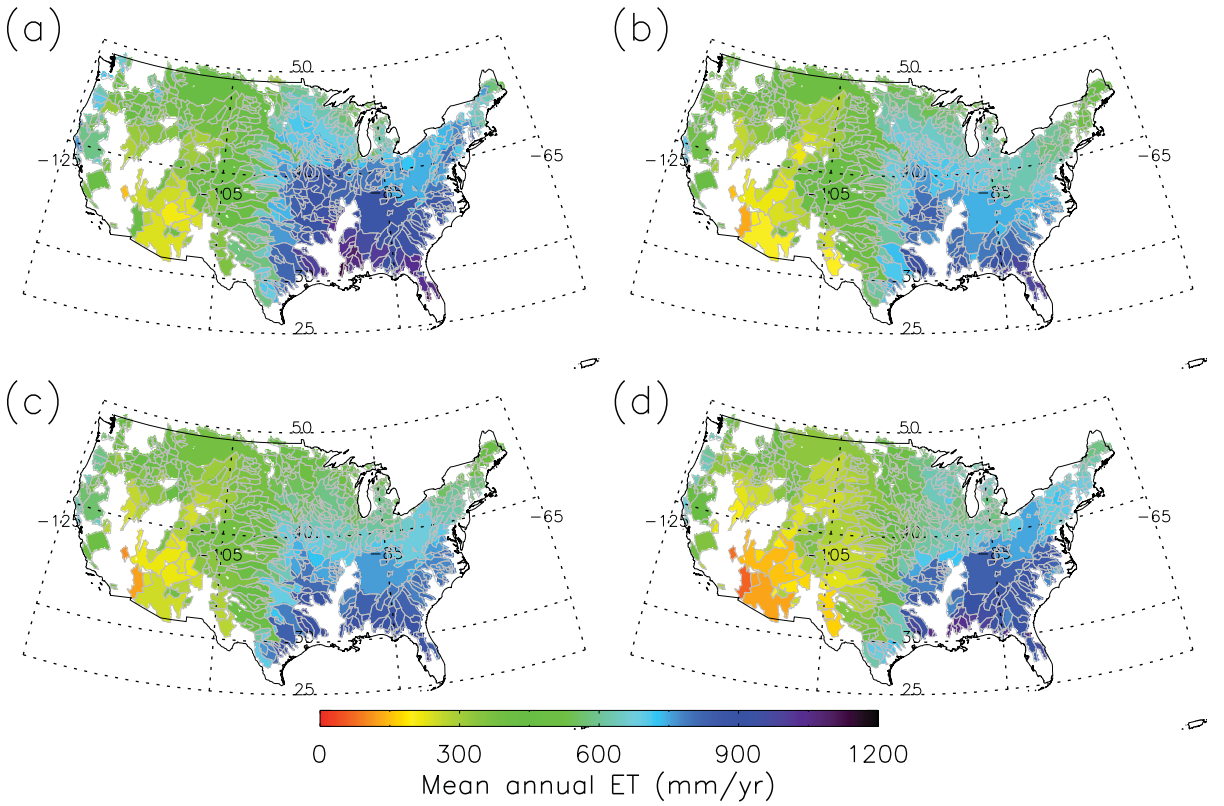

Figure 4. Spatial patterns of multiyear average annual ET from (a) the ensemble mean of water balance-based reconstructions, (b) a remote sensing-based estimate [*Zhang et al.*, 2010], (c) the data-driven upscaled estimate [*Jung et al.*, 2010], and (d) the MOD16A2 product [*Mu et al.*, 2011].

To assess the effectiveness of our downscaling method and the importance of monthly terrestrial water storage change, i.e., the ΔS term, in the water balance-based ET estimate, we produced two additional sets of monthly ET records: one is the water balance-based ET reconstruction by resampling the 1° GRACE data onto the 0.125° grid using the nearest neighbor method ($ET_{Resample}$), while the other is the ET estimate as the difference between P and R (ET_{P-R}). It is clear that the $ET_{Resample}$ shows substantially poorer agreements with the three independent ET records than the ET_{Recon} in terms of the scatterplots and the R^2 and RMSD metrics (Figures 7a–7c). This suggests that using the 1° GRACE data without downscaling it to derive subbasin level ET, in particular for regions that are less than 1° by 1°, will result in additional uncertainty and erroneously abnormal results as shown in Figures 7a–7c. In other words, our downscaling method has effectively disaggregated the coarser GRACE data to finer (0.125°) resolution, resulting in good-quality ET reconstruction. ET_{P-R} also shows degraded agreements with the three ET records similar to $ET_{Resample}$ in terms of the R^2 and RMSD metrics (Figures 7d–7f). Like the results of $ET_{Resample}$, the derived ET by ignoring the ΔS term can also result in erroneous and abnormal values such as negative values and erroneously high values as shown in Figures 7d–7f. Therefore, it is important to account for the ΔS term in order to provide accurate monthly ET estimates using the water balance approach. The downscaling approach implemented in this study is capable of disaggregating the coarser GRACE EWT to finer resolution to achieve reasonably good

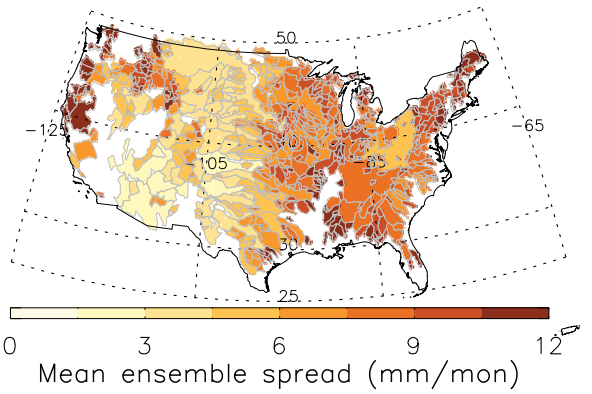

Figure 5. Mean ensemble spread of the reconstructed monthly ET.

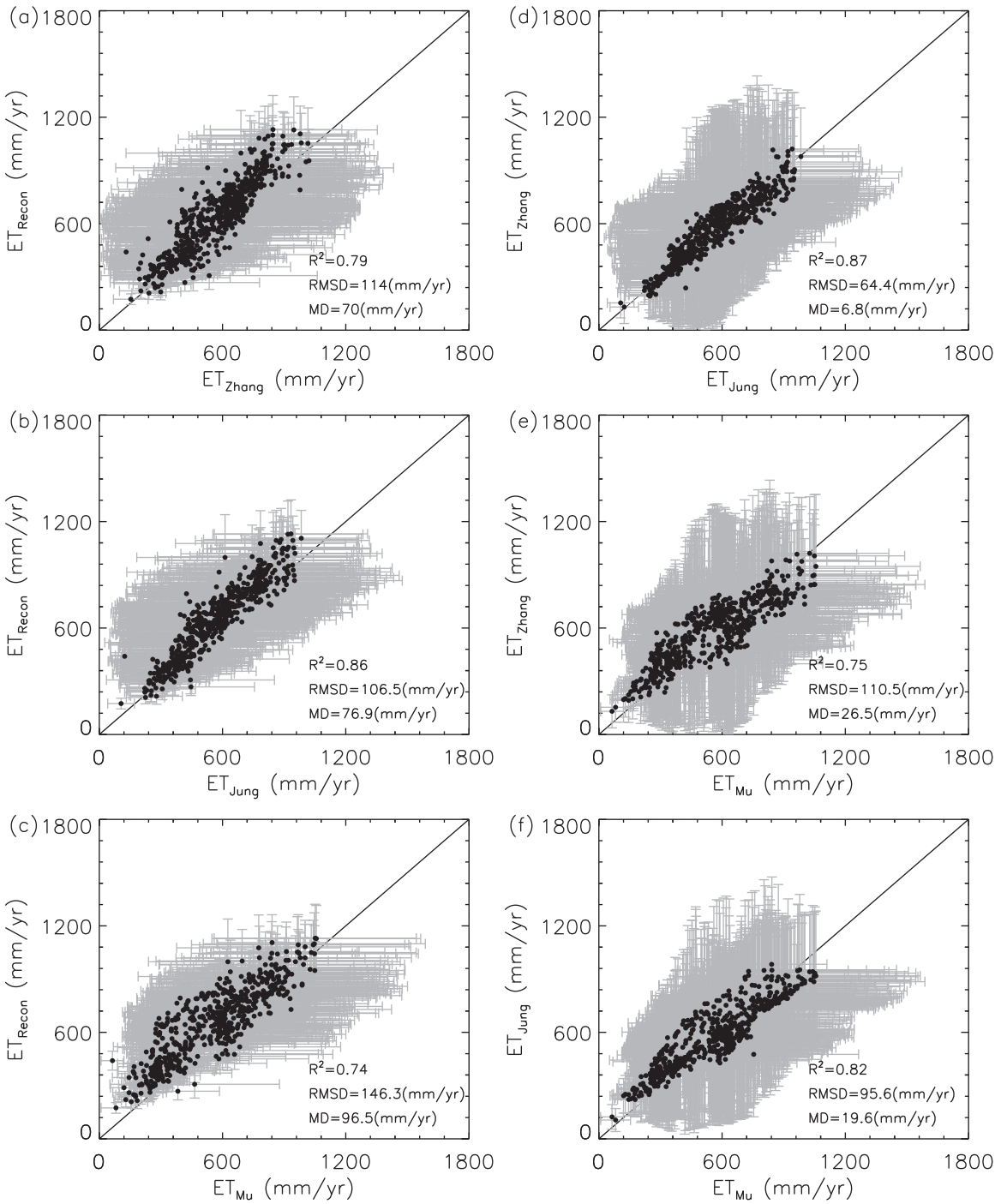

Figure 6. Intercomparisons (a) between mean annual ET estimates from the ensemble mean of water balance-based reconstruction (ET_{Recon}) and the remote sensing-based estimate by *Zhang et al*. [2010] (ET_{Zhang}), (b) between ET_{Recon} and the data-driven upscaled ET estimate by *Jung et al*. [2010] (ET_{Jung}), (c) between ET_{Recon} and the MOD16A2 ET by *Mu et al*. [2011] (ET_{Mu}), (d) between ET_{Zhang} and ET_{Jung}, (e) between ET_{Zhang} and ET_{Mu}, and (f) between ET_{Jung} and ET_{Mu} across 592 CONUS basins; black solid circles are basin-level mean annual ET, while grey error bars denotes interannual variability (standard deviation) of basin-level annual ET.

estimates of ΔS for subbasins that are even smaller than the footprint of the GRACE data. It is worthy to note that ET_{Recon}, $ET_{Resample}$, and $ET_{P\text{-}R}$ all have generally higher values than the three independent remote sensing-based ET products, suggesting that these remote sensing-based ET products may tend to understate the actual ET considering these products do not explicitly account for water balance closure and the effect of P on ET.

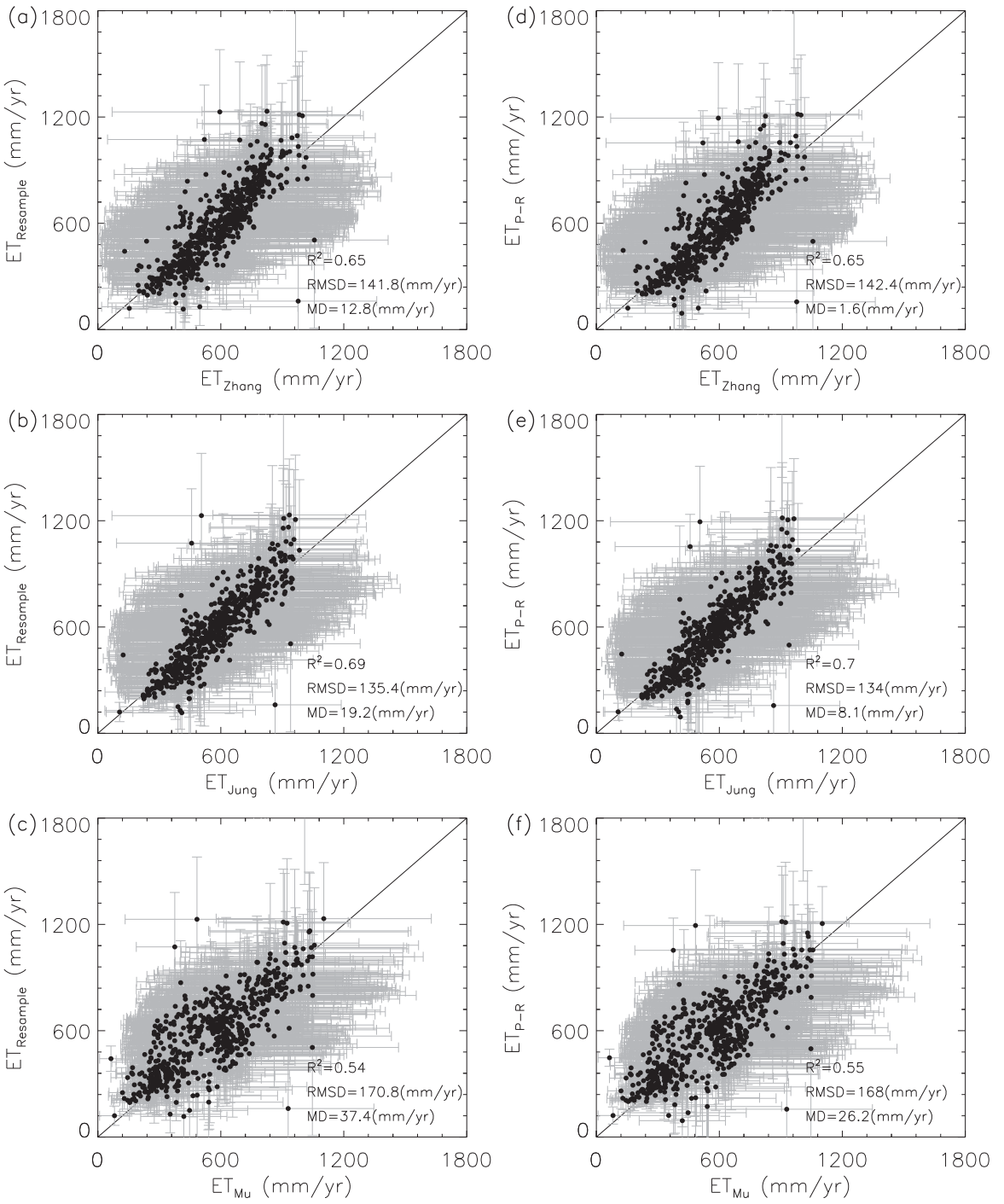

Figure 7. Same as Figure 6, but for (a–c) intercomparison between the water balance-based ET reconstruction by resampling the 1° GRACE data onto the 0.125° grid ($ET_{Resample}$) and the three independent ET records, and (d–f) intercomparison between the ET reconstruction by ignoring change in water storage (ET_{P-R}) and the three ET records.

We further investigated the agreement of seasonality among the four ET estimates, the ET_{Recon} and the three independent ET records, by comparing their 12 year mean monthly profiles. The results show that all the four ET estimates show similar monthly profiles with peak values in July when solar radiation, temperature and plant growth reach their peaks, and with minima in January when solar radiation and temperature reach their minima and most plants are dormant in the CONUS (Figure 8). Despite these similar monthly profiles, there are some noticeable differences. For example, the ET_{Recon} has generally higher values than the other three products, especially in the summer months. These differences imply that the existing three

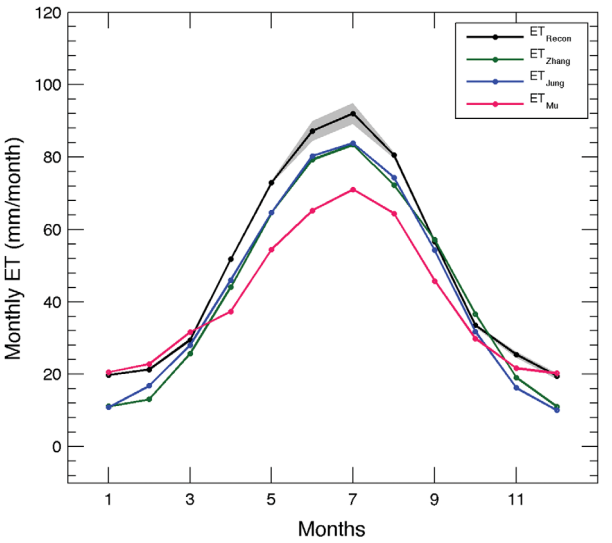

Figure 8. Comparison of mean monthly profile of actual ET from the ensemble mean of water balance-based reconstructions, remote sensing-based estimate [*Zhang et al*., 2010], data-driven upscaled estimate [*Jung et al*., 2010] and MOD16A product [*Mu et al*., 2011].

ET products may tend to underestimate the actual ET, because the existing ET products do not explicitly quantify some hydrological processes such as sublimation and snowmelt that impacts the ET, and the existing ET products can be also affected by satellite signal saturation during the peak of growing season. It is also notable that ET_{Mu} tends to have lower seasonal variability than the other products indicated by its higher minimum values and smaller peak values. In the rest of the months, ET_{Recon}, ET_{Zhang} and ET_{Jung} products have similar values, while ET_{Mu} have generally lower values than the other products (Figure 8).

4. Conclusion and Discussion

In this study, a new actual ET product across the CONUS has been derived from high quality satellite and ground observations, including the PRISM precipitation data, USGS observed streamflow data, and GRACE water storage data that has been downscaled using land surface models. This data set covers 73% of the CONUS and is available from April 2002 to September 2013. To our knowledge, this is the first study that derives decadal, continuous monthly ET values across the CONUS from observations using the subbasin water balance method. The method is unique in that it is observationally driven so that ET is computed as the residual in the water balance equation. This differs from past methods and models that often estimate ET using approximate methods and then compute the storage term as the residual in the water balance. The wide availability and accuracy of the GRACE observations enabled us to adopt a new approach in the terms of the water balance equation. The new ET product derived in this study shows high similarity with two existing, high quality ET products across the CONUS, indicating the reliability of the approach. Since the new ET product is derived from observations, it can be regarded as a benchmark data set to evaluate the existing and new model-based ET products. Moreover, we downscaled the GRACE data with the aid of four LSMs to produce a continuous daily equivalent water thickness data set with a spatial resolution of 0.125° and converted the USGS observed streamflow data to runoff depth. All the above products can serve as important hydro-meteorological data sets for assessment of hydrological and climatological changes, and evaluation of terrestrial water and energy cycle dynamics across the CONUS. These products will be also valuable for studies and applications in drought assessment, water resources management, climate change evaluation, and so on.

Although this new ET product is derived from ground and satellite observations, there are several limitations with this approach and the product. Further study is needed to thoroughly address these limitations. First, the reconstructed ET from the water balance method is a basin-mean product and correspondingly has variable spatial resolutions depending on the area of each individual subbasin. For example, the area of the 592 basins in this study ranges from 292 to 303,700 km^2. To produce gridded data, physical, or statistical methods need be developed to disaggregate the areal-average ET to individual grid cells; the distributed hydrologic models and land surface models may be useful for this.

Second, the ET reconstruction method does not account for the impacts of water transfer in or out of the subbasins by human activities such as irrigation and interbasin water diversions; therefore, the ET estimates in these areas heavily impacted by these human activities may have higher uncertainty. We derived a map showing these subbasins which have at least 10% of area controlled or affected by reservoirs and other human activities such as urbanization, mining, agricultural changes, and channelization using the USGS streamflow qualification codes for peak streamflow (http://nwis.waterdata.usgs.gov/nwis). 245 of the 592

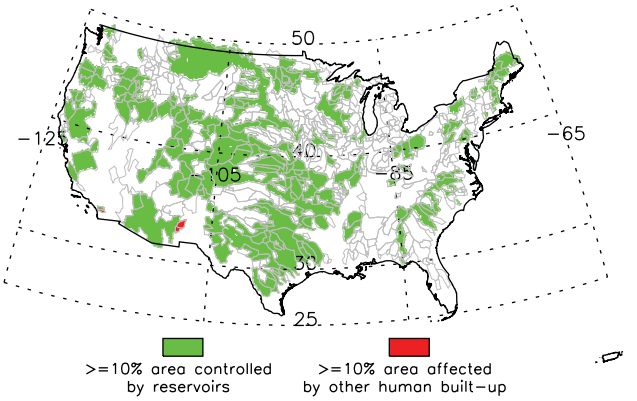

Figure 9. Locations of subbasins are impacted by reservoirs and other human activity such as urbanization, mining, agricultural changes, and channelization.

basins have more than 10% area controlled by reservoirs, while 3 basins have more than 10% impervious cover due to urbanization, mining, agricultural changes, channelization, or other anthropogenic activities (Figure 9). These basins impacted by human activities are largely located in the Midwest (Figure 9). The stream flow interruption caused by human activities do not impact the water balance-based ET reconstruction as long as no significant amount of water is diverted to another basin, because the ET in this study is derived on the basin level. However, the interbasin transfer of water definitely can cause large errors in the water balance-based ET calculation. It is impractical for us to quantify the impact of the interbasin transfer in this study due to lack of data. The general similar spatial patterns between ET derived in this study and the other three ET products from remote sensing and upscaled flux tower data in these basins impacted by human activities suggest that most of these basins do not experience substantial interbasin transfer of water (Figures 4 and 9).

Third, the ET estimate is directly calculated as the residual of all other water budget terms and inherits the measurement and processing errors in all other water budget terms. For example, some studies shows that the GRACE water thickness data can have an error of 2–3 cm [*Landerer and Swenson*, 2012]. Although the evaluation of the PRISM precipitation shows a near zero bias over the CONUS but may have relatively larger errors in some regions [*Daly et al*., 2008]. Finally, the availability of the ET reconstruction using this approach is limited by the availability of the measurements of the other water budget terms. However, the observation-based ET estimate in this study presents a best available ET estimate from the high quality observations. Therefore, there is strong reason to believe that this ET estimate is close to the "truth."

Acknowledgments

This research was funded by NOAA/Office of Oceanic and Atmospheric Research under NOAA-University of Oklahoma Cooperative Agreement NA14OAR4830100, U.S. Department of Commerce. Xianwu Xue was supported by a grant from DOI/USGS and South-Central Climate Science Center and Zhanming Wan was supported by the HyDROS Lab. The stream gauging data, observation-based precipitation data, and GRACE equivalent water thickness data are available at USGS National Water Data Center (http://waterdata.usgs.gov/nwis/rt), PRISM Climate Group (http://www.prism.oregonstate.edu), and JPL (http://grace.jpl.nasa.gov/data/), respectively. The rest data can be requested from Ke Zhang (kezhang@ou.edu). Zhanming Wan and Ke Zhang contributed equally to this work.

References

Allen, R. G., L. S. Pereira, D. Raes, and M. Smith (1998), Crop evapotranspiration: Guidelines for computing crop requirements, *FAO Irrig. Drain. Pap. 56*, Food and Agric. Organ. of the U.N., Rome.

Anderson, M. C., J. M. Norman, J. R. Mecikalski, J. A. Otkin, and W. P. Kustas (2007), 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.

Bastiaanssen, W. G. M., H. Pelgrum, J. Wang, Y. Ma, J. F. Moreno, G. J. Roerink, and T. van der Wal (1998), A remote sensing surface energy balance algorithm for land (SEBAL): 2. Validation, *J. Hydrol.*, *212–213*, 213–229.

Cai, X. M., Y. C. E. Yang, C. Ringler, J. S. Zhao, and L. Z. You (2011), Agricultural water productivity assessment for the Yellow River Basin, *Agric. Water Manage.*, *98*(8), 1297–1306.

Chambers, D. P. (2006), Evaluation of new GRACE time-variable gravity data over the ocean, *Geophys. Res. Lett.*, *33*, L17603, doi:10.1029/2006GL027296.

Chauhan, S., and R. K. Shrivastava (2009), Performance evaluation of reference evapotranspiration estimation using climate based methods and artificial neural networks, *Water Resour. Manage.*, *23*(5), 825–837.

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.

Cleugh, H. A., R. Leuning, Q. Mu, and S. W. Running (2007), Regional evaporation estimates from flux tower and MODIS satellite data, *Remote Sens. Environ.*, *106*, 285–304.

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.

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*, 901–919.

Freshwater Society (2013), *Minnesota's Groundwater: Is Our Use Sustainable?*, 26 pp., Excelsior, Minn.

Gillies, R. R., T. N. Carlson, J. Cui, W. P. Kustas, and K. S. Humes (1997), A verification of the 'triangle' method for obtaining surface soil water content and energy fluxes from remote measurements of the Normalized Difference Vegetation Index (NDVI) and surface radiant temperature, *Int. J. Remote Sens.*, *18*(15), 3145–3166.

Jung, M., et al. (2010), Recent decline in the global land evapotranspiration trend due to limited moisture supply, *Nature*, *467*(7318), 951–954.
Khan, S., Y. Hong, B. Vieux, and W. Liu (2010), Development evaluation of an actual evapotranspiration estimation algorithm using satellite remote sensing meteorological observational network in Oklahoma, *Int. J. Remote Sens.*, *31*(14), 3799–3819.
Koren, V., J. Schaake, K. Mitchell, Q. Y. Duan, F. Chen, and J. M. Baker (1999), A parameterization of snowpack and frozen ground intended for NCEP weather and climate models, *J. Geophys. Res.*, *104*(D16), 19,569–19,585.
Koren, V., M. Smith, Z. Cui, and B. Cosgrove (2007), Physically-based modifications to the Sacramento Soil Moisture Account Model: Modeling the effects of frozen ground on the rainfall-runoff process, *NOAA Tech. Rep. NWS 52*, Off. of Hydrol. Dev., NOAA Natl. Weather Serv., Silver Spring, Md.
Koster, R., M. Suarez, and M. Heiser (2000), Variance and predictability of precipitation at seasonal-to-interannual timescales, *J. Hydrometeorol.*, *1*, 26–46.
Landerer, F. W., and S. C. Swenson (2012), Accuracy of scaled GRACE terrestrial water storage estimates, *Water Resour. Res.*, *48*, W04531, doi:10.1029/2011WR011453.
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.
Liang, X., E. F. Wood, and D. P. Lettenmaier (1996), Surface soil moisture parameterization of the VIC-2L model: Evaluation and modification, *Global Planet. Change*, *13*(1–4), 195–206.
Liu, W. J., Y. Hong, S. Khan, M. B. Huang, T. Grout, and P. Adhikari (2011), Evaluation of global daily reference ET using Oklahoma's Environmental Monitoring Network-MESONET, *Water Resour. Manage.*, *25*(6), 1601–1613.
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.
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*, 1131–1151, doi:10.1002/2013WR014581.
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.
Mallick, K., et al. (2014), A Surface Temperature Initiated Closure (STIC) for surface energy balance fluxes, *Remote Sens. Environ.*, *141*, 243–261.
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.
Monteith, J. L. (1965), Evaporation and environment. The state and movement of water in living organisms, in *Symposium of the Society of Experimental Biology*, edited by G. E. Fogg, pp. 205–234, Cambridge Univ. Press, Cambridge, U. K.
Mu, Q., 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, J. M. Glassy, and S. W. Running (2003), Development of an evapotranspiration index from aqua/MODIS for monitoring surface moisture status, *IEEE Trans. Geosci. Remote Sens.*, *41*(2), 493–501.
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.
Schwalm, C. R., D. N. Huntinzger, A. M. Michalak, J. B. Fisher, J. S. Kimball, B. Mueller, K. Zhang, and Y. Q. Zhang (2013), Sensitivity of inferred climate model skill to evaluation decisions: A case study using CMIP5 evapotranspiration, *Environ. Res. Lett.*, *8*(2), 024028, doi:0.1088/1748-9326/8/2/024028.
Su, Z. (2002), The Surface Energy Balance System (SEBS) for estimation of turbulent heat fluxes, *Hydrol. Earth Syst. Sci.*, *6*(1), 85–99.
Swenson, S., and J. Wahr (2006), Post-processing removal of correlated errors in GRACE data, *Geophys. Res. Lett.*, *33*, L08402, doi:10.1029/2005GL025285.
Tang, Q. H., S. Peterson, R. H. Cuenca, Y. Hagimoto, and D. P. Lettenmaier (2009), Satellite-based near-real-time estimation of irrigated crop water consumption, *J. Geophys. Res.*, *114*, D05114, doi:10.1029/2008JD010854.
Tapley, B. D., S. Bettadpur, M. Watkins, and C. Reigber (2004a), The gravity recovery and climate experiment: Mission overview and early results, *Geophys. Res. Lett.*, *31*, L09607, doi:10.1029/2004GL019920.
Tapley, B. D., S. Bettadpur, J. C. Ries, P. F. Thompson, and M. M. Watkins (2004b), GRACE measurements of mass variability in the Earth system, *Science*, *305*(5683), 503–505.
Velpuri, N. M., G. B. Senay, R. K. Singh, S. Bohms, and J. P. Verdin (2013), A comprehensive evaluation of two MODIS evapotranspiration products over the conterminous United States: Using point and gridded FLUXNET and water balance ET, *Remote Sens. Environ.*, *139*, 35–49.
Wahr, J., M. Molenaar, and F. Bryan (1998), Time variability of the Earth's gravity field: Hydrological and oceanic effects and their possible detection using GRACE, *J. Geophys. Res.*, *103*(B12), 30,205–30,229.
Wahr, J., S. Swenson, and I. Velicogna (2006), Accuracy of GRACE mass estimates, *Geophys. Res. Lett.*, *33*, L06401, doi:10.1029/2005GL025305.
Wang, D. B., and N. Alimohammadi (2012), Responses of annual runoff, evaporation, and storage change to climate variability at the watershed scale, *Water Resour. Res.*, *48*, W05546, doi:10.1029/2011WR011444.
Wang, J., and R. L. Bras (2009), A model of surface heat fluxes based on the theory of maximum entropy production, *Water Resour. Res.*, *45*, W11422, doi:10.1029/2009WR007900.
Wang, J. F., and R. L. Bras (2011), A model of evapotranspiration based on the theory of maximum entropy production, *Water Resour. Res.*, *47*, W03521, doi:10.1029/2010WR009392.
Wu, H., J. S. Kimball, H. Y. Li, M. Y. Huang, L. R. Leung, and R. F. Adler (2012), A new global river network database for macroscale hydrologic modeling, *Water Resour. Res.*, *48*, W09701, doi:10.1029/2012WR012313.
Xia, Y., K. Mitchell, M. Ek, B. Cosgrove, J. Sheffield, L. Luo, C. Alonge, H. Wei, J. Meng, and B. Livneh (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., K. Mitchell, M. Ek, J. Sheffield, B. Cosgrove, E. Wood, L. Luo, C. Alonge, H. Wei, and J. Meng (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.

Xia, Y., M. T. Hobbins, Q. Mu, and M. B. Ek (2014), Evaluation of NLDAS-2 evapotranspiration against tower flux site observations, *Hydrol. Processes*, *29*(7), 1757–1771.

Zeng, Z. Z., S. L. Piao, X. Lin, G. D. Yin, S. S. Peng, P. Ciais, and R. B. Myneni (2012), Global evapotranspiration over the past three decades: Estimation based on the water balance equation combined with empirical models, *Environ. Res. Lett.*, *7*(1), 014026, doi:10.1088/1748-9326/7/1/014026.

Zhang, K., J. S. Kimball, Q. Mu, L. A. Jones, S. J. Goetz, and S. W. Running (2009), Satellite based analysis of northern ET trends and associated changes in the regional water balance from 1983 to 2005, *J. Hydrol.*, *379*(1-2), 92–110.

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.

Zhang, Y. Q., F. H. S. Chiew, L. Zhang, R. Leuning, and H. A. Cleugh (2008), Estimating catchment evaporation and runoff using MODIS leaf area index and the Penman-Monteith equation, *Water Resour. Res.*, *44*, W10420, doi:10.1029/2007WR006563.