Title
A methodology for evaluating evapotranspiration estimates at the watershed-scale using GRACE

Authors
Billah, Mirza M
Goodall, Jonathan L
Narayan, Ujjwal
et al.

Publication Date
2015-04-01

DOI
10.1016/j.jhydrol.2015.01.066

Peer reviewed

Journal of Hydrology 523 (2015) 574–586

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

A methodology for evaluating evapotranspiration estimates at the watershed-scale using GRACE

Mirza M. Billah [a], Jonathan L. Goodall [b,c,*], Ujjwal Narayan [d], J.T. Reager [e], Venkat Lakshmi [f], James S. Famiglietti [g,e]

[a] Bureau of Water, SC Department of Health and Environmental Control, Columbia, SC, USA
[b] Department of Civil and Environmental Engineering, University of Virginia, Charlottesville, VA, USA
[c] Department of Civil and Environmental Engineering, University of South Carolina, Columbia, SC, USA
[d] Con Edison, New York, NY, USA
[e] Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA
[f] Department of Earth and Ocean Sciences, University of South Carolina, Columbia, SC USA
[g] Department of Earth System Science, University of California, Irvine, CA, USA

ARTICLE INFO

Article history:
Received 28 August 2014
Received in revised form 20 January 2015
Accepted 31 January 2015
Available online 7 February 2015
This manuscript was handled by Konstantine P. Georgakakos, Editor-in-Chief, with the assistance of Matthew McCabe, Associate Editor

Keywords:
Evapotranspiration
Terrestrial water storage (TWS)
GRACE

SUMMARY

Accurate quantification of evapotranspiration (ET) at the watershed-scale remains an important research challenge. ET products from model simulations and remote sensing, even after incorporating in situ ET observations from flux towers in calibration or assimilation procedures, often produce different watershed areal-averaged ET estimates. These differences in ET estimates are magnified when they are integrated over time as part of water balance calculations. To address this challenge, we present a methodology for comparing watershed-average ET within a water balance framework that makes use of Gravity Recovery and Climate Experiment (GRACE)-observed terrestrial water storage change (TWSC). The methodology is demonstrated for South Carolina for a five-year period (2003–2007) using four different ET products: ET generated using a locally calibrated VIC model, a MODIS-derived ET product, and ET generated from two models (NOAH and VIC) as part of the North American Land Data Assimilation Systems 2 (NLDAS-2) project. The results of the example application suggest that the NLDAS-NOAH ET product is most consistent with GRACE-observed TWSC for the overall study region and time period. However, for periods of decreasing TWS, when ET becomes a more significant term in the water balance, the locally calibrated VIC model showed the most agreement with GRACE-observed TWSC. Application of the methodology for other regions and time periods can provide insight into different ET products when used for watershed-scale water resources management.

1. Introduction

Accurate quantification of evapotranspiration (ET) is essential for planning and design of water supply systems (Yeh et al., 1998), as well as for quantifying water availability and sustainable use of water resources (Loucks, 2000). Commonly used approaches for estimating ET rates at the watershed-scale include the use of land surface hydrologic models forced with weather data (e.g., Wood et al., 1992), fully-coupled land–atmosphere models in a reanalysis mode, or remote sensing (e.g., Mu et al., 2007; Rodell, 2004). These methods for ET estimation make use of in situ ET observations from flux towers within either calibration or data assimilation procedures (e.g., Pruitt and Angus, 1960; Running et al., 1999; Xu and Chen, 2005; Xu et al., 2006; Zhang et al., 2004). While these model and remote sensing-based approaches for estimating ET are widely used and well established, they often produce varying results when ET rates are integrated over watershed areas and over time. These differences in ET can be significant when performing watershed-scale water resource assessments and are difficult to evaluate with existing tools and methods.

GRACE provides monthly terrestrial water storage anomalies (TWSA) from observations of Earth's time-dependent gravity field (Wahr et al., 2004). TWSA, in the context of GRACE, is defined as a difference from the long-term mean of water stored in the terrestrial environment either as soil moisture, groundwater, snow, or surface water (Rodell and Famiglietti, 1999), among which soil moisture accounts for the largest portion in warmer regions (Rodell and Famiglietti, 2001). There has been significant work to

* Corresponding author at: PO Box 400742, Charlottesville, VA 22904, USA. Tel.: +1 (434) 243 5019.
E-mail address: goodall@virginia.edu (J.L. Goodall).

http://dx.doi.org/10.1016/j.jhydrol.2015.01.066
0022-1694/

extract hydrologic flux and state variables from GRACE data including estimating ET directly from GRACE as first shown by Rodell et al. (2004) and later by additional studies (e.g., Ramillien et al., 2006; Rodell et al., 2011; Syed et al., 2014). A new level 3 NASA GRACE RL05 release provides a 1×1 degree TWSA product that improves on leakage errors in prior GRACE TWSA products (Landerer and Swenson, 2012), making it applicable for regional-scale water resources assessments.

We build from this prior work of using GRACE to derive ET by instead using GRACE as a means for comparing different gridded ET products. The motivation for doing this is that remote sensing and modeling-based approaches for estimating watershed-scale ET provide more spatial resolution than GRACE-derived ET. However, at a regional-scale, GRACE TWSA provides a means for inter-comparing these gridded ET. In situ ET observations at flux towers have been the traditional approach for evaluating these regional scale ET gridded products. In fact, many of the gridded ET products make use of these in situ ET observations for data assimilation or calibration purposes. Despite this, and especially for regions with poor spatial coverage of in situ ET observations, there is a need for additional methods for evaluating gridded ET products used in water resources assessments.

The contribution of the work, therefore, is a methodology for comparing watershed-scale ET estimates using a water balance framework along with observed streamflow, precipitation, and Gravity Recovery and Climate Experiment (GRACE)-observed TWSA. The methodology consists of first identifying distinct periods of interest and computing TWS change (TWSC) for a collection of watersheds within the region for each period of interest. Then these TWSC values are estimated for each ET product being evaluated within an observations-based water balance framework. These ET-derived TWSC estimates are then compared to GRACE-derived TWSC estimates. Based on these comparisons, it is possible to gain insight into the ET products for the study area and analysis period.

We demonstrate the methodology for watersheds in South Carolina over the period 2003–2007 using four approaches for estimating watershed-scale ET. In the first approach, we use a locally calibrated Variable Infiltration Capacity (VIC) model (Abdulla and Lettenmaier, 1997a, 1997b) to estimate monthly ET rates. In the second approach we use an algorithm described by Mu et al. (2011) and that is based on the Penman–Monteith equation (Monteith, 1965) and Moderate Resolution Imaging Spectroradiometer (MODIS) remote sensing imagery to estimate monthly ET rates. The approach makes use of in situ ET observations for calibration (Mu et al., 2007) and has been shown to be an effective means for capturing regional-scale ET rates (Sheffield et al., 2009; Zhang et al., 2008). The third and fourth approaches use ET products from the National Land Dataset Assimilation System-2 (NLDAS-2) generated using the VIC and NOAH models (Mitchell, 2004; Xia et al., 2012a, 2012b). These NLDAS-2 ET datasets are produced using a data assimilation scheme and in-situ data including ET data from flux towers (Xia et al., 2012a).

While other methods for estimating watershed-scale ET exist, we have chosen to focus on these four methods in this study because we believe they are the most scale-appropriate and widely used ET data products for regional-scale hydrologic analysis. Other possible datasets include the North American Regional Reanalysis (NARR). However, NARR ET tends to over estimate ET rates, resulting in unrealistic decreasing trends in TWS when these ET estimates are used in water balance calculations (Billah and Goodall, 2011; Ruiz-Barradas and Nigam, 2006; Zeng et al., 2008). To correct for this, it is typical to apply a correction factor to NARR ET estimates to ensure no long-term trend in TWS. Furthermore, NARR estimates are at a coarse spatial scale (32.6 km), making the product unable to capture important regional-scale variability in ET rates important for regional water resources analysis. Therefore we have chosen to not use NARR ET as an approach for regional-scale estimation of ET rates in this study.

The remainder of the paper is organized as follows. The methodology is explained in Section 2 of the paper. A demonstration application of the methodology for watersheds in South Carolina is presented in Section 3. Finally, Section 4 provides a summary, discussion, and conclusion of the methodology and the South Carolina example application.

2. Methodology

2.1. Water balance framework

The basis of the methodology is to compare different ET products within the context of an observations-based water balance framework at the watershed-scale. Using each ET product, we estimate the incremental change in terrestrial water storage (TWSC) over some period of time for a watershed from the surface extending to a subsurface confining layer as

$$\mathrm{TWSC} = \overline{\{P_{obs}\}} - \overline{\{\mathrm{ET}_{est}\}} + \overline{\{R_{in}\}} - \overline{\{R_{out}\}} \quad (1)$$

where P_{obs} is observed precipitation, ET_{est} is estimated evapotranspiration, R_{in} is streamflow entering to the watershed, and R_{out} is streamflow exiting the watershed. The brackets signify a spatial integration over the watershed area and the bars signify a temporal integration over the time period. Eq. (1) assumes that the contribution of groundwater fluxes to TWSC is negligible, which we discuss in Section 4. Eq. (1) is applied for all watersheds within a study area for which the stream inflow (if not a headwater stream) and outflow were observed over the period of study.

2.2. Delineating watersheds for the water balance framework

Watersheds are delineated for a study region by first identifying streamflow stations with complete records over the period of analysis. Second, watersheds and sub-watersheds (the incremental drainage area between streamflow stations) are delineated using terrain processing or network tracing tools. In the US, it is possible to use the flow accumulation and flow direction grids from the National Hydrography Dataset (NHD) Plus program (McKay et al., 2012) to identify drainage areas between streamflow gauges. These grids have gone through a quality control process to ensure matching with the NHD vector-based hydrography data. For data management purposes, we recommend storing the watershed and time series data within a geodatabase with the data model described by Goodall and Maidment (2009) for later processing steps.

2.3. Water balance flux terms

Observed stream discharge data are obtained from observational networks including the National Water Information System in the US. Precipitation estimates are required on a monthly time step. In the US, the Parameter-elevation Regressions on Independent Slopes Model (PRISM) program is a widely used spatially interpolated precipitation dataset available on a monthly time step at a 4 km spatial resolution. Using the Zonal Statistics tool in ArcGIS, it is possible to estimate watershed areal-averaged precipitation rates for each watershed and for each month. These time series records are related to a specific watershed feature in the geodatabase.

One or more different ET products for the analysis are also generated as spatially interpolated datasets and aggregated to the

watershed-scale for each month following the same procedure as the precipitation dataset. Performing this calculation assumes that the ET products are available as gridded datasets or can be interpolated into gridded datasets. The end result of the data preparation step is a single geodatabase with watershed features and time series for streamflow, precipitation, and each of the different ET products on a monthly time step for each watershed feature.

2.4. Identification of analysis periods

Rather than simply integrating the monthly incremental TWSC estimated using the observational data over the entire period of analysis, we instead identify specific time periods within the record and calculate TWSC for each time period. We do this for two reasons: (i) to focus on specific periods in the record (e.g., wetting or drying periods) in order to understand similarities and differences in estimated TWSC using the ET products for these distinct periods; and (ii) to avoid the propagation of differences in ET where differences early in the time series influence latter portions of the time series.

For our demonstration application presented in Section 3, we first identify periods of approximately one-year each where GRACE TWSA begins and ends at average TWS levels during the GRACE period of record (i.e., where TWSA $\cong 0$). We also identify periods where GRACE TWSA begins at a local minimum (peak dry month) and extends to a GRACE TWSA local maximum (peak wet month). We call these periods of increasing TWS, or wetting periods. Finally we identify periods where GRACE TWSA begins at a local maximum (peak wet month) and extends to a GRACE TWSA local minimum (peak dry month). We call these periods of decreasing TWS, or drying periods. The periods used for the analysis are arbitrary and can be determined to address the needs of a specific application. For example, applications may benefit from setting analysis periods to seasons, water years, minimum to minimum GRACE-observed TWSA, or maximum to maximum GRACE-observed TWSA (see e.g., Castle et al., 2014, who analyzed groundwater depletion in the Colorado River basin during a 9-year drying period).

2.5. Calculation of ET-derived and GRACE-derived TWS change (TWSC)

For each of the periods identified for the analysis, TWS change (TWSC) is estimated using each ET product and the GRACE TWS anomaly (TWSA) product. ET-derived incremental TWSC estimates are first calculated on a monthly time step by applying Eq. (1). Then these monthly increment change estimates are summed for each period of analysis as

$$\mathrm{TWSC}_{\mathrm{ET}}(\Delta t) = \sum_{i=1}^{N} \mathrm{TWSC}(t_i) \tag{2}$$

where $\mathrm{TWSC}_{\mathrm{ET}}(\Delta t)$ is the ET-derived TWSC for the period Δt, $\mathrm{TWSC}(t_i)$ is the incremental TWSC for time t_i, and N is the total number of time increments in the period Δt. GRACE-derived TWSC estimates are taken as the difference between GRACE-observed TWS anomaly (TWSA) estimates at two different time periods as

$$\mathrm{TWSC}_{\mathrm{GRACE}}(\Delta t) = \mathrm{TWSA}(t_{i+\Delta t}) - \mathrm{TWSA}(t_i) \tag{3}$$

where $t_{i+\Delta t}$ and t_i represents the ending and starting times, respectively, in the time period Δt.

2.6. Analyzing ET-derived and GRACE-derived TWSC estimates

For each time period grouping (e.g., annual, wetting, and drying), we visualize and quantify the similarity between the ET-derived TWSC estimates and the GRACE-derived TWSC estimates. Because the study area is subdivided into watersheds each with observed stream in- and out-flow, it is possible to define groupings of watersheds in order to explore spatial-patterns in the ET estimates. For example, in the application for South Carolina presented in Section 3 we used four watershed groupings: (i) all watersheds, (ii) only watersheds draining to the streamflow station used for the local calibration of the VIC model used to estimate ET, (iii) only watersheds above the fall line, and (iv) only watersheds below the fall line. Finally, correlation coefficients between the different ET-derived TWSC estimates are also calculated to provide insight into their self-similarity for the different periods of analysis and watershed groupings.

3. Example application of methodology for South Carolina

As an example application of the methodology, we applied the approach for analyzing four ET products for watersheds in South Carolina. Because this application is part of a larger effort to study a period of drought in the region between 1998–2007, and because data from the Gravity Recovery and Climate Experiment (GRACE) that are available starting in 2003, the study period was set as 2003–2007 and we analyzed the data on a monthly time step.

3.1. Data preparation

3.1.1. Delineating watersheds

We identified 38 watersheds ranging from 30 to 7780 km^2 with observed inflow and outflow in our study region (Fig. 1). We did this by following the methodology described in Section 2.2 where we first found streamflow stations in the United States Geological Survey (USGS) National Water Information System (NWIS) with complete records over the study period (2003–2007). We then used geoprocessing tools within ArcGIS along with flow accumulation and flow direction grids from the National Hydrography Dataset Plus (NHD+) program (USEPA and USGS, 2005) to identify drainage areas between streamflow gauges.

3.1.2. Precipitation and streamflow

Estimates of precipitation and stream in- and out-flow were obtained for each basin (Fig. 2). Precipitation for each watershed was obtained from the Parameter-elevation Regressions on Independent Slopes Model (PRISM) dataset (Gibson et al., 2002). PRISM provides gridded precipitation estimates on a 4 km spatial resolution and on a monthly time step. By spatial averaging, we were able to obtain areal precipitation estimates for each watershed on a monthly time step. Monthly averaged streamflow estimates were obtained from the USGS National Water Information System (NWIS) for the stream gauges with complete records over the period of analysis.

3.1.3. VIC-derived ET

The first source for ET estimates used is from a locally calibrated Variable Capacity Infiltration (VIC) model. VIC is a semi-distributed macro-scale model that performs water and energy balances for a grid-based discretization of the landscape (Liang et al., 1994, 1996). ET in the VIC model is the combination of canopy layer evaporation (E_c), transpiration (E_t) from vegetation, and evaporation (E_l) from soil (Liang et al., 1994).

3.1.3.1. Model setup. We applied the VIC model to our study region using a 1/8° spatial grid. VIC was forced using meteorological data that includes precipitation, minimum and maximum temperature, and wind speed. Daily station observations of precipitation and minimum and maximum temperature obtained from the National Climatic Data Center (NCDC) were converted to a gridded input dataset of spatial resolution of 1/8° for the VIC model using the

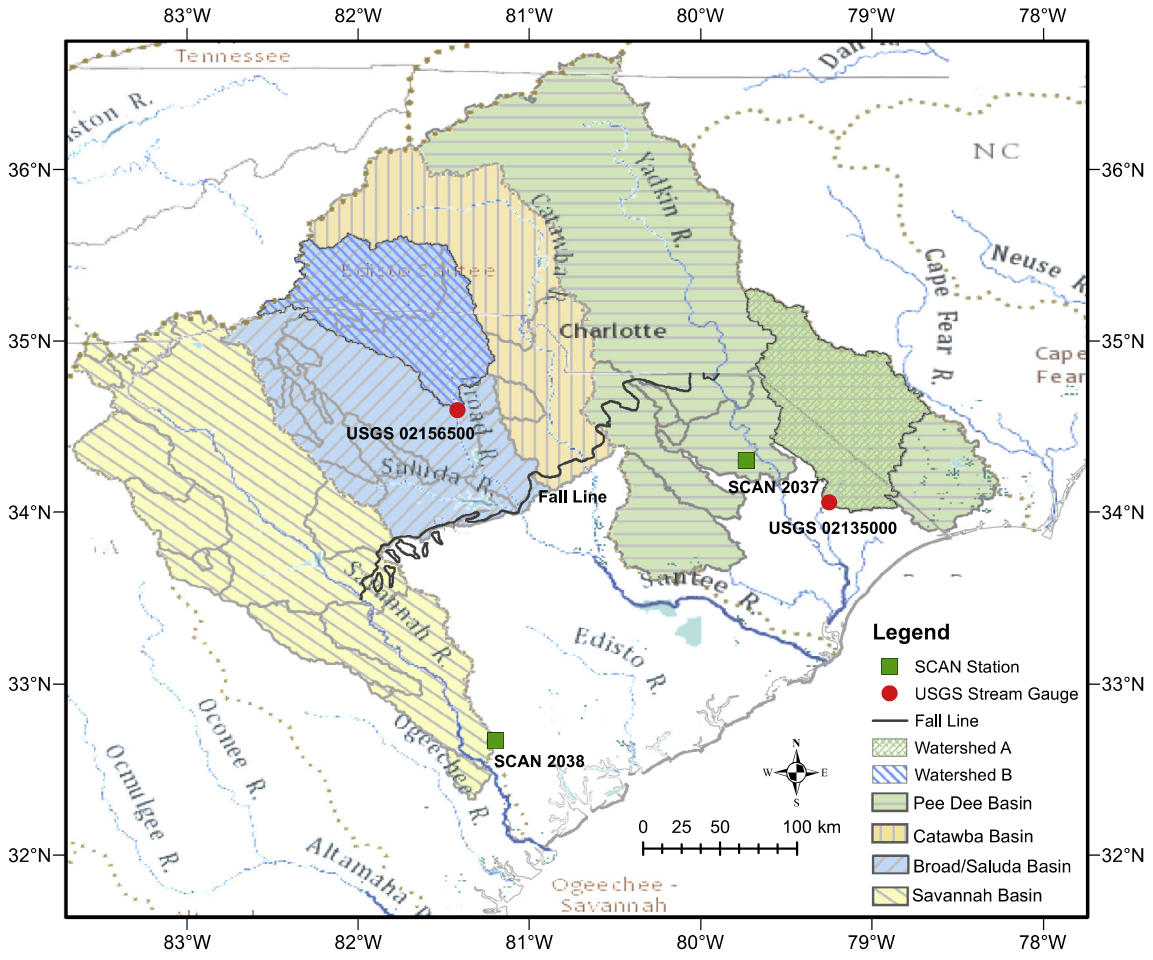

Fig. 1. The study area including the thirty-eight watersheds used in the analysis grouped into major river basin categories. Also shown are (i) the fall line, (ii) the two USGS streamflow gauging stations used for model calibration and evaluation along with the watersheds for these two gauging stations and (iii) the two SCAN soil moisture monitoring stations also used for model calibration and evaluation.

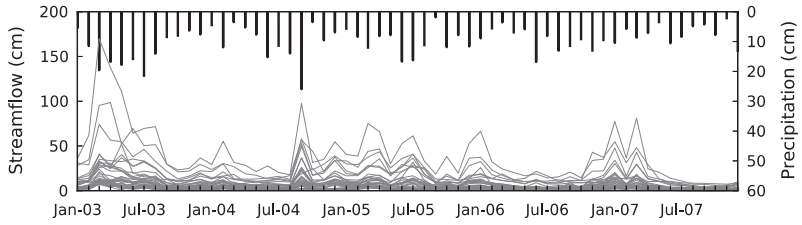

Fig. 2. Monthly average streamflow observations for each gage in the study area and areal averaged precipitation observations for the study area.

SYMAP interpolation algorithm (Shepard, 1984). Precipitation data were adjusted to match monthly means of the PRISM dataset to provide consistency in the water balance framework calculation that also uses PRISM data. The wind data were extracted from the National Centers for Environmental Prediction and National Center for Atmospheric Research (NCEP/NCAR) Reanalysis model and processed using linear interpolation to generate the gridded input dataset with spatial resolution of 1/8°.

We obtained soil and vegetation data from the Land Data Assimilation Systems (LDAS) (http://ldas.gsfc.nasa.gov/) at similar spatial resolution as precipitation, temperature, and wind data. The soil datasets obtained from Land Data Assimilation Systems (LDAS) were derived from 1-km Penn State STATSGO data that contains 11 individual layers to a depth of 2.5 m and 16 texture classifications ranging from sand to bedrock. Soil layer depths for the model were set to 0–10 cm for the top layer, 10–40 cm for the middle layer, and 40–100 cm for the deep layer. We selected the soil layer depths in part due to the model calibration and validation procedure described in the following paragraph. The vegetation parameter file was generated from LDAS vegetation data that contains 11 vegetation classes with spatial resolution of 1/8th degree. Among the vegetation classes, woodland (21%), wooded grassland (16%), mixed cover (17%), evergreen needleleaf forest (15%), and cropland (11%) covered most of the study area.

3.1.3.2. Model calibration. Because there are no known in situ ET observations within the study region and study period, we calibrated the VIC model using both soil moisture and streamflow observations and not ET observations. We began the calibration for a soil moisture station near Savannah, GA that is part of the Soil Climate Analysis Network (SCAN) (Fig. 1). The calibration objective function was to minimize the root mean square deviation (RMSD)

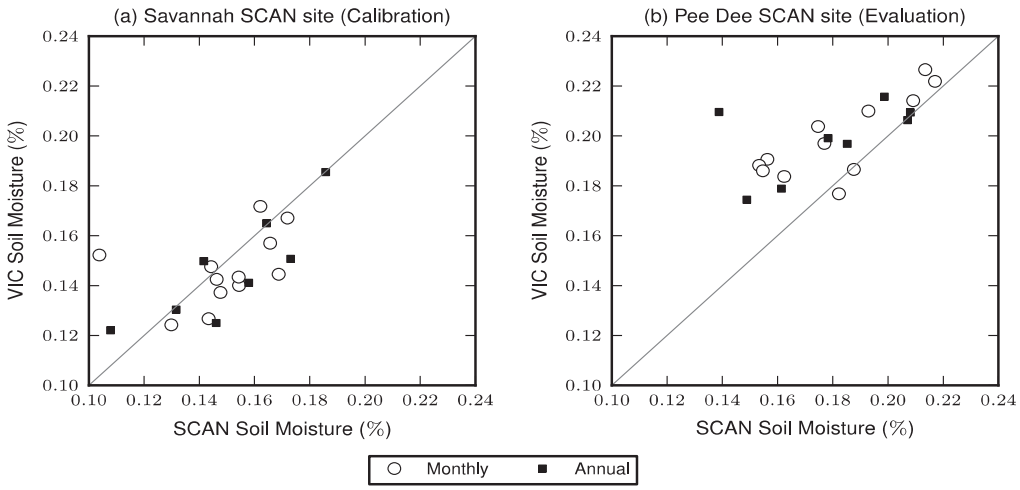

Fig. 3. Comparison of soil moisture estimated by the VIC model and soil moisture observed at the two Soil Climate Analysis Network (SCAN) monitoring stations shown in Fig. 1. The "Savannah SCAN site" was used for calibration and the "Pee Dee SCAN site" was used for evaluation.

between model-generated and SCAN-observed soil moisture for the area around the SCAN site. This was done for on both a monthly and annual aggregation (Fig. 3). We used VIC parameters typically used for model calibrate: variable infiltration curve (b), maximum base flow (D_{smax}), fraction of base flow where base flow occurs (D_s), fraction of maximum soil moisture content above which nonlinear base flow occurs (W_s), mid (d_2) and deep (d_3) soil layer depth, and minimum stomatal resistance (r_0) (Abdulla and Lettenmaier, 1997a, 1997b; Crow et al., 2003; Troy et al., 2008). The calibration was limited to a manual calibration due to computation demands of running the VIC model that prohibited more automated calibration techniques. The resulting RMSD was 0.022% for monthly means and 0.008% for annual means.

Starting with the parameter set obtained from the soil moisture calibration, we further calibrated the model using streamflow observations from the gage "Little Pee Dee River at Galivants Ferry" that is part of the USGS NWIS network (USGS 02135000; Watershed A in Fig. 1) for the period of 2003–2007. This stream gauge station has a drainage area of 7257 km^2 with sandy soil characteristics and was selected due to the fact that its one of the largest unmanaged basins within the study area. The calibration objective function was to maximize the Nash–Sutcliffe Efficiency (NSE) index between simulated streamflow using VIC routing scheme and observed streamflow for the two stations, which is the commonly used approach for evaluating VIC models. The NSE index estimated for the streamflow at this calibrated station was 0.67 and this index value was termed as good calibration according to published literature (Moriasi et al., 2007). Due to computational demands in running the VIC model, we were required to manually calibrate the model for a selected portion of the overall study region (e.g., only the model grid cells that drained to the monitoring station used for the calibration) in order to speedup model runtime. The parameter values used in the calibration, the range of values tested, and the final values of the parameters obtained from the calibration are provided in Table 1. A comparison of the monthly average streamflow for the calibrated VIC model and streamflow from USGS NWIS station are provided in Fig. 4.

3.1.3.3. Model evaluation. Following the calibration, we evaluated the model using soil moisture observed at a second SCAN site named the "Pee Dee SCAN site" (Fig. 1) and streamflow from a second USGS stream gauge location named "Broad River at Carlisle, SC" (USGS 02156500; Watershed B in Fig. 1). The stream gauge station has a drainage area of 7200 km^2 and was selected because it represents different geological conditions than the station used for calibration. The soil moisture estimates were compared to observed soil moisture values using a RMSD calculation and were found to be 0.027% for monthly means and 0.032% for annual means. The goodness-to-fit between observed and simulated results for monthly streamflow estimates were evaluated in terms of the Nash–Sutcliffe model efficiency coefficient and found to be a satisfactory model (0.59) for the period of 2003–2007 using the model classification scheme proposed by Moriasi et al. (2007).

3.1.4. MODIS-derived ET

The second source for ET estimates is from a remote sensing. Mu et al. (2011) describe an algorithm called MODIS16 for estimating ET based on MODIS satellite imagery that advances on the authors' earlier algorithm described in Mu et al. (2007). The original MODIS ET algorithm (Mu et al., 2007) is based on Penman–Monteith method and combines both meteorological observations and remote sensing data at a spatial resolution of 1 km. The meteorological observations include air pressure, minimum air temperature, humidity, and radiation. Remote sensing data

Table 1
Description of the parameters used in the VIC model calibration.

Parameter	Allowable range	Applied value	Units	Description
b	0.001–1.0	0.2	–	Variable infiltration curve parameter
D_{smax}	0.1–50.0	10	mm/d	Maximum base flow velocity
D_s	0.001–1.0	0.01	–	Fraction of D_{smax} where nonlinear base flow occurs
W_s	0.2–1.0	0.75	–	Fraction of maximum soil moisture above which nonlinear base flow occurs
d_2	0.1–3.0	0.3	m	Mid soil layer depth
d_3	0.1–3.0	0.6	m	Deep soil layer depth
r_0	~100	125–208	s/m	Stomatal resistance

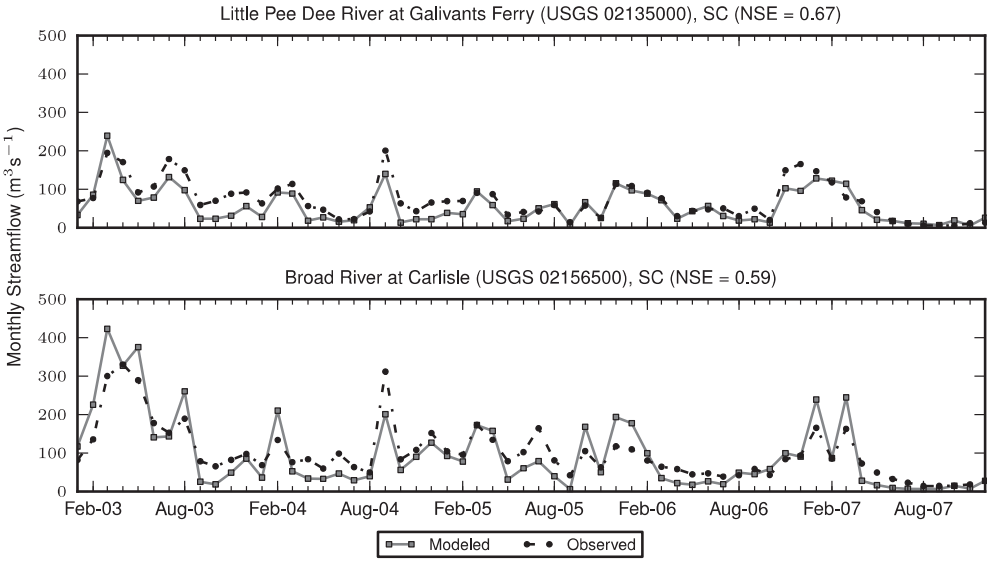

Fig. 4. Comparison of streamflow modeled using VIC coupled with a routing model and streamflow observed at the two USGS gauging stations shown in Fig. 1. The gauge "Little Pee Dee River at Galivants Ferry" was used for model calibration and the gauge "Broad River at Carlisle, SC" was used for model evaluation.

includes land cover, albedo, leaf area index (LAI), and Enhanced Vegetation Index (EVI). These data can be used to generate daily, monthly, or annual ET estimates. The MODIS ET algorithm was modified by Mu et al. (2011) to improve features of the ET estimates including estimates of nighttime ET, soil heat flux, and canopy transpiration. The improved algorithm was also able to estimate potential ET from saturated surfaces and actual ET from moist surfaces.

We used the MODIS16 product as an estimate of monthly ET at a 1 km spatial resolution for our study area. The monthly ET was imported using a geo-processing tool developed at the University of Montana and modified through this work to obtain all the monthly ET data for the study period in a single operation. The monthly mean ET in the extracted datasets were averaged for the watersheds in the study region to estimate rate of ET for the period of 2003–2007 using the approach described in Section 3.1.6.

3.1.5. NLDAS NOAH and VIC-derived ET

Two additional ET products were obtained from Phase 2 of the North American Land Assimilation Systems (NLDAS). NLDAS is a well known and widely used land surface modeling effort in the US that provides hydrologic flux estimates from 1979 to present on a 1/8° grid over North America generated using several different land surface models (LSMs) (Mitchell, 2004). For this study, we used monthly Evapotranspiration (total) product generated using the NLDAS-2 Noah model and VIC model. We selected these two models because Noah is widely used (e.g., Wei et al., 2013; Xia et al., 2013) and to compare the NLDAS-2 VIC output with our locally calibrated VIC output. We spatially averaged the output from the models to each of the 38 model watersheds as described in Section 3.1.6.

3.1.6. Watershed areal-averaged ET estimates

The gridded ET estimates were averaged for each watershed area for use in water balance framework. The spatial averaging can be described as

$$ET_i = \frac{1}{T}\int \{ET_{grid}\} dA_i \quad (4)$$

where ET_i is the incremental ET for watershed i [L] over the time period T, A_i is the area of watershed i [L^2], ET_{grid} is the gridded incremental ET [L] from one of the sources (i.e., VIC, MODIS, NLDAS-NOAH, and NLDAS-VIC) over the same time period T.

3.1.7. GRACE TWS anomaly (TWSA) product

We used the GRACE product to estimate TWSC for the entire study area (approximately 94,000 km^2) for the different analysis time periods. The gridded, column-integrated, monthly TWSA data was obtained from the NASA Jet Propulsion Laboratory (JPL) (ftp://podaac-ftp.jpl.nasa.gov/allData/tellus/L3/land mass/RL05) (Landerer and Swenson, 2012) in units of cm equivalent water thickness. These data must be user corrected with the supplied 1-deg scaling factors, derived by processing a model with similar truncation and filtering as the GRACE data and measuring the resulting signal decay. When applied, these scaling factors approximate a 1-deg gridded GRACE solution, and are accompanied by 1-deg GRACE error estimates that constrain the accuracy of the signal (Landerer and Swenson, 2012). This scaling is assumed to correct for regional leakage errors in the GRACE solutions by incorporating a priori information on the spatial pattern of the signal source.

The uncertainty in the GRACE TWSA estimates result mainly from GRACE measurement and leakage errors (for a more complete described of GRACE errors, please see Landerer and Swenson, 2012; Swenson and Wahr, 2002). The measurement error is caused by instrument limitations and is inversely related to spatial coverage (Rodell and Famiglietti, 2001). The leakage error is associated with truncation and filtering in the spectral domain. The total error is the combination of the two errors for each grid cell and measured by

$$error_{total,i} = \sqrt{(error_{measurement,i})^2 + (error_{leakage,i})^2} \quad (5)$$

where $error_{total,i}$ is the total error for grid cell i, $error_{measurement,i}$ is the measurement error at grid cell i, and $error_{leakage,i}$ is the error due to leakage at grid cell i (Landerer and Swenson, 2012). Because of spatial correlation in the error estimates, the regionally averaged error must be weighted when summed, resulting in smaller error estimates for larger areas (Landerer and Swenson, 2012). Using this approach we found the average error for the GRACE TWSA estimates in the study area to be approximately 2.3 cm for each month. This means that GRACE TWSC estimates could have twice this error, or 4.6 cm for each month, given that TWSC requires a difference calculation.

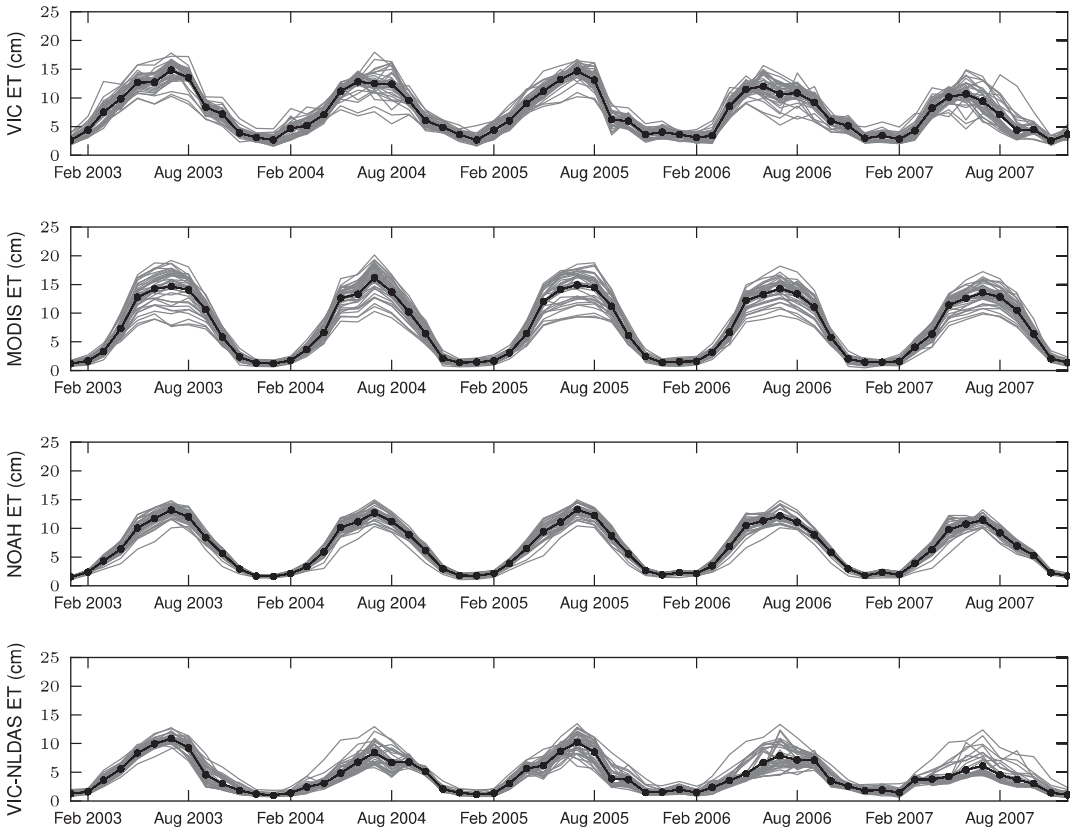

Fig. 5. Monthly average ET estimates from the four ET products. Each gray line represents areal-averaged ET for one of the 38 watersheds within the study area, and the black line on each subplot represents the mean across all 38 watersheds.

3.2. Results and discussion of the example application

3.2.1. Comparison of ET products

Fig. 5 shows the watershed-averaged ET rates for the four ET products over the period of analysis. The gray lines provide a measure of the spatial variability of ET rates across the 38 watersheds within the study region and the black line gives the average over all watersheds.

From Fig. 5, some of the differences in ET rates among the four products become clear. Looking at seasonal trends, MODIS-derived ET shows the largest summertime variability across the watersheds. NLDAS-VIC showed the lowest average summertime ET rates. The 2007 summer was drier than other summers in the period of analysis as evident by the streamflow observations during this period (Fig. 2). The impact of this dry period was most clearly evident in the NLDAS-VIC ET time series, although all ET products showed some impact through lower ET rates in the summer of 2007.

From the ET data shown in Fig. 5 alone, however, we have limited means for evaluating the different ET products. The motivation for developing this methodology was to provide a means to evaluate differences among ET products by using available observational data from other components of the water budget.

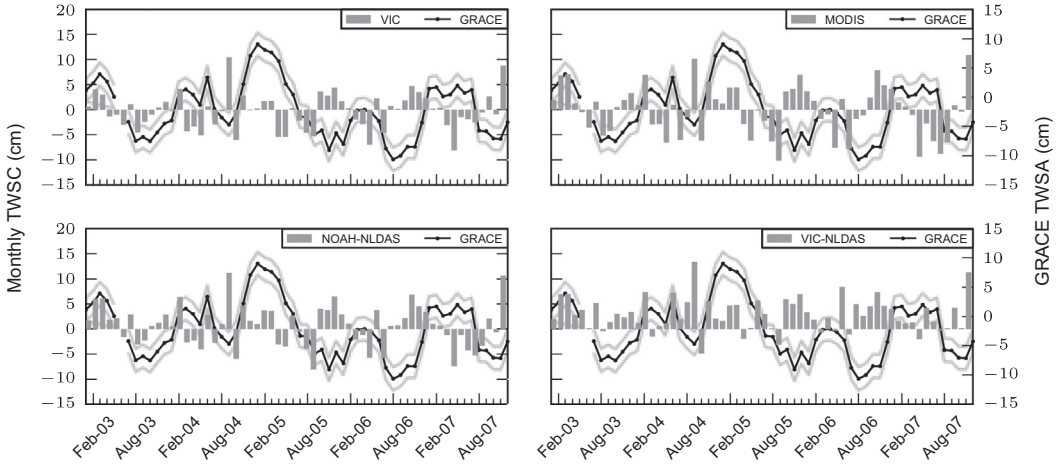

Fig. 6. The bar charts show monthly incremental TWSC estimated using each ET product within the water balance framework. The lines show the GRACE-observed TWSA over the same period.

Table 2
The time periods used for each scenario defined using the GRACE TWSA observations.

	2003	2004	2005	2006
(a) Annual periods: Periods between average GRACE TWSA				
Overall	07/03–07/04	07/04–07/05	07/05–04/06	04/06–07/07
Calibrated Watershed	05/03–05/04	05/04–10/05	10/05–12/06	12/06–12/07
Above Fall Line	07/03–07/04	07/04–07/05	07/05–04/06	04/06–07/07
Below Fall Line	07/03–07/04	07/04–07/05	07/05–04/06	04/06–08/07
(b) Wetting periods: Periods of increasing GRACE TWSA				
Overall	10/03–03/04	09/04–01/05	11/05–04/06	08/06–02/07
Calibrated Watershed	08/03–03/04	09/04–01/05	11/05–05/07	–
Above Fall Line	10/03–06/04	09/04–01/05	11/05–04/06	08/06–02/07
Below Fall Line	08/03–03/04	09/04–01/05	11/05–04/06	08/06–05/07

	2003	2004	2005	2006	2007
(c) Drying periods: Periods of decreasing GRACE TWSA					
Overall	03/03–10/03	06/04–09/04	01/05–11/05	04/06–8/06	02/07–11/07
Calibrated Watershed	03/03–08/03	03/04–09/04	01/05–11/05	05/07–11/07	–
Above Fall Line	03/03–10/03	06/04–09/04	01/05–11/05	04/06–08/06	02/07–10/07
Below Fall Line	03/03–08/03	03/04–09/04	01/05–11/05	04/06–08/06	05/07–11/07

3.2.2. Defining periods for TWSC analysis

Inserting each ET product within the observations-based water balance framework (Eq. (1)) results in estimated incremental TWSC for each watershed (Fig. 6). The bar charts in Fig. 6 show the monthly incremental TWSC accumulated across all watersheds within the study area. The line on each subplot in Fig. 6 shows the GRACE TWSA over the period of analysis.

Because TWSA is a measure of the difference between current TWS and a long-term mean TWS for the region, we would expect periods of increasing TWSA to correspond with periods of positive incremental TWSC estimated by the water balance framework. Likewise, periods of decreasing TWSA should correspond with periods of negative incremental TWSC estimated by the water balance framework. We would expect that the magnitude of TWSC during given periods from the GRACE TWSA, taken as the difference in TWSA between two points in time (Eq. (3)), to correspond to the accumulated incremental TWSC estimated by the water balance framework over that same period of time, taking the GRACE

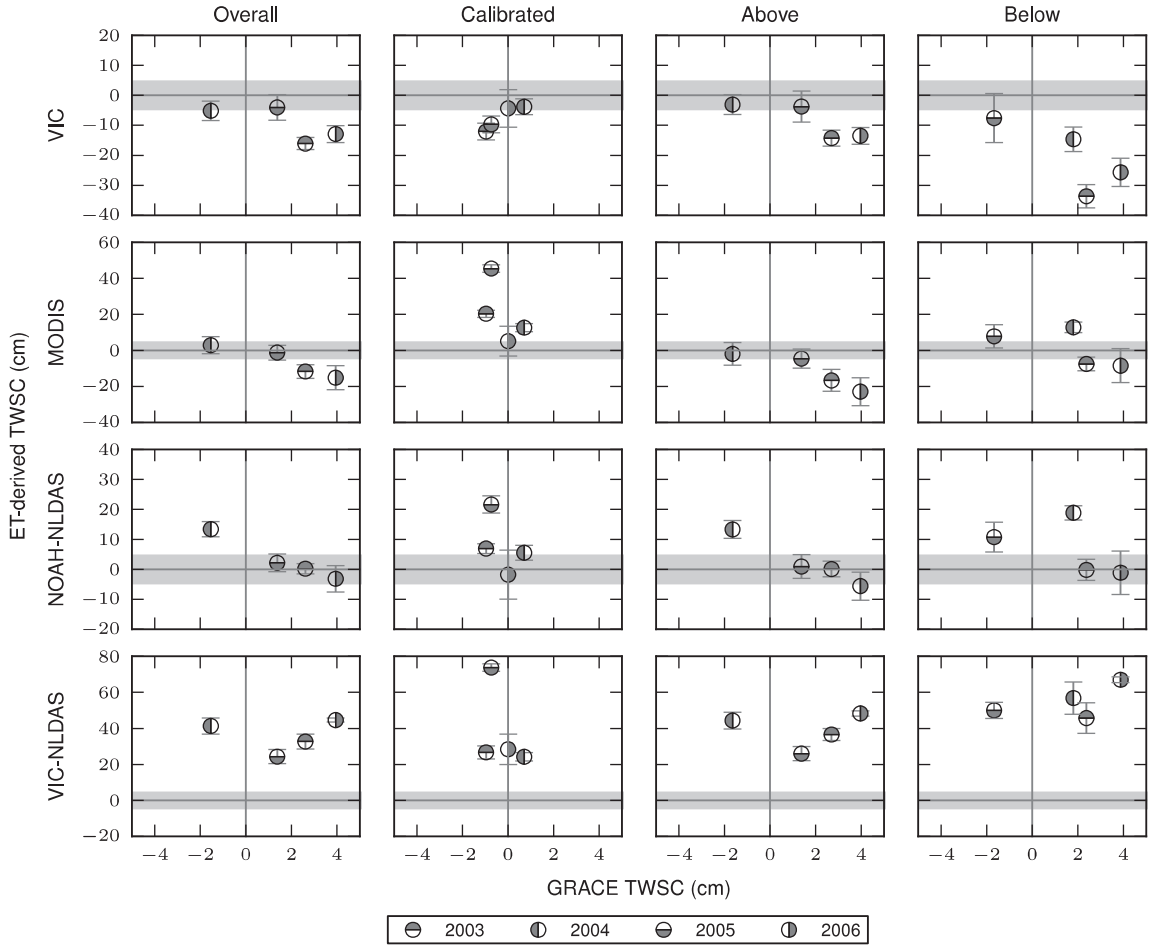

Fig. 7. GRACE-observed TWSC compared to ET-derived TWSC for periods between average TWS conditions. The shaded regions show where TWSC is ±5 cm (the approximate error of the GRACE TWSC measurement) using the ET products to aid in comparison to GRACE-observed TWSC.

TWSA uncertainties into account (shown in Fig. 6 as the region around the GRACE TWSA line plot).

Table 2 identifies different periods of analysis used to compare the results presented in Fig. 6. First, we identified four periods, approximately one-year each, where GRACE TWSA begins and ends at long-term average TWS levels (i.e., where TWSA ≅ 0 ± 5 cm). We did this first for all watersheds within the study region (the Overall row in Table 2a). We also grouped watersheds upstream of the streamflow station used to calibrate the locally calibrated VIC model (the Calibrated Watershed row in Table 2a), as well as watersheds that fall above or below the fall line (the Above Fall Line and Below Fall Line rows in Table 2a). We did this to better understand the similarities and differences between the ET products for these different regions within the study area.

Table 2a provides the start and end dates for each of these periods and the column headings give the start year for the period. This start year is used in later figures for referring to each time period. After identifying these annual periods, we also identified periods of increasing (Table 2b) and decreasing (Table 2c) TWSC from the GRACE TWSA record. For both the increasing TWS (or wetting) and decreasing TWS (or drying) periods, we again considered four different scenarios of watershed groupings within the study area.

3.2.3. ET-derived TWSC vs. GRACE-derived TWSC

We first compared the TWSC derived from each ET product to GRACE-observed TWSC for each time period and watershed-grouping scenario in Table 2. Fig. 7 shows the results of this analysis for the annual period analysis while Figs. 8 and 9 show the results of this analysis for the wetting and drying periods, respectively. The vertical error bars in the plots show the impact of including or excluding the start and end months in the ET-based TWSC time series as the variability in these estimates. In some cases, there was a large monthly incremental TWSC at the start or end of the period of analysis and, given possible lags between the ET-based TWSC estimates and the GRACE-based TWSC estimates, we wanted to capture the variability caused by the borderline incremental TWSC estimates in the analysis. The horizontal error bars show the uncertainty of the GRACE TWSC estimates for our study region discussed previously to be 4.6 cm.

The first column in Fig. 7 shows the results of the annual analysis for each ET product for all watersheds (the *overall* scenario) within the study region. The most obvious finding is that the VIC-NLDAS-derived estimates of TWSC are clearly higher than the TWSC observed by GRACE and the TWSC derived using the other ET products. This suggests that the VIC-NLDAS ET product is underestimating ET within the region and during the period of analysis. Also, Fig. 7 shows that the ET-derived TWSC estimates (with the exception of the VIC-NLDAS ET-derived TWSC estimate) are consistent for 2005, but differ for the other three years. Compared to the GRACE TWSC estimates, the NOAH ET-derived TWSC are within the 0 ± 5 cm range for three of the four years whereas the locally calibrated VIC and MODIS ET-derived TWSC are within the 0 ± 5 cm range for only two of the four years. Performing a root mean square deviation (RMSD) calculation for the data (Table 3) confirms that NLDAS-NOAH ET-derived TWSC matched most closely with GRACE TWSC (RMSD = 8.4) followed MODIS (RMSD = 12.2) and the locally calibrated VIC model (13.0). The NLDAS-VIC ET-derived TWSC is much further from the GRACE TWSC (RMSD = 35.1).

When looking at the different subgroups of watersheds in Fig. 7, we observe first that the locally calibrated VIC model, not surprisingly, performs best for the *calibrated watershed* scenario that

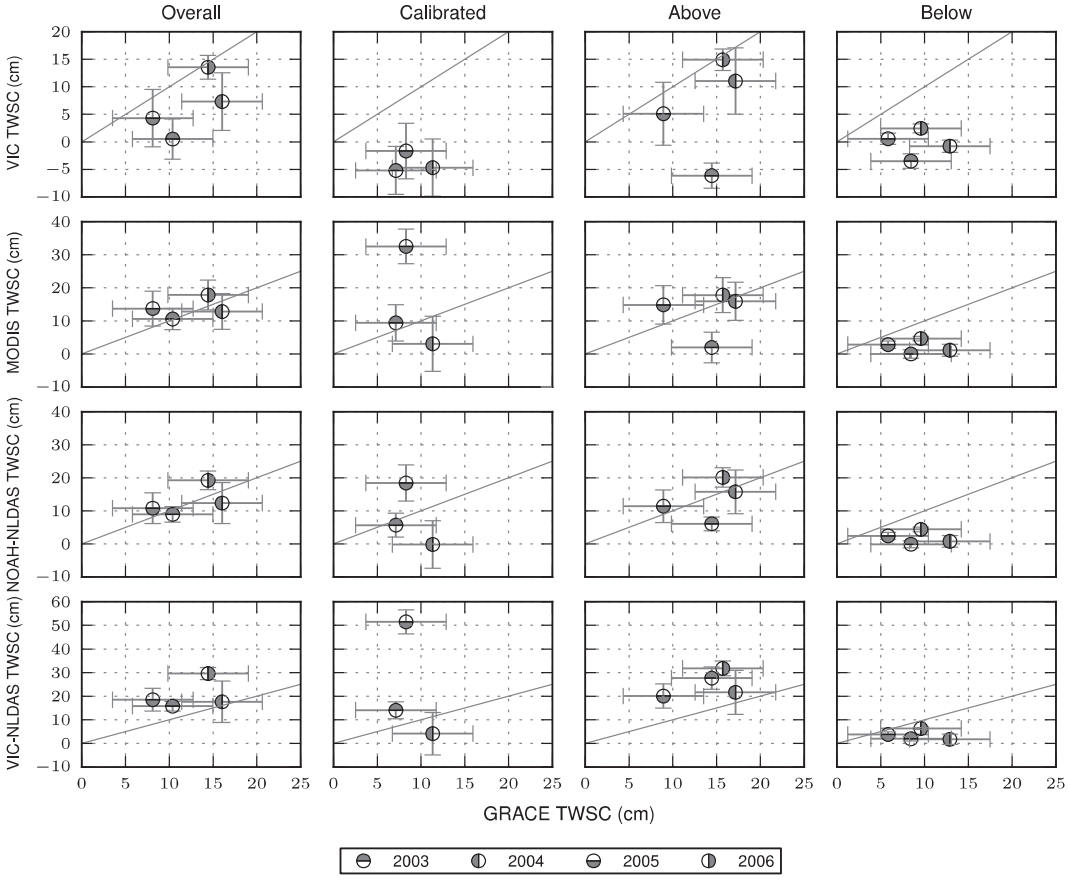

Fig. 8. Periods of increasing TWS observed by GRACE (wetting periods). 1:1 lines are included for interpreting cases of over and under prediction of TWSC using the different ET products.

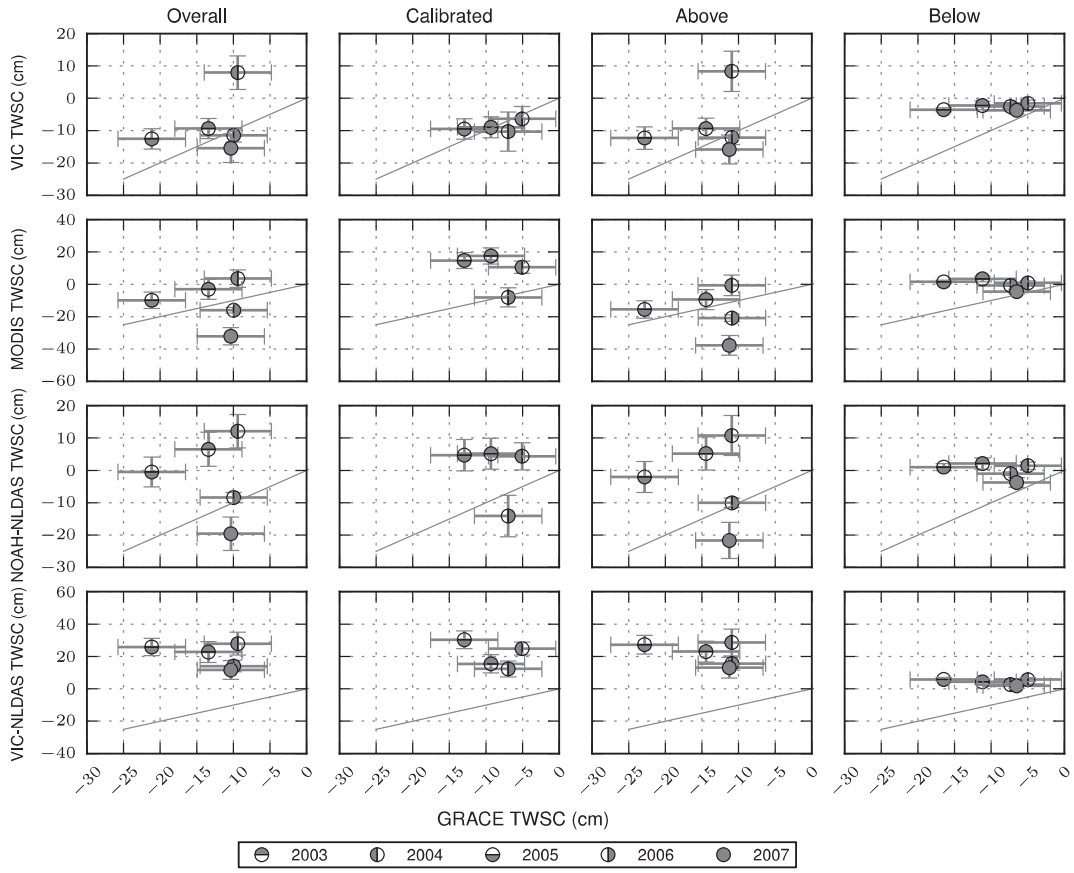

Fig. 9. Periods of decreasing TWS observed by GRACE (drying periods). 1:1 lines are included for interpreting cases of over and under prediction of TWSC using the different ET products.

includes only watersheds that drain to the streamflow gaging station used for the VIC model calibration. The fit according to a RMSD estimate is slightly better than the fit between NLDAS-NOAH and GRACE for the *overall* scenario described in the prior paragraph (7.8 vs. 8.4) (Table 3). The similarity between the other ET-derived TWSC estimates and GRACE TWSC for the *calibrated watershed* scenario increased compared to the *overall* scenario. Thus all ET products produced TWSC estimates that better aligned with GRACE-observed TWSC for this particular grouping of watersheds. This may be due to the fact that this is an unmanaged watershed whereas other watersheds in the region have more significant anthropogenic factors, as further discussed in Section 4.

For the *above* and *below the fall line* scenarios in Fig. 7, NLDAS-NOAH-derived TWSC was most similar to GRACE TWSC for both scenarios. The locally calibrated VIC-derived TWSC was similar to GRACE TWSC for the *above the fall line* scenario but not for the *below the fall line* scenario. Deep lateral groundwater flow is not well represented within our locally calibrated VIC model because it had only three layers representing the first 100 cm of the subsurface. For the below the fall line scenario, lateral groundwater flow between watersheds becomes a more dominate process because many of the watersheds in this grouping fall within the aquifer outcrop zone (Aucott and Speiran, 1985). Therefore, one possible explanation for this finding is the improved groundwater flow representation in NLDAS-NOAH compared to our locally calibrated VIC model. This hypothesis could be tested through future work to improve the groundwater flow representation within the locally calibrated VIC model and see if it results in ET-derived TWSC estimates that are more similar to GRACE-derived TWSC estimates.

Figs. 8 and 9 show the results of similar analyses that focus on periods of increasing (Fig. 8) or decreasing (Fig. 9) TWSC rather than annual periods of no net change in TWS (Fig. 7). Focusing first on periods of increasing TWS (Fig. 8, Table 3b), again the results show the strongest match between NLDAS-NOAH ET-derived TWSC and GRACE TWSC (RMSD = 3.4), but MODIS ET-derived TWSC has a similarly strong match (RMSD = 3.7). Not surprising is that all ET-derived TWSC values showed stronger similarity to GRACE TWSC over this wetting period compared to the annual period. During wet periods, ET is a less significant term in the water balance. Therefore the similarities between the ET-derived TWSC values and the GRACE TWSC values are being driven primarily by precipitation and streamflow estimates aligning with GRACE-observed changes in TWS during these wetting periods.

Table 3
Root mean squared deviation (RMSD) of predicted TWSC using each ET product to GRACE-observed TWSC over the same period. Units are cm of water.

	Overall	Calibrated	Above	Below
(a) Annual				
VIC	13.0	7.8	12.5	24.9
MODIS	12.2	26.2	16.8	10.7
NLDAS-NOAH	8.4	12.1	9.0	10.9
NLDAS-VIC	35.1	43.8	38.2	53.8
(b) Wetting periods				
VIC	6.9	13.0	10.9	10.1
MODIS	3.7	14.8	7.0	7.8
NLDAS-NOAH	3.4	8.9	4.9	8.0
NLDAS-VIC	9.6	25.6	12.0	6.7
(c) Drying periods				
VIC	9.1	2.5	10.3	7.6
MODIS	13.5	20.8	14.0	11.1
NLDAS-NOAH	16.6	12.9	16.7	10.7
NLDAS-VIC	34.5	30.6	36.8	14.3

Table 4
Correlation coefficients for TWSC estimated using each ET product (VIC, MODIS, NLDAS-NOAH, and NLDAS-VIC) and GRACE estimated TWSC for the (a) annual, (b) increasing TWS, and (c) decreasing TWS periods.

	MODIS	NLDAS-NOAH	NLDAS-VIC		MODIS	NLDAS-NOAH	NLDAS-VIC
(a) Annual							
Overall				*Calibrated Watershed*			
VIC	0.89	0.63	−0.30	VIC	0.99	0.92	−0.59
MODIS		0.88	−0.30	MODIS		0.90	−0.48
NLDAS-NOAH			0.11	NLDAS-NOAH			−0.46
Above Fall Line				*Below Fall Line*			
VIC	0.94	0.73	−0.37	VIC	0.85	0.77	0.02
MODIS		0.83	−0.45	MODIS		0.99	−0.16
NLDAS-NOAH			0.04	NLDAS-NOAH			−0.11
(b) Increasing TWS							
Overall				*Calibrated Watershed*			
VIC	0.94	0.98	0.91	VIC	0.94	0.90	0.95
MODIS		0.95	0.96	MODIS		0.99	1.00
NLDAS-NOAH			0.97	NLDAS-NOAH			0.99
Above Fall Line				*Below Fall Line*			
VIC	0.96	0.98	0.10	VIC	0.97	0.96	0.86
MODIS		0.89	−0.13	MODIS		1.00	0.95
NLDAS-NOAH			0.28	NLDAS-NOAH			0.97
(c) Decreasing TWS							
Overall				*Calibrated Watershed*			
VIC	0.77	0.78	0.67	VIC	0.40	0.57	0.39
MODIS		0.99	0.89	MODIS		0.98	0.55
NLDAS-NOAH			0.91	NLDAS-NOAH			0.66
Above Fall Line				*Below Fall Line*			
VIC	0.81	0.79	0.68	VIC	0.54	0.64	0.40
MODIS		0.99	0.87	MODIS		0.98	0.75
NLDAS-NOAH			0.89	NLDAS-NOAH			0.85

Periods of decreasing TWSC (Fig. 9, Table 3c) show times when ET will be a more significant part of the water balance because of lower precipitation and streamflow rates during dry periods. For these periods, the locally calibrated VIC ET-derived TWSC showed the most similarity to GRACE TWSC observations (RMSD = 9.1). The locally calibrated VIC ET-derived TWSC matched particularly well with GRACE observed TWSC for the *calibrated watershed* scenario (RMSD = 2.5). The NLDAS-NOAH ET-derived TWSC showed a poorer match with GRACE TWSC for these drying periods (RMSD for the *overall* scenario = 16.6). Given that these ET datasets are derived from complex models each with their own input datasets, it is difficult to determine the exact reason for this result. However, for water resources management applications dependent on ET estimates, insights like this for a particular region of interest could offer valuable information when deciding between competing ET products for a specific application.

The locally calibrated VIC model better matched GRACE TWSC for watersheds below (RMSD = 7.6) vs. above (RMSD = 10.3) the fall line for the drying periods. This is in contrast to the annual periods where the opposite result was found. Drying periods result in lower groundwater levels and less groundwater discharge into rivers within these watersheds. Groundwater discharge to streams can be significant for aquifer outcrop areas below the fall line (Aucott et al., 1987), but is not well represented within the VIC model because it includes only the first 100 cm of the subsurface environment. Thus, this finding of the locally calibrated VIC model better matching GRACE TWSC for these drying conditions could be explained by the fact that groundwater discharge is less significant during these conditions.

For the *below the fall line* scenario, the ET-derived TWSC estimates are consistently lower than the GRACE-observed TWSC estimates for the wetting periods (Fig. 8) and consistently higher than the GRACE-observed TWSC estimates for the drying periods (Fig. 9). We would expect that during wetting periods lateral groundwater flow entering these watersheds is discharging from the watersheds as streamflow. If this is in fact the case, then the ET-derived TWSC estimates would be over estimating the GRACE observed TWSC for these watersheds because Eq. (1), which is used to calculate the ET-derived TWSC estimates, assumes groundwater flow exchanges are negligible. Likewise for drying periods, these below the fall line watersheds are likely losing water through groundwater recharge that is then transferred through groundwater flow out of the watershed. This would explain the underestimation of TWSC compared to GRACE observed TWSC for these watersheds during drying periods, again because of the assumption of negligible groundwater flow exchanges in Eq. (1).

3.2.4. Correlation between ET-derived TWSC

Given these complexities with both the assumption in the water balance framework (Eq. (1)) of negligible groundwater flow exchanges between watersheds and the spatial limits of GRACE-observed TWSA, another means for comparison of the ET products is through self-similarity of their derived TWSC estimates. While correlations between the ET time series would arrive at likely the similar result, comparing ET rates using TWSC takes into account periods where ET is a more or less important term in the water balance. Table 4 provides correlations between the ET-derived TWSC values for all three analysis periods and for all four watershed grouping scenarios.

From the annual time period (Table 4a), there is a very strong correlation between the locally-calibrated VIC ET-derived TWSC and the MODIS-derived TWSC for the *calibrated watershed* scenario and very strong correlation between the MODIS-derived TWSC and the NLDAS-NOAH-derived TWSC for the *below the fall line* scenario. For the wetting period (Table 4b), not surprisingly we see strong correlations across all comparisons with the exception of the NLDAS-VIC ET-derived TWSC estimates for the above the fall line scenario. Again this is due to ET not being a dominant term in the water balance for wetting periods. For the drying periods (Table 4c) we see strong correlation between MODIS ET-derived TWSC and NLDAS-NOAH ET-derived TWSC for all scenarios.

For the *calibrated watershed* scenario for which the locally calibrated VIC ET-derived TWSC showed strong agreement with GRACE TWSC, there is little correlation between the locally

calibrated VIC ET-derived TWSC and the other three ET-derived TWSC estimates. However, for the same scenario, the MODIS ET-derived TWSC and the NLDAS-NOAH ET-derived TWSC have a very strong correlation. Thus either both the MODIS and NLDAS-NOAH ET rates are highly correlated but incorrect, or the GRACE-observed TWSC is not measuring the true TWSC for this *calibrated watershed* scenario. This is a good example of how the methodology is not designed to provide definite answers, because in the absence of a dense in situ ET observation network at the watershed-scale, which is rarely available, there will be uncertainty as to the 'true' ET rates. Instead, it provides a tool for exploring watershed-scale ET rates for water resource managers that must make decisions despite these uncertainties.

4. Summary, discussion, and conclusions

The primary contribution of this work is a methodology for evaluating different watershed-scale ET products. This methodology proposes a means for subdividing the landscape into watersheds, combining observed precipitation, streamflow, and GRACE TWS in a water balance framework, and methods for defining time periods over which to evaluate differences between ET-derived and GRACE-observed TWSC. The demonstration of the methodology for watersheds in South Carolina shows how the approach provides insight into ET differences within the larger context of watershed-scale water balances.

Comparison of the ET-derived TWSC estimates to GRACE TWSC estimates has some limitations that are important to understand when interpreting results from this or any application of the methodology. First, comparisons for small regions could be problematic given the 1 deg resolution of the GRACE TWSC level 3 product used in the analysis. For example, local anthropogenic impacts on water resources may not be detected in the GRACE signal given the 1 deg resolution. Also, in our water balance framework we assumed groundwater exchanges between watersheds to be a negligible component of the water budget. The assumption may be valid for the study region overall, but may not hold for certain subgroupings of watersheds, in particular the *below the fall line* scenario that includes watersheds in the aquifer outcrop zone. Relaxing this assumption in Eq. (1) would be ideal, but data to quantify this subsurface flux between watersheds will rarely be available. In such cases, it may be necessary to rely on correlation-based approaches for inter-comparing watershed-scale TWSC estimates generated using different ET products.

While we discussed uncertainties in the GRACE TWSA estimates throughout the paper, there are also uncertainties in the observed data including uncertainties in precipitation areal-estimates and discharge observations. For example, stream discharge at USGS gauging stations often rely on rating curves to estimate discharge from river stage observations. In some cases, such as GRACE TWS, it is possible to quantify uncertainties in the data and models used for this analysis. In other cases, such as the ET products from the land surface models, errors are not easily quantified without significant effort to explore structural, parameter, calibration, and other possible sources of error. Thus care should be taken when drawing conclusions from results of the methodology to be consistent with the level of confidence in the data and models used for the analysis.

Based on the results of applying the methodology to South Carolina watersheds we found that, overall, NLDAS-NOAH ET produced TWSC estimates that were most similar to GRACE TWSC. However, limiting the study to only periods of decreasing TWSC, when ET becomes a more significant term in the water balance equation, the locally calibrated VIC model produced ET estimates that, when inserted into the water balance framework, better aligned with GRACE observed TWSC. We also found that for the specific watershed used to calibrate the locally calibrated VIC model, the ET estimates generated for this watershed produced TWSC estimates that showed the strongest agreement with GRACE TWSC observations, especially for the period of decreasing TWSC. Therefore, we suspect that a more thorough local calibration of the VIC model using multiple monitoring locations throughout the region in a multi-objective calibration routine could potentially improve the ET estimates generated by the locally calibrated VIC model.

We also inter-compared the ET-derived TWSC estimates by calculating calibration coefficients. The assumption is that, if one or more methods converge on the same ET time series for the watersheds, it should increase our confidence in the ET estimate in the absence of in situ watershed-scale ET observations. This means of comparison is better suited for small regions below the spatial limitations of GRACE or for watersheds that have a significant groundwater exchanges that are assumed to be negligible within the water balance framework comparison methodology. For example, in the application of the methodology for watersheds in South Carolina, watersheds below the fall line, where surface water/groundwater interactions are more significant, showed less agreement with GRACE TWSC during wetting and drying periods presumably due to groundwater exchanges not accounted for in the water balance framework.

The example application of the methodology for watersheds in South Carolina also provides further evidence of the difficultly in determining watershed-scale ET rates. Without in situ ET estimates with sufficient coverage for measuring watershed-scale ET fluxes, which are not available in this case and rarely available for other watersheds throughout the world, there will still be uncertainty as to the true ET rates at this scale. With the exception of the NLDAS-VIC model, it is not possible to definitely conclude which of the three other ET products is most accurate for the South Carolina example application. We suspect this will be true for other applications of the methodology as well, based on the variability in regional ET rates on a monthly time scale found through the NLDAS-2 validation efforts (Xia et al., 2012b). Furthermore, it may be that certain ET modeling methods perform better under certain situations compared to others (e.g., wet vs. dry periods; above vs. below the fall line), as we saw in the South Carolina application, based on the underlying assumptions and data used for each prediction method. Despite these challenges, applying the methodology presented in this paper to a region and time period provides insight into these characteristics of different ET products that are valuable for water resources management and not evident through direct analysis of ET time series alone.

Acknowledgement

This work was supported in part by the US National Science Foundation under the award number ACI:0940841.

References

Abdulla, F.A., Lettenmaier, D.P., 1997a. Development of regional parameter estimation equations for a macroscale hydrologic model. J. Hydrol. 197, 230–257. http://dx.doi.org/10.1016/S0022-1694(96)03262-3.

Abdulla, F.A., Lettenmaier, D.P., 1997b. Application of regional parameter estimation schemes to simulate the water balance of a large continental river. J. Hydrol. 197, 258–285. http://dx.doi.org/10.1016/S0022-1694(96)03263-5.

Aucott, W.R., Speiran, G.K., 1985. Ground-water flow in the coastal plain aquifers of South Carolina. Ground Water 23, 736–745. http://dx.doi.org/10.1111/j.1745-6584.1985.tb01952.x.

Aucott, W.R., Meadows, R.S., Patterson, G.G., 1987. Regional Ground-water Discharge to Large Streams in the Upper Coastal Plain of South Carolina and Parts of North Carolina and Georgia (No. 86-4332), U.S. Geological Survey Water-Resources Investigations Report, Columbia, SC.

Billah, M.M., Goodall, J.L., 2011. Annual and interannual variations in terrestrial water storage during and following a period of drought in South Carolina, USA. J. Hydrol. 409, 472–482. http://dx.doi.org/10.1016/j.jhydrol.2011.08.045.

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: groundwater loss in Colorado River Basin. Geophys. Res. Lett. http://dx.doi.org/10.1002/2014GL061055.

Crow, W.T., Wood, E.F., Pan, M., 2003. Multiobjective calibration of land surface model evapotranspiration predictions using streamflow observations and spaceborne surface radiometric temperature retrievals. J. Geophys. Res. 108. http://dx.doi.org/10.1029/2002JD003292.

Gibson, W.P., Daly, C., Kittel, T., Nychka, D., Johns, C., Rosenbloom, N., McNab, A., Taylor, G., 2002. Development of a 103-year high-resolution climate data set for the conterminous United States. In: 13th AMS Conference on Applied Climatology. American Meteorological Society, Portland, OR, pp. 181–183.

Goodall, J.L., Maidment, D.R., 2009. A spatiotemporal data model for river basin-scale hydrologic systems. Int. J. Geogr. Inf. Sci. 23, 233–247. http://dx.doi.org/10.1080/13658810802032193.

Landerer, F.W., Swenson, S.C., 2012. Accuracy of scaled GRACE terrestrial water storage estimates. Water Resour. Res. 48. http://dx.doi.org/10.1029/2011WR011453.

Liang, X., Lettenmaier, D.P., Wood, E.F., Burges, S.J., 1994. A simple hydrologically based model of land surface water and energy fluxes for general circulation models. J. Geophys. Res. 99, 14415. http://dx.doi.org/10.1029/94JD00483.

Liang, X., Wood, E.F., Lettenmaier, D.P., 1996. Surface soil moisture parameterization of the VIC-2L model: evaluation and modification. Global Planet. Change 13, 195–206. http://dx.doi.org/10.1016/0921-8181(95)00046-1.

Loucks, D.P., 2000. Sustainable water resources management. Water Int. 25, 3–10. http://dx.doi.org/10.1080/02508060008686793.

McKay, L., Bondelid, T., Dewald, T., Rea, A., Johnston, C., Moore, R., 2012. NHDPlus Version 2: User Guide.

Mitchell, K.E., 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. http://dx.doi.org/10.1029/2003JD003823.

Monteith, J., 1965. Evaporation and environment. Symp. Soc. Exp. Biol. 19, 205–234.

Moriasi, D.N., Arnold, J.G., Van Liew, M.W., Bingner, R.L., Harmel, R.D., Veith, T.L., 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Trans. ASABE 50, 885–900. http://dx.doi.org/10.13031/2013.23153.

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 Sens. Environ. 111, 519–536. http://dx.doi.org/10.1016/j.rse.2007.04.015.

Mu, Q., Zhao, M., Running, S.W., 2011. Improvements to a MODIS global terrestrial evapotranspiration algorithm. Remote Sens. Environ. 115, 1781–1800. http://dx.doi.org/10.1016/j.rse.2011.02.019.

Pruitt, W.O., Angus, D.E., 1960. Large weighing lysimeter for measuring evapotranspiration. Trans. ASABE 3, 13–15.

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 Resour. Res. 42 (10). http://dx.doi.org/10.1029/2005WR004331.

Rodell, M., 2004. Basin scale estimates of evapotranspiration using GRACE and other observations. Geophys. Res. Lett. 31. http://dx.doi.org/10.1029/2004GL020873.

Rodell, M., Famiglietti, J.S., 1999. Detectability of variations in continental water storage from satellite observations of the time dependent gravity field. Water Resour. Res. 35, 2705–2723. http://dx.doi.org/10.1029/1999WR900141.

Rodell, M., Famiglietti, J.S., 2001. An analysis of terrestrial water storage variations in Illinois with implications for the Gravity Recovery and Climate Experiment (GRACE). Water Resour. Res. 37, 1327–1339. http://dx.doi.org/10.1029/2000WR900306.

Rodell, M., Famiglietti, J., Chen, J., Seneviratne, S., Viterbo, P., Holl, S., Wilson, C., 2004. Basin scale estimates of evapotranspiration using GRACE and other observations. Geophys. Res. Lett. 31.

Rodell, M., McWilliams, E.B., Famiglietti, J.S., Beaudoing, H.K., Nigro, J., 2011. Estimating evapotranspiration using an observation based terrestrial water budget. Hydrol. Process. 25, 4082–4092. http://dx.doi.org/10.1002/hyp.8369.

Ruiz-Barradas, A., Nigam, S., 2006. Great Plains hydroclimate variability: the view from North American Regional Reanalysis. J. Clim. 19, 3004–3010. http://dx.doi.org/10.1175/JCLI3768.1.

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

Sheffield, J., Ferguson, C.R., Troy, T.J., Wood, E.F., McCabe, M.F., 2009. Closing the terrestrial water budget from satellite remote sensing. Geophys. Res. Lett. 36. http://dx.doi.org/10.1029/2009GL037338.

Shepard, D.S., 1984. Computer mapping: the SYMAP interpolation algorithm. In: Spatial Statistics and Models. Springer, pp. 133–145.

Swenson, S., Wahr, J., 2002. Methods for inferring regional surface-mass anomalies from Gravity Recovery and Climate Experiment (GRACE) measurements of time-variable gravity. J. Geophys. Res. 107. http://dx.doi.org/10.1029/2001JB000576.

Syed, T.H., Webster, P.J., Famiglietti, J.S., 2014. Assessing variability of evapotranspiration over the Ganga river basin using water balance computations. Water Resour. Res. 50, 2551–2565. http://dx.doi.org/10.1002/2013WR013518.

Troy, T.J., Wood, E.F., Sheffield, J., 2008. An efficient calibration method for continental-scale land surface modeling. Water Resour. Res. 44. http://dx.doi.org/10.1029/2007WR006513.

USEPA, USGS, 2005. National hydrography dataset plus - NHDPlus. Available from: <ftp://ftp.horizon-systems.com/NHDPlus/documentation/metadata.pdf> (verified 09.08.2010).

Wahr, J., Swenson, S., Zlotnicki, V., Velicogna, I., 2004. Time-variable gravity from GRACE: first results. Geophys. Res. Lett. 31. http://dx.doi.org/10.1029/2004GL019779.

Wei, H., Xia, Y., Mitchell, K.E., Ek, M.B., 2013. Improvement of the Noah land surface model for warm season processes: evaluation of water and energy flux simulation. Hydrol. Process. 27, 297–303. http://dx.doi.org/10.1002/hyp.9214.

Wood, E.F., Lettenmaier, D.P., Zartarian, V.G., 1992. A land-surface hydrology parameterization with subgrid variability for general circulation models. J. Geophys. Res. 97, 2717. http://dx.doi.org/10.1029/91JD01786.

Xia, Y., Mitchell, K., Ek, M., Cosgrove, B., Sheffield, J., Luo, L., Alonge, C., Wei, H., Meng, J., Livneh, B., Duan, Q., Lohmann, D., 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. http://dx.doi.org/10.1029/2011JD016051.

Xia, Y., Mitchell, K., Ek, M., Sheffield, J., Cosgrove, B., Wood, E., Luo, L., Alonge, C., Wei, H., Meng, J., Livneh, B., Lettenmaier, D., Koren, V., Duan, Q., Mo, K., Fan, Y., Mocko, D., 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. http://dx.doi.org/10.1029/2011JD016048.

Xia, Y., Ek, M., Sheffield, J., Livneh, B., Huang, M., Wei, H., Feng, S., Luo, L., Meng, J., Wood, E., 2013. Validation of Noah-simulated soil temperature in the North American Land Data Assimilation System Phase 2. J. Appl. Meteorol. Climatol. 52, 455–471. http://dx.doi.org/10.1175/JAMC-D-12-033.1.

Xu, C.-Y., Chen, D., 2005. Comparison of seven models for estimation of evapotranspiration and groundwater recharge using lysimeter measurement data in Germany. Hydrol. Process. 19, 3717–3734. http://dx.doi.org/10.1002/hyp.5853.

Xu, C., Gong, L., Jiang, T., Chen, D., Singh, V.P., 2006. Analysis of spatial distribution and temporal trend of reference evapotranspiration and pan evaporation in Changjiang (Yangtze River) catchment. J. Hydrol. 327, 81–93. http://dx.doi.org/10.1016/j.jhydrol.2005.11.029.

Yeh, P.J.-F., Irizarry, M., Eltahir, E.A.B., 1998. Hydroclimatology of Illinois: a comparison of monthly evaporation estimates based on atmospheric water balance and soil water balance. J. Geophys. Res. 103, 19823. http://dx.doi.org/10.1029/98JD01721.

Zeng, N., Yoon, J.-H., Mariotti, A., Swenson, S., 2008. Variability of basin-scale terrestrial water storage from a PER water budget method: the Amazon and the Mississippi. J. Clim. 21, 248–265. http://dx.doi.org/10.1175/2007JCLI1639.1.

Zhang, L., Hickel, K., Dawes, W.R., Chiew, F.H.S., Western, A.W., Briggs, P.R., 2004. A rational function approach for estimating mean annual evapotranspiration. Water Resour. Res. 40. http://dx.doi.org/10.1029/2003WR002710.

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