A Climate Data Record (CDR) for the global terrestrial water budget: 1984-2010

Yu Zhang[1], Ming Pan[1], Justin Sheffield[1], Amanda Siemann[1], Colby Fisher[1], Miaoling Liang[2], Hylke Beck[1], Niko Wanders[1], Rosalyn MacCracken[3], Paul R. Houser[3], Tian Zhou[4], Dennis P. Lettenmaier[5], Rachel T. Pinker[6], Janice Bytheway[7], Christian D. Kummerow[7], Eric F. Wood[1]

[1]Department of Civil and Environmental Engineering, Princeton University, Princeton, NJ 08544, USA
[2]National Meteorological Center, China Meteorological Administration, Beijing, 100081, China
[3]George Mason University, Fairfax, VA 22030, USA
[4]Pacific Northwest National Laboratory, Richland, WA 99352, USA
[5]Department of Geography, University of California - Los Angeles, Los Angeles, CA 90095, USA
[6]Department of Meteorology, University of Maryland, College Park, MD 20742, USA
[7]Department of Atmospheric Science, Colorado State University, Fort Collins, CO 80523, USA

Correspondence to: Ming Pan (mpan@princeton.edu)

Abstract. Closing the terrestrial water budget is necessary to providing consistent estimates of budget components for understanding water resources and changes over time. Given the lack of in-situ observations of budget components at anything but local scale, merging information from multiple data sources (e.g. in-situ observation, satellite remote sensing, land surface model and reanalysis) through data assimilation techniques that optimize the estimation of fluxes is a promising approach. In this study, a systematic method is developed to optimally combine multiple available data sources for precipitation (P), evapotranspiration (ET), runoff (R) and the total water storage change ($TWSC$) at 0.5° spatial resolution globally and to obtain water budget closure (i.e. to enforce $P - ET - R - TWSC = 0$) through a Constrained Kalman Filter (CKF) data assimilation technique. The resulting long-term (1984-2010), monthly, 0.5° resolution global terrestrial water cycle Climate Data Record (CDR) dataset is developed under the auspices of the National Aeronautics and Space Administration (NASA) Earth System Data Records (ESDRs) program. This dataset serves to bridge the gap between sparsely gauged regions and the regions with sufficient in-situ observations in investigating the temporal and spatial variability in the terrestrial hydrology at multiple scales. The CDR created in this study is validated against in-situ measurements like river discharge from the Global Runoff Data Centre (GRDC) and the United States Geological Survey (USGS) and *ET* from FLUXNET. The dataset is shown to be reliable and can serve the scientific community in understanding historical climate variability in water cycle fluxes and stores, benchmarking the current climate, and validating models.

1 Introduction

Quantification of the terrestrial water budget and its evolution over time at fine spatial resolutions is critical to understanding the availability and variability of Earth's terrestrial water budget and the exchanges and interactions among the terrestrial,

atmospheric and oceanic branches of the hydrosphere, and to assess the risk of hydrological extremes such as floods and droughts at regional to global scales. Understanding the mean state and variability of the terrestrial water budget is also one of the primary goals of World Climate Research Programme's (WCRP) Global Energy and Water EXchanges (GEWEX, (Morel, 2001)) Project and the National Aeronautics and Space Administration (NASA) Energy and Water cycle Study (NEWS: (NASA NEWS Science Integration Team, 2007). The overarching goal of GEWEX is to "reproduce and predict, by means of suitable models, the variations of the global hydrological regime, its impact on atmospheric and surface dynamics, and variations in regional hydrological processes and water resources and their response to changes in the environment, such as the increase in greenhouse gases" (http://www.gewex.org/gewex_overview.html). The grand challenge of the NEWS project is "to document and enable improved, observationally-based, predictions of energy and water cycle consequences of Earth system variability and change" (http://www.nasa-news.org). Toward these goals, a number of Earth System Data Records (ESDRs) for the major components of the terrestrial water budget are developed under NASA's Making Earth Science data record for Use in Research Environments (MEaSUREs) program. While the MEaSUREs program refers to long-term, satellite based data records as Earth Science Data Records (ESDR), they are generally referred to as Climate Data Records (CDR) following the National Research Council report where a CDR is defined as "a time series of measurements of sufficient length, consistency and continuity to determine climate variability and change" (National Research Council, 2004). We will refer to the data set developed and described in this paper as a CDR.

The terrestrial water budget consists of four major components: precipitation (P), evapotranspiration (ET), runoff (R) and total water storage change ($TWSC$) as shown in Eq. (1). The total water storage change over a time interval ($TWSC$) is balanced by the difference between the incoming water flux of precipitation (P) and outgoing water fluxes of evapotranspiration (ET) and surface and sub-surface runoff (R) for a control volume from the Earth's surface to a lower bound at depth:

$$TWSC = P - ET - R \quad (1)$$

In-situ observations are often considered as the ground "truth" to quantitatively estimate the water budget terms. However, limited network coverage, especially for data sparse regions, has resulted in a long-time challenge for assessing the terrestrial water budget. Presently, satellite remote sensing has become a major data source to measure the various terms because of its generally global coverage and sufficient temporal repeat times. A number of satellite based products have been developed to estimate precipitation over the globe, including the Tropical Rainfall Measuring Mission (TRMM) Multi-satellite Precipitation Analysis (TMPA: (Huffman et al., 2007;Huffman et al., 2010)), the Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks-Cloud Classification System (PERSIANN-CCS: (Hong et al., 2007)), and the Climate Prediction Center morphing method (CMORPH: (Joyce et al., 2004)). For evapotranspiration, global estimates can be derived from a combination of satellite surface radiation budget (SRB), surface meteorology and vegetation cover (Fisher et al., 2008;Mu et al., 2007;Vinukollu et al., 2011;Zhang et al., 2010;Zhang et al., 2015). With the NASA Gravity Recovery And Climate Experiment (GRACE) mission, launched in March, 2002, (Landerer and Swenson, 2012;Tapley et al., 2004;Wahr et al., 2004), the changes in gravity detected by the GRACE satellites can be used to derive estimates of TWSC, albeit at relatively coarse scale. GRACE has been widely used to study changes in the terrestrial water storage (Rodell et al., 2009;Rodell et al.,

2011); the terrestrial water budget (Long et al., 2014a;Long et al., 2014b;Pan et al., 2012;Sahoo et al., 2011;Gao et al., 2010;Sheffield et al., 2009;Wang et al., 2014) and hydrological extremes such as droughts (Thomas et al., 2014;Famiglietti, 2014). For runoff, earlier studies estimated the global mean terrestrial runoff by simply calculating the differences between precipitation and evapotranspiration under the assumption of negligible long-term total water storage change (Berner and Berner, 1987;Baumgartner and Reichel, 1975). But this "inferred" runoff estimation approach can only be applied to estimate the long term mean since water storage change cannot be neglected at short temporal scales, e.g. daily, monthly or seasonally. Furthermore, human interaction with the storage might also play an important role. For example, reservoir filling after construction, inter-annual reservoir storage changes, and groundwater pumping (Rodell et al., 2009;Famiglietti, 2014;Voss et al., 2013) can significantly contribute to observed storage changes at regional scales. As an alternative, river discharge can be estimated from satellite altimetry (Birkett et al., 2002;Berry et al., 2011), for example, the future Surface Water Ocean Topography (SWOT) (Durand et al., 2010) mission. These satellite missions provide a promising and cost-efficient way of estimating individual water budget components. However, when combined together, they do not close the water budget because of errors in the individual component estimates. (Sheffield et al., 2009) found that high bias in satellite precipitation, particularly in the summer, was the major factor in budget non-closure over the Mississippi River basin. (Gao et al., 2010) also concluded that water budget closure over 13 large continental rivers in the US was not achieved using remote sensing data mainly due to the biases in precipitation and *ET*.

In addition to space-borne satellites, our understanding of the hydrological cycle in data-scarce regions has also depended on other data sources such as land surface models (LSM) (Trenberth et al., 2007;Trenberth and Fasullo, 2013b), and weather/climate reanalysis (Reichle et al., 2011). Off-line LSM simulations can provide long-term budget estimates with closure by design (Nijssen et al., 2014;Sheffield and Wood, 2007;Trenberth et al., 2007;Oki and Kanae, 2006). Reanalysis model output provides information that can be used to estimate the water budget at basin to continental (Betts et al., 2003a;Betts et al., 2003b;Betts et al., 2005) and global (Reichle et al., 2011;Balsamo et al., 2015) scales. These large scale land surface and reanalysis models have pushed the global water budget inventories into a new era where sparse traditional in-situ observations are supplemented.

However, different types of uncertainties exist in these sources of information including those in the parameterizations (satellite retrieval algorithms, LSM and reanalysis process representations), in LSM parameters such as soil and vegetation properties, and forcing data (surface radiation and meteorology) and in reanalysis data assimilation procedures. Therefore, an optimal "combination" of all data sources, including in-situ and remote sensing, LSM and reanalysis data, with their extensive spatial coverage and fine resolution, has the potential to overcome the limitation of relying on a single data source, and to offer improved accuracy, spatial and temporal coverage, and consistency in creating long-term, large scale water budget information at fine spatial resolutions (Pan et al., 2012).

To address the non-closure problem, techniques have been developed to assess the uncertainties of each budget component and to enforce water budget closure from either multiple data sources (Pan et al., 2012) or single source (Sahoo et al., 2011), usually at the scale of major river basins across the globe. For example, (Rodell et al., 2015) recently quantified the mean

annual and monthly water budgets over continents and ocean basins for the first decade of the 21st century by using data sets that combine satellite remote sensing and conventional observations. In this study, the Constrained Kalman Filter (CKF), which is a simplified version (non-ensemble) of the constrained ensemble Kalman filter (CEnKF, (Pan and Wood, 2006)), is chosen to close water balance. The CKF is a non-ensemble form, and is a standalone procedure after a regular Kalman Filter update, thus it is ideal for closing water balance without filtering or data assimilation.

Building on an increasingly available inventory of global water budget data sets from in-situ, satellite, reanalysis and land surface models, this study reported here has five advances over previously reported work. These are to: (1) expand the use of the CKF data assimilation technique in closing the water budget from that reported by (Pan et al., 2012) and (Sahoo et al., 2011), (2) extend the data records back in time to 1984 (versus to 2000 in (Rodell et al., 2015) and forward to 2010 (vs previous analyses which usually stop near the turn of the 21st century), (3) refine the spatial resolution to 0.5° for the land surface (versus basin scale analysis in (Pan et al., 2012) and (Sahoo et al., 2011), and continental and oceanic analysis in (Rodell et al., 2015) and (Trenberth and Fasullo, 2013a)), and account for the oblateness of Earth, (4) develop a harmonized global terrestrial water cycle CDR by merging the full combination of in-situ and satellite remote sensing observations, LSM simulations, and reanalysis model outputs at monthly and 0.5º spatial resolution for the period 1984 – 2010. The CDR data set includes estimates for all major terrestrial water budget variables (i.e. *P*, *ET*, *R* and *TWSC*) with budget closure at the grid-scale, and finally (5) validate the CDR against in-situ observations not used in the development of the data set.

To the authors' knowledge, this paper presents the first attempt to estimate over multiple decades the global terrestrial water budget (Greenland and Antarctica excluded), with closure at a 0.5° grid scale using this diversity of observational data sources. The data set provides comprehensive and detailed information for water budget analyses over land, and will be of particular significance in those sparsely or ungauged regions for understanding historical climate variability of the water cycle, and for benchmarking and validating climate models.

In developing the dataset, significant challenges are faced that need to be addressed. These included: (1) How consistent are the different products at different spatial scales? (2) What is the best approach to assess the uncertainty of each individual product and then optimally merge them? And, (3) what is the spatial and temporal variability of the non-closure errors and how to attribute them? Given the developed CDR, a key question is whether the merged dataset is in agreement with in-situ observations, and thus be able to capture historical hydro-climatological events (e.g. floods and droughts)?

Section 2 introduces the data sources and the methodology. Section 3 carries out a consistency and uncertainty analysis for the multiple input data sources, and investigates the spatial variability of the non-closure errors and their attribution during the budget closure enforcement process. Budget estimates based on the closure constrained dataset are presented at global, continental and large basin scales. Then the CDR is validated against in-situ runoff and *ET* in section 4. Conclusions from the research and future work are discussed in Section 5.

2 Data Description, Analysis and Methodology

In this study, the water budget is estimated and constrained at 0.5°, monthly, for the global land area excluding Antarctica and Greenland. In addition, continental and basin-scale budget estimates are also provided, including six continents and 32 major basins (Figure 1) from across the world with a range of climatic regimes. Information about the input data sources (data length, original spatial and temporal resolution, and references) is listed in Table. 1. The Community Land Model (CLM) and NOAH land surface models are used for seasonal cycle analysis but are not included later in the merging and constraining algorithm because of significant disagreement between their seasonal cycles and observations, as discussed in sections 2.1.3 and 2.1.4. The 27-year period is divided into four consecutive sub-periods (1984-1997, 1998-2002, 2003-2007, and 2008-2010) based on the data availability and overlap (Table. 2). Note that the total water storage from GRACE for the initial year 2002 is excluded from the study due to missing values.

2.1 Input Datasets

2.1.1 Precipitation

A set of precipitation products is evaluated including the remote sensing precipitation product from Colorado State University (CSU: (Bytheway and Kummerow, 2013)) with uncertainty estimates, the gauge-based Global Precipitation Climate Centre (GPCC) product (Schneider et al., 2014), the multi-source merged products of the Princeton Global Forcing dataset (PGF: (Sheffield et al., 2006)), and the Climate Hazard group InfraRed Precipitation with Stations (CHIRPS: (Funk et al., 2014)). Please refer to Supplement I for more information on these data sets.

Figures 2 and 3 show the seasonal cycles of these four precipitation products over six continents and over twelve selected representative basins distributed in different continents and climate regimes, for their overlapping period of 1998-2010. The coefficient of variation (CV), calculated as the standard deviation divided by the ensemble mean, is plotted to quantify the uncertainties among the precipitation product ensemble of PGF, CSU, GPCC and CHIRPS. The CV is first calculated for each grid cell and then averaged over continents or basins. There is no spatial coverage beyond 50°N to 50°S from CSU or CHIRPS. Therefore, only the grids between 50°N and 50°S are used to calculate the seasonal cycles in Figure 2. Likewise in Figure 3, only PGF and GPCC are compared over those basins which are either outside or extend poleward of 50°N (or 50°S) (e.g. Lena and Mackenzie river basins). Similar to the conclusion of (Pan et al., 2012), who examined a different set of datasets, the spread among these four products is higher in the densely gauged continents in Europe and North America (and basins in those two continents such as the Danube and Mississippi), with a CV ranging from 5-12% and 2-8%, respectively (Figures 2 and 3), than in the sparsely gauged regions, such as Amazon (Figure 3). There is an “abnormal” high spread (high CV) for the Niger River basin (sparsely gauged) during the dry season because the ensemble mean of the four precipitation products, is close to zero (Figure 3). The uncertainties are also high for the Mekong River basin where the rainfall totals are high and dominated by the monsoon season (Figure 3). The high uncertainties in less densely gauged regions could originate from the different

gauge densities from different products, or the ways in which the data are merged and gridded. It is interesting to note that the average discrepancy between the highest estimates (CSU) and the lowest (CHIRPS) over Europe is around 15mm/month throughout the year (Figure 2). This discrepancy is more prominent at basin scales; for example, the monthly mean difference between CSU and CHIRPS in the densely gauged basins such as Danube and Mississippi is around 20 mm/month (Figure 3). CHIRPS is a blended precipitation product (e.g. precipitation climatology, remote sensing from multiple sources, seasonal forecast form Climate Forecast System Version 2 (CFSv2), and in situ observations) but it is dominated by gauge corrections in regions with higher gauge density such as Europe and North America, and therefore in basins such as the Danube and Mississippi. The differences among the three gauge-merged products PGF, GPCC and CHIRPS might possibly be from the different data sources that they merge rather than from gauge observations, different numbers of gauges used and under-catch corrections. The seasonal cycles in Figures 2 and 3 are consistent with the climate regimes, e.g., the inversed seasonality in the Murray-Darling basin, the high peak in South America in March, and wet summer in low latitudes.

2.1.2 Evapotranspiration

Unlike precipitation with relatively dense in-situ observations, especially for developed regions, in-situ based evapotranspiration estimations (from flux towers) are very sparse. Here we collect ten gridded global terrestrial evapotranspiration (*ET*) products, of which five are satellite derived, two are reanalysis products and three are from land surface models. One satellite product is the Global Land-surface Evaporation: the Amsterdam Methodology (GLEAM: (Miralles et al., 2011). As parts of the MeaSUREs products, the four other satellite products are derived using two algorithms, the Penman-Monteith (PM) and Priestly-Taylor (PT), cross-combined with two forcing inputs that are different from the other six *ET* products, the SRB-CFSR (Surface Radiation Budget – Climate Forecast System Reanalysis) and SRB-PGF. These four products are referred as: SRB-CFSR-PM, SRB-CFSR-PT, SRB-PGF-PM and SRB-PGF-PT (Vinukollu et al., 2011). Satellite remote sensing, carries the mission of observing Earth at fine spatial resolution and comprehensive coverage and makes it possible to estimate water budget in sparsely gauged regions. Therefore, 5 satellite *ET* products are merged into the CDR. The two reanalysis *ET* products are from the ERA-interim (Simmons et al., 2006) and NASA's Modern-Era Retrospective Analysis for Research and Application (MERRA: (Rienecker et al., 2011). The LSM *ET* datasets are from the Variable Infiltration Capacity model (VIC v4.0.6), CLM v3.5 and NOAH v3.4 forced by an updated version of PGF. Please refer to Supplement I for more information.

The ten *ET* products show less consistency in the seasonal cycle (Figure 4 and 5) than the precipitation datasets (Figures 2 and 3). At continental scales (Figure 4), the reanalysis *ET* products (ERA-Interim and MERRA) generally have relatively high values for the six continents, while the LSMs generally predict lower values over Asia, Europe and North America. Most of the satellite *ET* products (i.e., GLEAM, SRB-CFSR-PM, SRB-CFSR-PT, SRB-PGF-PM, and SRB-PGF-PT) lie between the reanalysis and LSMs in Asia, Europe and North America. More striking is the relative lack of consistency among those ten *ET* products for the wet tropical basins (Amazon, Congo and Mekong). The seasonality of *ET* over these basins is complex because of the overall energy limitation but seasonally and spatially varying moisture limitation (Guan et al., 2015). These results imply

that the ten approaches have significant differences in their derived surface radiation budget and meteorology as well as the parameterizations of evaporative processes (potential *ET*, transpiration, interception, and soil evaporation) and their interaction with phenological and environmental controls. The relatively higher consistency of the remotely sensed algorithms for these basins is in part a result of using the same (or closely similar) surface radiation budget but different meteorological forcings.

2.1.3 Runoff

The three LSMs are forced by the same meteorological forcing from PGF to simulate global runoff over land. The VIC simulation was calibrated over 43 well-distributed major global basins against the measured streamflow data from the Global Runoff Data Center (GRDC, http://grdc.bagf.de) while CLM and Noah are un-calibrated. Please refer to Supplement I for additional model information under the evapotranspiration section. Figures 6 and 7 display the seasonal cycles over the six continents and the twelve representative major river basins. Noah shows opposite seasonal cycle against VIC and CLM in Europe and North America, which include high latitude regions (Figure 6). Unlike VIC and Noah, CLM almost shows no seasonal cycle in Oceania (Figure 6). The disagreement between the LSMs can also be found at basin scales (e.g. Danube, Lena. Mackenzie, Yukon and Murray-Darling in Figure 7). Additionally, Figure 8 shows the verification of the runoff from the LSMs against GRDC observations for 26 basins that have available data records longer than 3 years during 1984-2010. Noah shows a negative runoff bias against GRDC for most of the mid to high latitude basins (Columbia, Danube, Indigirka, Kolyma, Kena, Mackenzie, Northern Dvina, Ob, Olenek, Pechora, Yenisei and Yukon; Figure 8). CLM has better performance over high latitude basins than Noah but it shows a high overestimation of runoff for the Danube, and Don (Figure 8). None of the LSMs capture the seasonal cycles for the Indus and Senegal basins. Nonetheless, the authors recognize that runoff estimates using a number of LSMs (e.g. (Haddeland et al., 2011)) can provide uncertainty estimates (i.e. spread or standard deviation among different data sources) in simulated runoff. However, CLM and Noah runoff estimates are not merged into the CDR developed in this study in order to avoid the large biases from their un-calibrated parameters. Additional reasons for not merging CLM and Noah are discussed in section 2.1.4.

2.1.4 Total Water Storage Change

The total water storage change ($TWSC$), which measures the changes in total water storage during specific unit period, is taken from the LSMs and the GRACE data. The GRACE monthly total water storage anomaly ($TWSA$) time series, which are anomalies relative to the 2004-2009 time-mean baseline from ReLease 05 (RL05) that are processed by three centers, Geoforschungs Zentrum Potsdam (GFZ), Center for Space Research at University of Texas, Austin (CSR), and Jet Propulsion Laboratory (JPL), are used to calculate the $TWSC$ via the backward difference equation in Eq.(2) and central difference equation in Eq.(3). Comparisons indicate that the central difference calculation Eq. (3) is in better agreement with VIC inferred $TWSC$. Therefore, the central difference $TWSC$ has been used.

$$TWSC = (TWSA_t - TWSA_{t-1})/\Delta t \quad (2)$$

$$TWSC = (TWSA_{t+1} - TWSA_{t-1})/2\Delta t \quad (3)$$

Different parameters and solution strategies were explored and applied by these three processing centers and the differences between the centers have generally decreased over the Releases (https://grace.jpl.nasa.gov/data/choosing-a-solution/). Even though VIC only computes the water storage in the upper few meters of the soil column (depending on the calibrated storage capacity in its 2nd and 3rd layers), this is the most active part of the soil column. Therefore studies (e.g. (Gao et al., 2010;Tang et al., 2010) found reasonable agreement between changes in *TWSC* from GRACE and the VIC model. Similar results were also found in this study: $TWSC$ from VIC and GRACE (from GFZ, CSR, and JPL) are in good agreement at both continental (Figure 9) and basin scales (Figure 10) except for some timing lags in the high latitude basins of the Lena and Yukon. This lag between GRACE derived minimum $TWSC$ and VIC inferred minimum $TWSC$ suggests that the snowmelt (and subsequent runoff) starts earlier in VIC than observed by GRACE, or that more snowmelt ponds into wetlands or discharges into lakes, neither of which are well represented in VIC as its snow melt discharges more directly into rivers. In contrast, Noah shows a reversed seasonal cycle in those high latitude continental regions and basins such as Asia, North America, Danube, Lena, Mackenzie and Yukon, while CLM shows disagreement in the seasonal cycle in Oceania as well as in the Danube and Mississippi basins relative to GRACE observations. Not surprisingly, the spread within the three GRACE products is very small compared to the differences against VIC (Figures 9 and 10). (Sakumura et al., 2014) found that the ensemble mean (simple arithmetic mean of JPL, CSR, GFZ) was the most effective method in reducing the noise in the gravity field solutions within the available scatter of the solutions. Therefore, the ensemble mean of the $TWSC$ from GFZ, CSR, and JPL is taken as the best $TWSC$ product derived from GRACE, and this is used in the later water budget analysis together with *TWSC* from VIC.

2.2 Methods

All the datasets, as listed in Table 1, are first aggregated or disaggregated to 0.5°spatially and monthly values using bilinear interpolation, then the errors/uncertainties of each product are assessed. Estimates for the same water budget variable are then merged following the algorithm described in (Luo et al., 2007). The merged water budget estimates are further adjusted to ensure closure at every grid using the Constrained Kalman Filter (CKF) approach of (Pan et al., 2012). Then, the unconstrained and constrained water budgets are analyzed at different scales. Figure 11 provides a flow chart of the procedure.

2.2.1 Uncertainty estimation and data merging technique

There is no best estimate or observation of each individual water budget component at the grid scale over the globe due to the limited spatial coverage of in-situ measurements. This is especially true for evapotranspiration observations from the flux tower networks. Thus, the limited availability of gridded ground observations makes it impossible to quantify the error in each water budget component. Therefore, in this study, the deviation from the ensemble mean of all data sources for the same budget variable is used as a proxy of the uncertainty/error in individual products. The merging procedure for each budget component

is a weighted, averaging where the optimal merging weight w_i is given by the following equation (Luo et al., 2007;Sahoo et al., 2011):

$$w_i = \frac{1}{\sigma_i^2} / \sum_{i=1}^{n} \frac{1}{\sigma_i^2} \quad (4)$$

in which w_i is the merging weight for product i, σ_i^2 is the error variance of product i calculated against the ensemble mean, and n is the total number of products. Note that $\sum w_i$ equals to 1. The larger the error variance of product i, the lower is its weight.

The number of products merged into single water budget estimate varies in the different sub-periods due to the data availability (Table 2). A "data consistency adjustment" is applied after the data merging process in order to guarantee the consistency of the climate data record (CDR) estimated in this study. Taking precipitation as an example, first, for the period with complete data records (i.e. 1998-2008), the inner-annual monthly mean precipitation merged from all the available products (i.e. PGF, GPCC, CHIRPS, CSU), and the mean precipitation merged from the available products (i.e. PGF, GPCC, CHIRPS) during the incomplete data records period (i.e. 1984-1997 during which the CSU is not available) are calculated, respectively. Then the inter-annual monthly climatological bias, which is the monthly mean precipitation merged from PGF, GPCC, CHIRPS and CSU minus that merged from PGF, GPCC and CHIRPS, is simply added to the inter-annual monthly mean precipitation during the incomplete data records period (i.e. 1984-1997). This "data consistency" approach aims to avoid the "jump" in the merged precipitation time series in the year 1998 when the CSU became available. The same procedure is then applied to adjust the data consistency for *ET* during 2008-2010 and *TWSC* during 1984-2002. We contend that this is a key step, as the temporal consistency of the CDR will impact the reproduction of historical hydrological extremes and the analysis of long-term trends for all the available water budget variables.

2.2.2 Enforcing water budget closure using CKF

In short, CKF redistributes the non-closure errors back onto the various water budget components according to their error levels and correlations. We define the water balance residual as $r = P - ET - R - TWSC$. If we write the budget components as a column vector x: $x = [P, ET, R, TWSC]^T$, then the residual of the water balance can be expressed as a linear function of the vector, $r = G\,x$, where $G = [1, -1, -1, -1]$. The error covariance matrix of x is calculated as $\varepsilon_{xx} = \overline{(\hat{x} - x)(\hat{x} - x)^T}$, where $\hat{x}$ is an estimate of x, its "true value". In this study, $(\hat{x} - x)$ is replaced with the spread of the ensemble in each water budget component. This uncertainty estimation method was first proposed by (Adler et al., 2001) and then applied in (Tian and Peters-Lidard, 2010) to generate a global precipitation uncertainty map for a variety of satellite remote sensing products. ε_{xx} has dimensions of 4×4 since x consists of 4 budget variables. Then the balance-constrained estimate is calculated from $\hat{x}' = \hat{x} - \varepsilon_{xx} G^T (G \varepsilon_{xx} G^T)^{-1} \hat{r}$. The residual term $\hat{r}$ is redistributed back onto the various water budget variables through the above equation. Mathematically, the CKF algorithm mimics assimilating a "perfect" (zero-error) observation of $r = 0$. Further details are presented in (Pan and Wood, 2006). In this study, the error of runoff is simply assumed as 10% as VIC is the single source of runoff. The water budget closure is done monthly based on variational error from month to month.

3 Water Budget Merging and Constraint

3.1 Data Merging

All the products for the same water budget component are merged into a single estimate based on their uncertainties/errors relative to their ensemble mean as described in section 2.2.1. The values in Table 2 summarize the mean merging weights of each individual product for different periods. Please refer to Figures S1 to S3 in Supplement II for the spatial maps of the merging weights from different products. The global mean merging weights for the precipitation are calculated over 50°N-50°S during 1984-1997 and 1998-2010. CHIRPS and CSU only cover 50°N-50°S; therefore, for those regions outside 50°N-50°S, PGF and GPCC are merged with equal weights (50%). Before the availability of the CSU product in 1998, the average merging weights of PGF, GPCC, and CHIRPS over 50°N-50°S (land) are 29.6%, 34.6% and 35.8%. CHIRPS is closest to the ensemble mean especially for the Amazon basin, and therefore has a higher weight in that region (Figure S1 in Supplement II). For the period 1998-2010 when CSU becomes available, CHIRPS (26.5%), GPCC (26.8%) and CSU (26.0%) have similar weights. Note that the weights vary with time and location. The annual mean of the merged precipitation is 767.0 mm for 1984-1997, 792.7 mm for 1998-2002, 786.7 mm for 2003-2007 and 802.9 mm for 2008-2010 (Table 3). Equivalent numbers at monthly scale are displayed in Figure 12 in terms of global maps. The values from Table 3 and Figure 12 are calculated using the data consistency adjustment described in section 2.2.1.

For evapotranspiration, the averaged merging weights over land for each product are: VIC (11.3%), ERA (10.8%), MERRA (6.6%), GLEAM (12.8%), SRB-PGF-PM (17.2%), SRB-PGF-PT (15.9%), SRB-CFSR-PM (13.9%), and SRB-CFSR-PT (11.5%) during their common period 1984-2007 (Figure S2 in Supplement II). Among the eight ET products, MERRA has the lowest averaged merging weight as it has a relatively larger deviation from other ET products at both continental (Figure 4) and basin scales (Figure 5). In the Amazon basin, MERRA shows nearly an opposite seasonal cycle against other ET products (Figure 5) and thus its merging weight is extremely low there (Figure S2 in Supplement II). The merged ET in the unconstrained budgets, averaged over land, are 518.0 mm/year, 523.6 mm/year, 516.0 mm/year and 522.0 mm/year throughout those four sub-periods (Table 3, the spatial maps can be found in Figure 12).

The runoff simulated from VIC is used as the "merged" terrestrial runoff at the grid scale since the gauge observations are discrete and spatially incomplete. The annual averaged runoff over the globe is 338.9 mm/year during 1984-2010 (Table 3, see Figure 12 for the spatial maps for the sub-periods).

For the total water storage change, the uncertainty in VIC inferred storage change and GRACE derived storage change are simply assumed to be 5% and 10% of their actual values due to the lack of a better source for their validation (Pan et al., 2012). Consequently, the higher merging weight from VIC (67.1%) and lower merging weight from GRACE (32.9%) in Table 2 (and Figure S3 in Supplement II for the spatial maps of merging weights) are a result of the assigned error ratios (i.e. 5% and 10%). Given the good agreement in $TWSC$ between VIC and GRACE (Figures 9 and 10), the impact of such a subjective error assignment is relatively small. But for high-latitude basin such as Yukon where VIC and GRACE have relative large

discrepancy, the error is relatively high. Globally, the monthly mean of $TWSC$ is almost zero during the four sub-periods as shown in the fourth row of Figure 12. Nonetheless, multi-year variability due to drought and wet periods is observable. For example, the long-term drought in the central U.S. and Canadian prairies over the 1998-2002 period shows up as do the Brazilian droughts in 1994-1995 and 2004-2005 that extended into Argentina (2004-2006). Also seen in Figure 12 is the wetting trend over the last two decades of the Sahel since the severe mid-1980's drought as well as the floods in Brazil in 2008.

3.2 Data Assimilation to close the water budget

The last row of Figure 12 shows the global maps of the non-closure errors for the sub-periods. The long-term, mean non-closure error relative to precipitation is around -9.8% over land during 1984-2010 (Table 3). The annual mean imbalance over land ranges from -55.3 mm/year to -80.6 mm/year during the four sub-periods (Table 3).

Figure 13 shows an example of the unconstrained (left) and constrained (right) water budgets for the Amazon basin together with imbalances and their attribution (bottom). Over the Amazon basin where the total precipitation is large and the gauges are sparse, the precipitation uncertainty is higher. This results in precipitation being the main recipient of the non-closure error attribution (third row on the right in Figure 13), receiving around 50% of the non-closure error for each of the sub-periods as well as the complete analysis period. Due to the "inconsistencies" in terms of different numbers of available data sources merged into the budget during the four consecutive sub-periods, the imbalance/non-closure error (third row on the left in Figure 13) does not show a regular seasonal cycle and a continuous pattern of imbalance.

The annual mean water budget in terms of *P*, *ET*, *R* and *TWSC* after water balance constraint are 781.8 mm, 463.9 mm, 318.0 mm and 0 mm during 1984-2010, respectively (Table 3). Note that direct application of CKF to enforce the water balance without other constraints may possibly lead to a non-zero *TWSC* over long term, and sometimes a negative runoff. Therefore, two additional "filters" are added after the CKF. First, if the runoff is negative, we will re-run the CKF and only re-distribute the non-closure error onto the other three budget components. Second, for each grid cell if the long-term mean *TWSC* over 1984-2010 is not zero, the monthly long term mean *TWSC* will be subtracted from the *TWSC* and added to the precipitation and evapotranspiration month by month during 1984-2010. Figure S4 in Supplement II shows the mean water budget components after the CKF water balance enforcement in addition to the mean water budget components before the enforcement in Figure 12. The long term mean of *TWSC* at each grid cell is zero over the entire 27 years after the second filter, which is also named as "*TWSC* de-trending". Though at regional scales, some places have experienced groundwater depletions such as US high plains and central valley, western Iran, India etc., starting from different years. One of the challenges is a lack of data on groundwater extractions. Therefore, from the global perspective, for almost three decades during the study period 1984-2010 covered by this study, the authors assume the long term *TWSC* to be zero thus apply the de-trending, after which the spatial variability of *TWSC* still exists during the four sub-periods (Figure S4). A more comprehensive comparison of the water budget estimation before and after the closure enforcement is listed in Table 4 at both the continental and basin scales. These water budget component values are spatially and temporally aggregated for each continent or basin over the analysis period of 1984-2010.

The attribution of the non-closure term for each water budget variable is based on the uncertainties among different products. The results from this study are in general agreement with (Pan et al., 2012) where the authors showed that ET has a high non-closure attribution in a large portion of the 32 river basins that they analyzed. The average attribution of non-closure errors to ET over the globe is 45.4% during 1984-2010 compared to 38.4% for P, 4.9% for R and 11.2% for $TWSC$ (see Table 3). For most of the regions ET receives the highest attribution of the non-closure error, particularly, in Africa (50% attributed to ET vs. 37% to precipitation, 3% to runoff, and 10% to $TWSC$; see Table 4), and Oceania (46% attributed to ET vs. 41% to precipitation, 2% to runoff, and 10% to $TWSC$; see Table 4). Figure S5 additionally shows the global maps of the mean water budget non-closure error attribution during different sub-periods. Higher attributions to precipitation occur in basins in mid to high latitudes such as the Danube (42% to precipitation vs. 38% to ET, 6% to runoff, and 12% to $TWSC$; see Table 4), and Don (42% to precipitation vs. 39% to ET, 3% to runoff, and 16% to $TWSC$; see Table 4), where the estimation of extreme rainfall rates remain less well resolved (Huffman et al., 2007;Yong et al., 2014). High non-closure attributions to precipitation also occur in tropical basins such as the Amazon (46% to precipitation vs. 33% to ET, 9% to runoff, and 12% to $TWSC$; see Table 4) and Congo (46% to precipitation vs. 37% to ET, 6% to runoff, and 11% to $TWSC$; see Table 4) because the precipitation is large and the gauges are scarce in these basins. The attribution to the total water storage change is generally small except for the northern regions where snow, ice melt and seasonal storage changes in wetlands dominate the water budgets (Figure S5 in Supplement II). Runoff receives the smallest attribution of the imbalance among the four water budget components for most regions over the globe, which is in agreement with what was concluded in (Sahoo et al., 2011). The mean attributions of each water budget component over different continents and basins over 1984-2010 are listed in Table 4 as well.

4 Validation of the MEaSUREs Global Terrestrial Water Budget CDR

The final CDR, which is the constrained global water budget with closure, is validated against in-situ observations in terms of runoff and ET at multiple spatial scales.

4.1 Runoff Verification

In-situ river discharge observations are collected from three major data sources: (1) Global Runoff Data Center (GRDC), (2) United States Geological Survey (USGS), and (3) Australian Land and Water Resources Audit project (Peel et al., 2000). The observations were collected from GRDC for a total number of 32 large basins and 26 of them are used (as shown in Figure 8) after filtering out those basins with less than three years of data during 1984-2010. Figure 1 provides the locations of these basins. 165 out of a total of 362 medium sized basins (5,000 to 10,000 km^2, 331 from GRDC and 31 from USGS) were selected for validations. For validation over small basins, discharge data for 862 basins (1,000 to 5,000 km^2) were collected from GRDC, USGS and the Australian Land and Water Resources Audit project. Basins under any one or more of the following conditions were excluded: (1) GRDC basins for which the catchment boundaries could not be reliably determined; (2) basins with large dams (reservoir capacity greater than 10% of annual streamflow); (3) basins with urban areas greater than 2% (using the

''artificial areas'' class of the map from GlobCover, version 2.3; (Bontemps et al., 2011)); (4) basins with irrigated areas greater than 2% (using the Global Irrigated Area Map; http://www.iwmigiam.org); and (5) basins with either a gain or loss forest (change in land cover) > 20% of the basin area. For both the medium and small basins, those basins with data records length less than 5 years were also excluded. Figure 14(a) displays the locations of medium and small basins. The observed discharge data were converted to runoff by dividing by the basin area upstream of the gauge location.

The seasonal cycles of runoff from the CDR created in this study over the 26 large basins, are compared against the GRDC observations as shown in Figure 8. Not surprisingly, the runoff estimated from the constrained system (grey dashed line) is not much different from the runoff estimated in the unconstrained system (which is VIC runoff shown by the solid blue line) as we assign a small error (10%) on the runoff component within the budget constraint algorithm. In general, VIC outperforms the other two LSMs as VIC was calibrated over 43 major global river basins (Sheffield and Wood, 2007) although the calibration periods varied. Therefore, we believe that VIC can provide a reliable grid-scale estimate of runoff budget. Note that the seasonal peaks from Noah and VIC are in agreement for the Indus basin but their peaks precede the peak from the GRDC observations, which strangely happen in November. Comparing to other studies for the Indus River (Bookhagen and Burbank, 2010), show that the discharge peak occurs in the summer time , which is consistent with VIC and Noah. Likewise for Senegal River, records from regional studies (Andersen et al., 2001) and (Stisen et al., 2008) show runoff peaks in August to September instead of April to May from the GRDC record. In summary, we believe that our CDR provides good runoff estimates over the Amur, Danube, Mackenzie, Mekong, Mississippi, Pearl, Pechora, Yangtze and Yenisei rivers but unsatisfactory estimates over the Congo, Lena, Murray-Darling and Yellow rivers in that the predicted seasonal discharge differ significantly from the observed seasonal cycle. The reasons for this are due to water management not being included in the VIC model (e.g. Murray-Darling and Yellow rivers), a combination of scarce data and not including large wetlands (e.g. the Congo and Lena basins).

By filtering out those basins with non-significant correlations, Figure 14 compares CDR estimated runoff against in-situ observations for 165 medium basins and 862 small basins in terms of correlation coefficient (CC, Figure 14 (a) and (b)), and scatter plots (Figure 14 (c) and (d)), at the monthly scale. Again, the observed discharge measurements are converted to runoff using the basin area. 84 out of 165 medium basins (~ 51%) and 625 out of 862 small basins (~ 73%) have CC values that are larger than 0.5 as shown in Figure 14 (a) and (b). There are some medium basins with extreme low CC values (red dots in Figure 14(a)) in northern Canada where the lake/wetland influences are not modelled, and in south Africa where the sporadic rainfall is not picked up and the model fails to replicate the quick runoff. The runoff from the CDR have CC values of 0.86 and 0.83 for the same medium and small basins as shown in Figure 14 (a) and (b), and have a Bias Ratio of 6% for medium basins and -16% for small basins (Figure 14 (c) and (d)). There is also a tendency for the model to underestimate runoff in the small basins in wetter regions (Figure 14 (d)). This scatter may be due forcing uncertainty, model calibration or omitted processes like water management (reservoirs, irrigation), all which might shift the timing of the runoff peak, particularly on a monthly basis. For the small basins, though they were filtered in an attempt to remove basins impacted by factors such as reservoirs, irrigation, urbanization, and so forth, they might be impacted by the scaling issues. The CDR was computed at 0.5°

grid resolution, which is approximately 50km near the equator. The small basins range from 1,000 – 5,000 km^2 so that the small basins only cover a maximum of two grid pixels and a minimum of 0.2 of a grid pixel for the smallest basin. The basin masks were extracted at a higher spatial resolution and then aggregated onto the 0.5° grids with the fractional area for the basin in order to minimize the impact of spatial mismatch. Nonetheless, the coarser spatial resolution of the CDR still affects the comparison of the runoff estimates with small-scale basin observations. No estimate of this resolution effect has been determined but the results shown in Figure 14(d) suggest that the effect is limited to a small number of basins.

4.2 *ET* Verification

Estimated *ET* is verified by two different approaches: first against an inferred *ET* that is computed from the difference between observed precipitation minus observed discharge (*P-R*) at the annual scale to minimize the effect of seasonal *TWSC.* This is done for the 25 large, 169 medium and 813 small basins which are selected by the criteria of no less than 5 years' annual records. And it is then secondly verified against in-situ observations from FluxNet tower data (Baldocchi et al., 2001).

The precipitation used in computing the inferred *ET* is from GPCC, which is a gridded rain gauge analysis that merges around 67,000 gauge measurements globally (Schneider et al., 2014). The observed runoff, *R*, is from the same sources as used in section 4.1. As shown in Figure 15, the correlation coefficients between MEaSUREs CDR *ET* and inferred *ET* from observed *P-R* are 0.97, 0.96 and 0.76 for those large, medium and small basins. For some of the MEaSUREs CDR *ET* over medium basins, particularly wetter basins, do not match well with the observed *P-R*, are attributed to the effects of water management on our estimates of *R.* Essentially, if the CDR runoff that doesn't reflect water management is too large, then the estimates of *ET* will be too low, which is what is seen in Figure 15(b). The small basins show worse agreements with the inferred ET with 20% Bias (versus 4% Bias for large basins and -4% Bias for medium basins in Figure 15) that we attribute to scaling effects in estimating *R* than for the medium basins.

The *ET* estimates from the CDR are further assessed by comparing the grid-scale estimates with observations from 47 FluxNet towers, which measures the turbulent latent flux using the eddy covariance technique. Those 47 Flux Towers were selected based on data availability (Michel et al., 2015) in terms of the meteorological variables and radiations, and the final selection represents a variety of biomes and dry/wet climate regimes. The raw data are at 3 hourly and the most complete data were recorded during the warm seasons. Therefore, the comparisons are made only over the summer (warm) seasons by filtering out those years with less than 70% data records based on the data availability at each tower. The 47 flux towers are located in four continents (North America, Europe, Asia and Africa) as shown under different land covers that are defined by the International Geosphere-Biosphere International Program (IGBP, (Loveland et al., 2000)) in Figure 16(a). The tower stations are also described in Table 5. The validations against the FluxNet observations are only carried out during the warm season when *ET* is more dominant and when there are fewer missing values. From the 47 flux towers, we found out that our *ET* estimates from the CDR are in high agreement with FluxNet observations under the land cover types WSA (Woody Savannas, one station in Africa and the other one in the US) and EBF (Evergreen Broadleaf Forest, only one station in France, Figure 16(b)). In general, our CDR *ET* matches well with the observation with a correlation coefficient of around 0.77 and a Bias

Ratio of 11% except for some over estimations for the stations, most of which are under the land cover CRO (Cropland) and ENF (Evergreen Needleleaf Forest, Figure 16(b)). The positive Bias of MEaSUREs CDR *ET* relative to FluxNet observations are attributed to the tower management – during the rainy days in the summer the flux towers are usually turned off thus underestimate the actual *ET* during the rainy days.

5 Discussion and Future Work

A well-constrained, global inventory of the historical terrestrial water budget at fine resolution is essential to understanding the terrestrial hydrological cycle, its partitioning into individual components and their variability at regional to global scales. In this study, the consistency and uncertainties of multiple hydrological data products are investigated, with precipitation found to have the highest consistency among the available products at both continental and basin scales compared to ET and $TWSC$. Data products from multiple sources that include in-situ and satellite remote sensing observations, land surface model estimates, and reanalysis model outputs are combined to create homogenized terrestrial water budget estimates at 0.5° spatial and monthly temporal scales for the period 1984-2010. This long-term water budget data record has both spatial and temporal consistency, and is part of NASA's the Earth System Data Records (ESDRs) program. The CDR data set was created by applying a water balance closure constraint using the CKF data assimilation technique of (Pan and Wood, 2006). For the individual data products, their ensemble mean is taken as the best estimate for the variable, and the ensemble spread against the ensemble mean as a proxy for their uncertainty. These estimates of the mean and uncertainty for the product are important assumptions underlying the development of data records. The CDR is validated against ground observations, i.e. GRDC, USGS and Australian Land and Water Resources Audit project for runoff and FluxNet for *ET*, which seem not independent from the merged and constrained CDR. However, data developed from either satellite remote sensing or model are often calibrated against "ground truth", i.e., gauge observations, which are also the best "ground truth" that are normally used for verification. The "ground truth" is no way independent from those remote sensing or modeled data, particularly for global data validation. Nevertheless, we believe these data records represent the best, current knowledge for the global terrestrial water budget at the 0.5° and monthly scale over the 27-year period of 1984-2010.

Additionally, the developed data set allows for the documentation of the water budget at continental and basin scales resulting in a better depiction across multiple scales. The attribution analysis of the budget imbalance (non-closure) shows that ET receives the largest adjustment in most regions –particularly in Africa and Oceania. In contrast, runoff tends to receive the lowest attribution of the non-closure error, in part due to the calibrated land surface model estimates from 43 large global basins. $TWSC$ receives larger adjustments in high latitude regions, which we attribute to the impacts from snowmelt and seasonal dynamics of wetlands and small lakes that are not well represented in VIC LSM.

Currently the authors are carrying out another study in comparing the CDR water budget records against around 20 high-impacted studies, at multiple spatial scales (i.e. continental and global). This on-going study is the first attempt to gather and compare global water budget estimates from studies as early as 1974 (i.e. Budyko 1974) to the current study in order to provide

a comprehensive overview of global water budget estimates, even though the studies focused on different periods using different data sources and have different global coverage (e.g. some of them exclude Antarctica or Greenland or both). Figure S6 in supplement II gives an example comparison with (Trenberth et al., 2007, T2007 hereafter), which estimated the water budget during 1979-2000 and excluded Antarctica. The total precipiation is quite close to this study ($114 \times 10^3 km^3/yr$) to T2007 ($113 \times 10^3 km^3/yr$). By converting the water budgets into mm/yr based on the global coverage information available in each of those studies, the long-term mean precipiation is around 28 mm/yr (vs. 32 mm/yr in the CDR and 27 mm/yr from T2007), *ET* is around 78 mm/yr (vs. 78 mm/yr in the CDR and 77 mm/yr from T2007), and runoff is around 47 mm/yr (vs. 46mm/yr in this study and T2007). Figure S7 further provides an example of how the CDR precipiation time series captured the 1998-1999 US drought. The 6-month SPIs exceeds the threshold of exceptional drought (which is defined by the US drought Monitor system; http://droughtmonitor.unl.edu/AboutUs/ClassificationScheme.aspx) around the year 1998. The CDR developed in this study, as a time series of measurements of sufficient length, consistency and continuity, can also be applied to determine climate variability. Figure S8 in the supplement II, as an example, provides the inter-annual variability of the available water (*P-ET*) over the globe during the CDR period 1984 -2010.

The major challenge for the creation of ESDR/CDR of the terrestrial water budget (and potentially the terrestrial surface energy budget) is the lack of "ground truth" observations that can serve as reference data sets for bias correction. The sparseness of the observations in accessible data archives (e.g. GRDC: Global Runoff Data Centre for river discharge, GPCC: Global Precipitation Climatology Centre for precipitation and publically accessible and quality-controlled FluxNet data) is both a scientific and institutional challenge. Many additional gauge locations and data records exist and could contribute to the development of improved CDR and our understanding of climate variability and change but have not been made available. Besides these operationally-focused observations, the relative inaccessibility of global FluxNet tower observations is also disappointing, although this situation has improved over the recent past. Even though there are over 650 towers in 30 regional networks covering 5 continents, the free fair-use subset of the La Thuile FluxNet dataset (which has been harmonized, standardized and gap filled) contains only 154 stations of which 47 were deemed useful for the validation presented here, based on quality assessment (e.g. closure of the energy budget) and record length. Data availability and accessibility challenges need to be at the top of the agendas of the world's major space agencies (ESA, NASA, JAXA), international data programs such as the Global Climate Observing System (GCOS), Global Energy and Water EXchange (GEWEX) project of WCRP, Global Earth Observing System of Systems (GEOSS), and international agencies like the World Meteorological Organization. The "standard" statements and claims about "free and open access" to climate data from these programs have not resulted in improved access. If the needed improvements to CDR are to occur, and must occur to better assess the impacts from global environmental change, improved in-situ data archiving and access by the scientific community is imperative for a more accurate analysis of climate variability and change.

The CDR developed in this study – the global terrestrial water budget at 0.5°, monthly for 1984-2010 is currently archived on our public server and available at http://stream.princeton.edu:8080/opendap/MEaSUREs/WC_MULTISOURCES_WB_050/ and will be formally archived at

the NASA Goddard Earth Science Data and Information Services Center (GES DISC) for the future use of climate and water management communities, and will advance our understanding climate variability and trends at multiple spatial scales.

As the authors are aware, essential directions in global water and energy cycle research are towards improved understanding historical climate, benchmarking future climate predictions, validating models, and improving the understanding of the interactions among land, ocean and atmosphere hydrospheres. Future work will be targeted on extending the data sets to even longer period, and at finer resolutions, by combining upcoming new satellite missions and the analysis and predictions from more advanced modelling systems.

6 Acknowledgements

This study was made possible under the support of NASA grants NNX08AN40A (Developing Consistent Earth System Data Records for the Global Terrestrial Water Cycle) under NASA's Making Earth Science data record for Use in Research Environments (MEaSUREs) program, and NNX09AK35G (Development and diagnostic analysis of a multi-decadal global evaporation product for NEWS) under the NASA Energy and Water System (NEWS) program. The support from these programs is highly appreciated.

7 References

Adler, R. F., Kidd, C., Petty, G., Morissey, M., and Goodman, H. M.: Intercomparison of global precipitation products: The third Precipitation Intercomparison Project (PIP-3), Bulletin of the American Meteorological Society, 82, 1377-1396, 2001.

Andersen, J., Refsgaard, J. C., and Jensen, K. H.: Distributed hydrological modelling of the Senegal River Basin—model construction and validation, Journal of Hydrology, 247, 200-214, 2001.

Baldocchi, D., Falge, E., Gu, L., Olson, R., Hollinger, D., Running, S., Anthoni, P., Bernhofer, C., Davis, K., Evans, R., Fuentes, J., Goldstein, A., Katul, G., Law, B., Lee, X., Malhi, Y., Meyers, T., Munger, W., Oechel, W., Paw, K. T., Pilegaard, K., Schmid, H. P., Valentini, R., Verma, S., Vesala, T., Wilson, K., and Wofsy, S.: FLUXNET: A New Tool to Study the Temporal and Spatial Variability of Ecosystem–Scale Carbon Dioxide, Water Vapor, and Energy Flux Densities, Bulletin of the American Meteorological Society, 82, 2415-2434, 10.1175/1520-0477(2001)082<2415:FANTTS>2.3.CO;2, 2001.

Balsamo, G., Albergel, C., Beljaars, A., Boussetta, S., Brun, E., Cloke, H., Dee, D., Dutra, E., Muñoz-Sabater, J., and Pappenberger, F.: ERA-Interim/Land: a global land surface reanalysis data set, Hydrology and Earth System Sciences, 19, 389-407, 2015.

Baumgartner, A., and Reichel, E.: The world water balance: Mean annual global, continental and maritime precipitation evaporation and run-off, Elsevier Science Inc, 1975.

Berner, E. K., and Berner, R. A.: Global water cycle: geochemistry and environment, Prentice-Hall, 1987.

Berry, P. A. M., Salloway, M. K., Smith, R. G., and Benveniste, J.: Global inland water monitoring from satellite radar altimetry—a glimpse into the future, IAHS-AISH publication, 104-109, 2011.

Betts, A. K., Ball, J. H., Bosilovich, M., Viterbo, P., Zhang, Y., and Rossow, W. B.: Intercomparison of water and energy budgets for five mississippi subbasins between ecmwf reanalysis (era‐40) and nasa data assimilation office fvgcm for 1990‐1999, Journal of Geophysical Research: Atmospheres (1984–2012), 108, 2003a.

Betts, A. K., Ball, J. H., and Viterbo, P.: Evaluation of the ERA-40 surface water budget and surface temperature for the Mackenzie River basin, Journal of Hydrometeorology, 4, 1194-1211, 2003b.

Betts, A. K., Ball, J. H., Viterbo, P., Dai, A., and Marengo, J.: Hydrometeorology of the Amazon in ERA-40, Journal of Hydrometeorology, 6, 764-774, 2005.

Birkett, C. M., Mertes, L. A. K., Dunne, T., Costa, M. H., and Jasinski, M. J.: Surface water dynamics in the Amazon Basin: Application of satellite radar altimetry, Journal of Geophysical Research: Atmospheres (1984–2012), 107, LBA 26-21-LBA 26-21, 2002.

Bontemps, S., Defourny, P., Bogaert, E. V., Arino, O., Kalogirou, V., and Perez, J. R.: GLOBCOVER 2009-Products description and validation report, 2011.

Bookhagen, B., and Burbank, D. W.: Toward a complete Himalayan hydrological budget: Spatiotemporal distribution of snowmelt and rainfall and their impact on river discharge, Journal of Geophysical Research: Earth Surface (2003–2012), 115, 2010.

Budyko M. 1., 1974: Climate and life. International Geophysics Series. 1978;18.

Bytheway, J. L., and Kummerow, C. D.: Inferring the uncertainty of satellite precipitation estimates in data-sparse regions over land, Journal of Geophysical Research: Atmospheres, 118, 9524-9533, 10.1002/jgrd.50607, 2013.

Durand, M., Fu, L.-L., Lettenmaier, D. P., Alsdorf, D. E., Rodriguez, E., and Esteban-Fernandez, D.: The surface water and ocean topography mission: Observing terrestrial surface water and oceanic submesoscale eddies, Proceedings of the IEEE, 98, 766-779, 2010.

Famiglietti, J. S.: The global groundwater crisis, Nature Clim. Change, 4, 945-948, 10.1038/nclimate2425, 2014.

Fisher, J. B., Tu, K. P., and Baldocchi, D. D.: Global estimates of the land–atmosphere water flux based on monthly AVHRR and ISLSCP-II data, validated at 16 FLUXNET sites, Remote Sensing of Environment, 112, 901-919, 2008.

Funk, C. C., Peterson, P. J., Landsfeld, M. F., Pedreros, D. H., Verdin, J. P., Rowland, J. D., Romero, B. E., Husak, G. J., Michaelsen, J. C., and Verdin, A. P.: A quasi-global precipitation time series for drought monitoring, U.S. Geological Survey Data Series, 832, 4p, http://dx.doi.org/10.3133/ds832, 2014.

Gao, H., Tang, Q., Ferguson, C. R., Wood, E. F., and Lettenmaier, D. P.: Estimating the water budget of major US river basins via remote sensing, International Journal of Remote Sensing, 31, 3955-3978, 2010.

Guan, K., Pan, M., Li, H., Wolf, A., Wu, J., Medvigy, D., Caylor, K. K., Sheffield, J., Wood, E. F., and Malhi, Y.: Photosynthetic seasonality of global tropical forests constrained by hydroclimate, Nature Geoscience, 8, 284-289, 2015.

Haddeland, I., Clark, D. B., Franssen, W., Ludwig, F., Voß, F., Arnell, N. W., Bertrand, N., Best, M., Folwell, S., and Gerten, D.: Multimodel estimate of the global terrestrial water balance: setup and first results, Journal of Hydrometeorology, 12, 869-884, 2011.

Hong, Y., Gochis, D., Cheng, J.-t., Hsu, K.-l., and Sorooshian, S.: Evaluation of PERSIANN-CCS rainfall measurement using the NAME event rain gauge network, Journal of Hydrometeorology, 8, 469-482, 2007.

Huffman, G. J., Bolvin, D. T., Nelkin, E. J., Wolff, D. B., Adler, R. F., Gu, G., Hong, Y., Bowman, K. P., and Stocker, E. F.: The TRMM multisatellite precipitation analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales, Journal of Hydrometeorology, 8, 38-55, 2007.

Huffman, G. J., Adler, R. F., Bolvin, D. T., and Nelkin, E. J.: The TRMM multi-satellite precipitation analysis (TMPA), in: Satellite rainfall applications for surface hydrology, Springer, 3-22, 2010.

Joyce, R. J., Janowiak, J. E., Arkin, P. A., and Xie, P.: CMORPH: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution, Journal of Hydrometeorology, 5, 487-503, 2004.

Landerer, F. W., and Swenson, S. C.: Accuracy of scaled GRACE terrestrial water storage estimates, Water Resources Research, 48, 2012.

Long, D., Longuevergne, L., and Scanlon, B. R.: Uncertainty in evapotranspiration from land surface modeling, remote sensing, and GRACE satellites, Water Resources Research, 50, 1131-1151, 2014a.

Long, D., Shen, Y., Sun, A., Hong, Y., Longuevergne, L., Yang, Y., Li, B., and Chen, L.: Drought and flood monitoring for a large karst plateau in Southwest China using extended GRACE data, Remote Sensing of Environment, 155, 145-160, http://dx.doi.org/10.1016/j.rse.2014.08.006, 2014b.

Loveland, T. R., Reed, B. C., Brown, J. F., Ohlen, D. O., Zhu, Z., Yang, L., and Merchant, J. W.: Development of a global land cover characteristics database and IGBP DISCover from 1 km AVHRR data, International Journal of Remote Sensing, 21, 1303-1330, 2000.

Luo, L., Wood, E. F., and Pan, M.: Bayesian merging of multiple climate model forecasts for seasonal hydrological predictions, Journal of Geophysical Research: Atmospheres (1984–2012), 112, 2007.

Michel, D., Jiménez, C., Miralles, D. G., Jung, M., Hirschi, M., Ershadi, A., Martens, B., McCabe, M. F., Fisher, J. B., and Mu, Q.: The WACMOS-ET project–Part 1: Tower-scale evaluation of four remote sensing-based evapotranspiration algorithms, Hydrology and Earth System Sciences Discussions, 12, 10739-10787, 2015.

Miralles, D. G., De Jeu, R. A., Gash, J. H., Holmes, T. R., and Dolman, A. J.: Magnitude and variability of land evaporation and its componentsat the global scale, Hydrology and Earth System Sciences, 2011.

Morel, P.: Why GEWEX? The agenda for a global energy and water cycle research program, GEWEX News, 11, 7-11, 2001.

Mu, Q., Heinsch, F. A., Zhao, M., and Running, S. W.: Development of a global evapotranspiration algorithm based on MODIS and global meteorology data, Remote Sensing of Environment, 111, 519-536, 2007.

NASA NEWS Science Integration Team: Predicting Energy and Water Cycle Consequences of Earth System Variability and Change, 89, http://news.cisc.gmu.edu/doc/NEWS_implementation.pdf, 2007.

National Research Council: Climate Data Records from Environmental Satellites: Interim Report, Committee on Climate Data Records from NOAA Operational Satellites, The National Academies Press, Washington, DC, 150 pp., 2004.

Nijssen, B., Shukla, S., Lin, C., Gao, H., Zhou, T., Ishottama, Sheffield, J., Wood, E. F., and Lettenmaier, D. P.: A Prototype Global Drought Information System Based on Multiple Land Surface Models, Journal of Hydrometeorology, 15, 1661-1676, 10.1175/JHM-D-13-090.1, 2014.

Oki, T., and Kanae, S.: Global hydrological cycles and world water resources, science, 313, 1068-1072, 2006.

Pan, M., and Wood, E. F.: Data assimilation for estimating the terrestrial water budget using a constrained ensemble Kalman filter, Journal of Hydrometeorology, 7, 534-547, 2006.

Pan, M., Sahoo, A. K., Troy, T. J., Vinukollu, R. K., Sheffield, J., and Wood, E. F.: Multisource estimation of long-term terrestrial water budget for major global river basins, Journal of Climate, 25, 3191-3206, 2012.

Peel, M. C., Chiew, F. H. S., Western, A. W., and McMahon, T. A.: Extension of unimpaired monthly streamflow data and regionalisation of parameter values to estimate streamflow in ungauged catchments, Report to the National Land and Water Resources Audit, 2000.

Reichle, R. H., Koster, R. D., De Lannoy, G. J. M., Forman, B. A., Liu, Q., Mahanama, S. P. P., and Touré, A.: Assessment and Enhancement of MERRA Land Surface Hydrology Estimates, Journal of Climate, 24, 6322-6338, 10.1175/JCLI-D-10-05033.1, 2011.

Rienecker, M. M., Suarez, M. J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., Bosilovich, M. G., Schubert, S. D., Takacs, L., Kim, G.-K., Bloom, S., Chen, J., Collins, D., Conaty, A., da Silva, A., Gu, W., Joiner, J., Koster, R. D., Lucchesi, R., Molod, A., Owens, T., Pawson, S., Pegion, P., Redder, C. R., Reichle, R., Robertson, F. R., Ruddick, A. G., Sienkiewicz, M., and Woollen, J.: MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications, Journal of Climate, 24, 3624-3648, 10.1175/JCLI-D-11-00015.1, 2011.

Rodell, M., Velicogna, I., and Famiglietti, J. S.: Satellite-based estimates of groundwater depletion in India, Nature, 460, 999-1002, 2009.

Rodell, M., Chambers, D. P., and Famiglietti, J. S.: Groundwater and terrestrial water storage, Bulletin of the American Meteorological Society, 92, 2011.

Rodell, M., Beaudoing, H. K., L'Ecuyer, T. S., Olson, W. S., Famiglietti, J. S., Houser, P. R., Adler, R., Bosilovich, M. G., Clayson, C. A., and Chambers, D.: The observed state of the water cycle in the early 21st century, J Clim. doi, 10, 1175, 2015.

Sahoo, A. K., Pan, M., Troy, T. J., Vinukollu, R. K., Sheffield, J., and Wood, E. F.: Reconciling the global terrestrial water budget using satellite remote sensing, Remote Sensing of Environment, 115, 1850-1865, 2011.

Sakumura, C., Bettadpur, S., and Bruinsma, S.: Ensemble prediction and intercomparison analysis of GRACE time‐variable gravity field models, Geophysical Research Letters, 41, 1389-1397, 2014.

Schneider, U., Becker, A., Finger, P., Meyer-Christoffer, A., Ziese, M., and Rudolf, B.: GPCC's new land surface precipitation climatology based on quality-controlled in situ data and its role in quantifying the global water cycle, Theor Appl Climatol, 115, 15-40, 10.1007/s00704-013-0860-x, 2014.

Sheffield, J., Goteti, G., and Wood, E. F.: Development of a 50-year high-resolution global dataset of meteorological forcings for land surface modeling, Journal of Climate, 19, 3088-3111, 2006.

Sheffield, J., and Wood, E. F.: Characteristics of global and regional drought, 1950–2000: Analysis of soil moisture data from off-line simulation of the terrestrial hydrologic cycle, Journal of Geophysical Research: Atmospheres, 112, D17115, 10.1029/2006JD008288, 2007.

Sheffield, J., Ferguson, C. R., Troy, T. J., Wood, E. F., and McCabe, M. F.: Closing the terrestrial water budget from satellite remote sensing, Geophysical Research Letters, 36, 2009.

Simmons, A., Uppala, S., Dee, D., and Kobayashi, S.: ERA-Interim: New ECMWF reanalysis products from 1989 onwards., ECMWF Newsletter, 110, 26-35, 2006.

Stisen, S., Jensen, K. H., Sandholt, I., and Grimes, D. I. F.: A remote sensing driven distributed hydrological model of the Senegal River basin, Journal of Hydrology, 354, 131-148, 2008.

Tang, Q., Gao, H., Yeh, P., Oki, T., Su, F., and Lettenmaier, D. P.: Dynamics of terrestrial water storage change from satellite and surface observations and modeling, Journal of Hydrometeorology, 11, 156-170, 2010.

Tapley, B. D., Bettadpur, S., Ries, J. C., Thompson, P. F., and Watkins, M. M.: GRACE measurements of mass variability in the Earth system, Science, 305, 503-505, 2004.

Thomas, A. C., Reager, J. T., Famiglietti, J. S., and Rodell, M.: A GRACE ‐ based water storage deficit approach for hydrological drought characterization, Geophysical Research Letters, 41, 1537-1545, 2014.

Tian, Y., and Peters ‐ Lidard, C. D.: A global map of uncertainties in satellite ‐ based precipitation measurements, Geophysical Research Letters, 37, 2010.

Trenberth, K. E., Smith, L., Qian, T., Dai, A., and Fasullo, J.: Estimates of the global water budget and its annual cycle using observational and model data, Journal of Hydrometeorology, 8, 758-769, 2007.

Trenberth, K. E., and Fasullo, J. T.: North American water and energy cycles, Geophysical Research Letters, 40, 365-369, 2013a.

Trenberth, K. E., and Fasullo, J. T.: Regional Energy and Water Cycles: Transports from Ocean to Land, Journal of Climate, 26, 7837-7851, 10.1175/JCLI-D-13-00008.1, 2013b.

Vinukollu, R. K., Meynadier, R., Sheffield, J., and Wood, E. F.: Multi-model, multi-sensor estimates of global evapotranspiration: climatology, uncertainties and trends, Hydrological Processes, 25, 3993-4010, 10.1002/hyp.8393, 2011.

Voss, K. A., Famiglietti, J. S., Lo, M., Linage, C., Rodell, M., and Swenson, S. C.: Groundwater depletion in the Middle East from GRACE with implications for transboundary water management in the Tigris ‐ Euphrates ‐ Western Iran region, Water resources research, 49, 904-914, 2013.

Wahr, J., Swenson, S., Zlotnicki, V., and Velicogna, I.: Time ‐ variable gravity from GRACE: First results, Geophysical Research Letters, 31, 2004.

Wang, S., Huang, J., Li, J., Rivera, A., McKenney, D. W., and Sheffield, J.: Assessment of water budget for sixteen large drainage basins in Canada, Journal of Hydrology, 512, 1-15, 2014.

Yong, B., Liu, D., Gourley, J. J., Tian, Y., Huffman, G. J., Ren, L., and Hong, Y.: Global view of real-time TRMM Multi-satellite Precipitation Analysis: implication to its successor Global Precipitation Measurement mission, Bulletin of the American Meteorological Society, 2014.

Zhang, K., Kimball, J. S., Nemani, R. R., and Running, S. W.: A continuous satellite ‐ derived global record of land surface evapotranspiration from 1983 to 2006, Water Resources Research, 46, 2010.

Zhang, K., Kimball, J. S., Nemani, R. R., Running, S. W., Hong, Y., Gourley, J. J., and Yu, Z.: Vegetation Greening and Climate Change Promote Multidecadal Rises of Global Land Evapotranspiration, Scientific reports, 5, 2015.

Table 1. Summary of the gridded data used in this study (The study period is 1984 – 2010; CLM and NOAH in grey are analyzed but NOT merged into the final water budget CDR in this study)

Dataset	Period	Spatial Resolution	Temporal Resolution	References
		Precipitation		
CSU	1998-2010	0.25°	3 h	Bytheway et al. (2013)
PGF	1948-2010	0.25°	3 h	Sheffield et al. (2006)
CHIRPS	1981-present	0.5°	monthly	Funk et al. (2014)
GPCC(v6)	1901-2010	0.5°	monthly	Schneider et al.(2014)
		Evapotranspiration		
VIC	1948-2010	0.25°	3 h	Sheffield and Wood (2007)
CLM	1948-2010	1°	3 h	-
NOAH	1948-2010	1°	3 h	-
ERA-Interim	1979-present	T255	-	Simmons et al. (2006)
MERRA	1979-present	2/3*1/2 H grids	-	Rienecker et al. (2011)
GLEAM(V2A)	1981-2011	0.25°	daily	Miraless et al. (2011)
SRB-PGF PM PT	1984-2010	0.5°	daily	Vinukollu et al. (2011)
SRB-CFSR PM PT	1984-2007	0.5°	daily	Vinukollu et al. (2011)
		Runoff		
VIC	1948-2010	0.25°	3 h	Sheffield and Wood (2007)
CLM	1948-2010	1°	3 h	-
NOAH	1948-2010	1°	3 h	-
		Total Water Storage (TWS)		
VIC	1948-2010	0.25°	3 h	Sheffield and Wood (2007)
CLM	1948-2010	1°	3 h	-
NOAH	1948-2010	1°	3 h	-
GRACE	2002-present	1°	monthly	Landerer et al. (2012)

Table 2. Data sources of merged water budgets with their averaged merging weights in brackets throughout different sub-periods (*TWSC* from GRACE in 2002 is uncompleted that GRACE for 2002 is excluded; the spatial maps of merging weights over the globe can be found in Figure S1 to S3 in Supplement II)

	1984-1997	1998-2002	2003-2007	2008-2010
P	PGF (29.6%), GPCC (34.6%), CHIRPS (35.8%)	PGF (20.7%), GPCC (26.8%), CHIRPS (26.5%), CSU (26.0%),		
ET	VIC (11.3%), ERA-Interim (10.8%), MERRA (6.6%), GLEAM (12.8%), SRB-PGF-PM (17.2%), SRB-PGF-PT (15.9%), SRB-CFSR-PM (13.9), SRB-CFSR-PT (11.5%)			VIC (17.4%), ERA-Interim (19.6%), MERRA (11.4%), GLEAM (20.3%), SRB-PGF-PM (18.0%), SRB-PGF-PT (13.2%)
R	VIC			
TWSC	VIC		VIC (67.1%), GRACE (32.9%)	

Table 3. Annual mean water budgets (mm/year) over the globe (Greenland and Antarctica excluded) before (normal font) /after (in bold) water balance constraint, and their attributions (in italic) to non-closure error throughout sub-periods

	1984-1997	1998-2002	2003-2007	2008-2010	1984-2010
P	767.0	792.7	786.7	802.9	779.4
	776.0	**798.0**	**779.1**	**787.0**	**781.8**
	42.9%	*35.3%*	*34.5%*	*29.4%*	*38.4%*
ET	518.0	523.6	516.0	522.0	519.1
	464.7	**467.0**	**457.6**	**465.2**	**463.9**
	42.0%	*47.7%*	*48.2%*	*53.0%*	*45.4%*
R	333.0	352.7	343.8	335.3	338.9
	312.1	**330.5**	**322.0**	**318.0**	**318.0**
	4.6%	*5.2%*	*5.2%*	*5.2%*	*4.9%*
TWSC	-3.4	-1.7	-2.4	0.9	-2.4
	-0.8	**0.6**	**-0.5**	**3.8**	**0.0**
	10.5%	*11.9%*	*12.1%*	*12.4%*	*11.2%*
Closure	-80.6	-82.0	-70.8	-55.3	-76.2

Table 4. The summary table of annual mean water budgets (mm/year) before and after water balance constraint and their corresponding attributions (%) to the non-closure error at both continental and basin scales (Greenland is excluded for North America)

	water budget terms before constrain					water budget terms after constrain				Attribution (%)			
	P	**ET**	**R**	**TWSC**	**Closure**	**P**	**ET**	**R**	**TWSC**	**P**	**ET**	**R**	**TWSC**
Africa	650	524	179	-6	-48	652	478	174	0	37	50	3	10
Asia	685	409	346	2	-91	686	359	327	0	37	45	5	12
Europe	676	410	276	4	-18	653	395	258	0	36	41	7	14
NorthAmer	686	461	318	1	-95	679	395	284	0	37	42	6	12
SouthAmer	1555	958	685	-11	-77	1575	907	668	0	42	40	7	11
Oceania	450	404	129	-6	-76	468	346	122	0	41	46	2	10
Amazon	2160	1173	1048	-14	-46	2182	1153	1029	0	46	33	9	12
Amur	405	325	135	-4	-52	424	295	129	0	36	52	4	8
Aral	239	235	75	0	-71	266	197	69	0	40	45	2	13
Columbia	581	416	318	-1	-152	624	331	293	0	37	42	7	14
Congo	1454	1063	407	-16	0	1449	1045	404	0	46	37	6	11
Danube	771	505	272	-2	-3	768	503	265	0	42	38	6	13
Dnieper	605	429	117	13	47	577	461	115	0	35	44	4	17
Don	514	402	96	19	-3	498	404	94	0	42	39	3	16
Indigirka	240	173	132	0	-65	258	138	120	0	30	50	4	17
Indus	388	338	154	-11	-94	425	277	148	0	41	46	3	11
Kolyma	276	194	125	3	-46	283	167	116	0	29	48	4	18
Lena	366	267	142	-7	-36	379	245	134	0	30	51	4	15
Limpo	506	571	42	-5	-103	537	496	41	0	38	54	1	8
Mackenzie	388	293	189	10	-105	413	241	173	0	32	49	6	13
Mekong	1496	955	655	-8	-106	1518	883	634	0	39	39	7	15
Mississippi	789	597	220	-11	-16	792	577	215	0	42	42	5	10
Murray-Darling	439	440	42	-3	-40	452	411	41	0	41	49	1	8
Niger	595	426	198	-1	-28	595	401	194	0	35	49	3	13
Nile	522	464	97	-3	-36	521	426	96	0	32	54	2	12
Noethern Dvina	598	355	318	16	-91	618	324	294	0	32	43	10	16
Ob	468	340	159	14	-45	470	323	147	0	34	45	5	17
Olenek	284	177	114	1	-8	280	174	106	0	28	47	4	21
Parana	1201	893	278	-17	47	1171	892	279	0	39	47	4	10
Pearl	1425	794	747	9	-126	1438	732	706	0	38	41	10	10
Pechora	563	246	342	57	-82	552	244	308	0	33	38	8	20
Senegal	283	238	57	-5	-7	274	218	55	0	37	50	1	10
Ural	303	286	59	-6	-36	317	260	57	0	35	44	3	18
Volga	560	391	196	3	-30	563	375	188	0	36	40	7	17
Yangtze	1017	623	522	-5	-123	1061	554	507	0	37	44	9	9
Yellow	416	372	99	-1	-54	431	336	95	0	36	51	4	9
Yenisei	437	300	217	22	-102	460	265	195	0	31	48	6	15
Yukon	307	222	149	7	-71	314	175	139	0	39	43	4	14

Table 5. Flux towers information list. From left to right the station name; data available time span; latitude; longitude; International Geosphere-Biosphere International Program (IGBP) land cover (Loveland et al. 2000)

Name	Availabile Years	Lat	Lon	VegeType
DE-Geb	2004 - 2006	51.1	10.91	CRO
DE-Kli	2004 - 2006	50.89	13.52	CRO
FR-Lam	2005 - 2005	43.49	1.24	CRO
IT-Bci	2004 - 2006	40.52	14.96	CRO
US-ARM	2003 - 2006	36.61	-97.49	CRO
US-Bo1	1997 - 2006	40.01	-88.29	CRO
US-Bo2	2004 - 2006	40.01	-88.29	CRO
IT-Noe	2004 - 2006	40.61	8.15	CSH
CA-Oas	1997 - 2005	53.63	-106.2	DBF
DE-Hai	2000 - 2006	51.08	10.45	DBF
IT-Col	1996 - 2005	41.85	13.59	DBF
IT-Ro1	2000 - 2006	42.41	11.93	DBF
US-MMS	1999 - 2005	39.32	-86.41	DBF
US-Moz	2004 - 2006	38.74	-92.2	DBF
US-WCr	1999 - 2006	45.81	-90.08	DBF
FR-Pue	2000 - 2006	43.74	3.6	EBF
CA-Ca1	1998 - 2005	49.87	-125.33	ENF
CA-Obs	1999 - 2005	53.99	-105.12	ENF
CA-Ojp	1999 - 2005	53.92	-104.69	ENF
CA-Qcu	2002 - 2006	49.27	-74.04	ENF
CA-Qfo	2004 - 2006	49.69	-74.34	ENF
DE-Tha	1997 - 2006	50.96	13.57	ENF
DE-Wet	2002 - 2006	50.45	11.46	ENF
FR-LB	1997 - 2006	44.72	-0.77	ENF
IL-Yat	2001 - 2006	31.34	35.05	ENF
IT-Lav	2001 - 2006	45.96	11.28	ENF
NL-Loo	1997 - 2006	52.17	5.74	ENF
RU-Fyo	1998 - 2006	56.46	32.92	ENF
SE-Nor	1996 - 2005	60.09	17.48	ENF
US-NR1	1999 - 2003	40.03	-105.55	ENF
US-Wrc	1998 - 2006	45.82	-121.95	ENF
DE-Meh	2004 - 2006	51.28	10.66	GRA
ES-VDA	2004 - 2006	42.15	1.45	GRA
IT-Mbo	2003 - 2006	46.02	11.05	GRA
NL-Ca1	2003 - 2006	51.97	4.93	GRA
PT-Mi2	2005 - 2006	38.48	-8.02	GRA
US-Aud	2003 - 2006	31.59	-110.51	GRA
US-Bkg	2004 - 2006	44.35	-96.84	GRA
US-CaV	2004 - 2004	39.06	-79.42	GRA
US-Fpe	2000 - 2006	48.31	-105.1	GRA
US-Goo	2002 - 2006	34.25	-89.87	GRA
US-Wkg	2004 - 2006	31.74	-109.94	GRA
JP-Tom	2001 - 2003	42.74	141.51	MF
CA-Mer	1998 - 2005	45.41	-75.52	WET
CN-Do2	2005 - 2005	31.58	121.9	WET

BW-Ma1	2000	-	2001	-19.92	23.56	WSA
US-SRM	2004	-	2006	31.82	-110.87	WSA

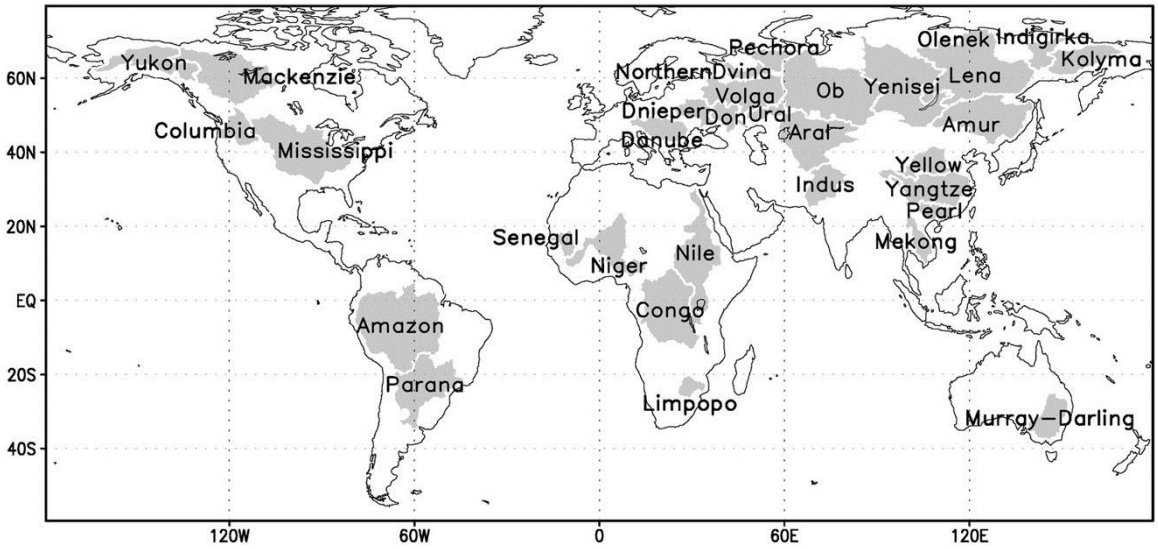

Figure 1. Locations of 32 selected large basins (Pan et al. (2012))

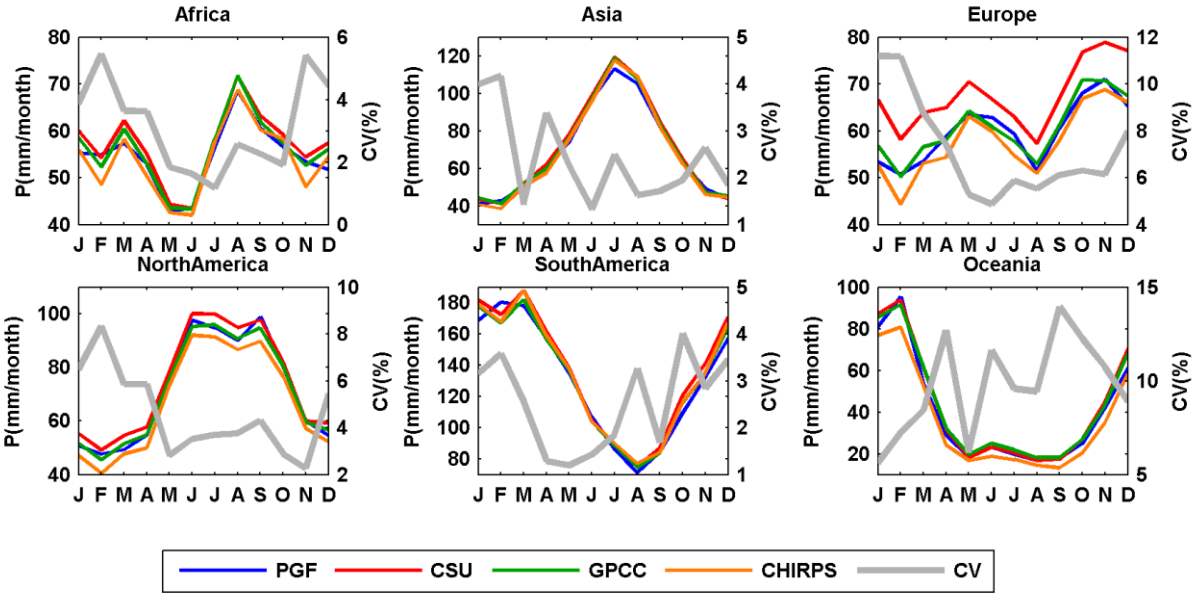

Figure 2. Seasonal cycles of precipitation from different products over the six continents for 1998-2010 (CHIRPS and TMPART only cover the region between 50°N-50°S; therefore, only the grids between 50°N-50°S are counted into the calculation of the seasonal cycle; the Coefficient of Variance (CV, %) is calculated as the standard deviation divided by the ensemble mean of all the products (The same for Figures 3-9))

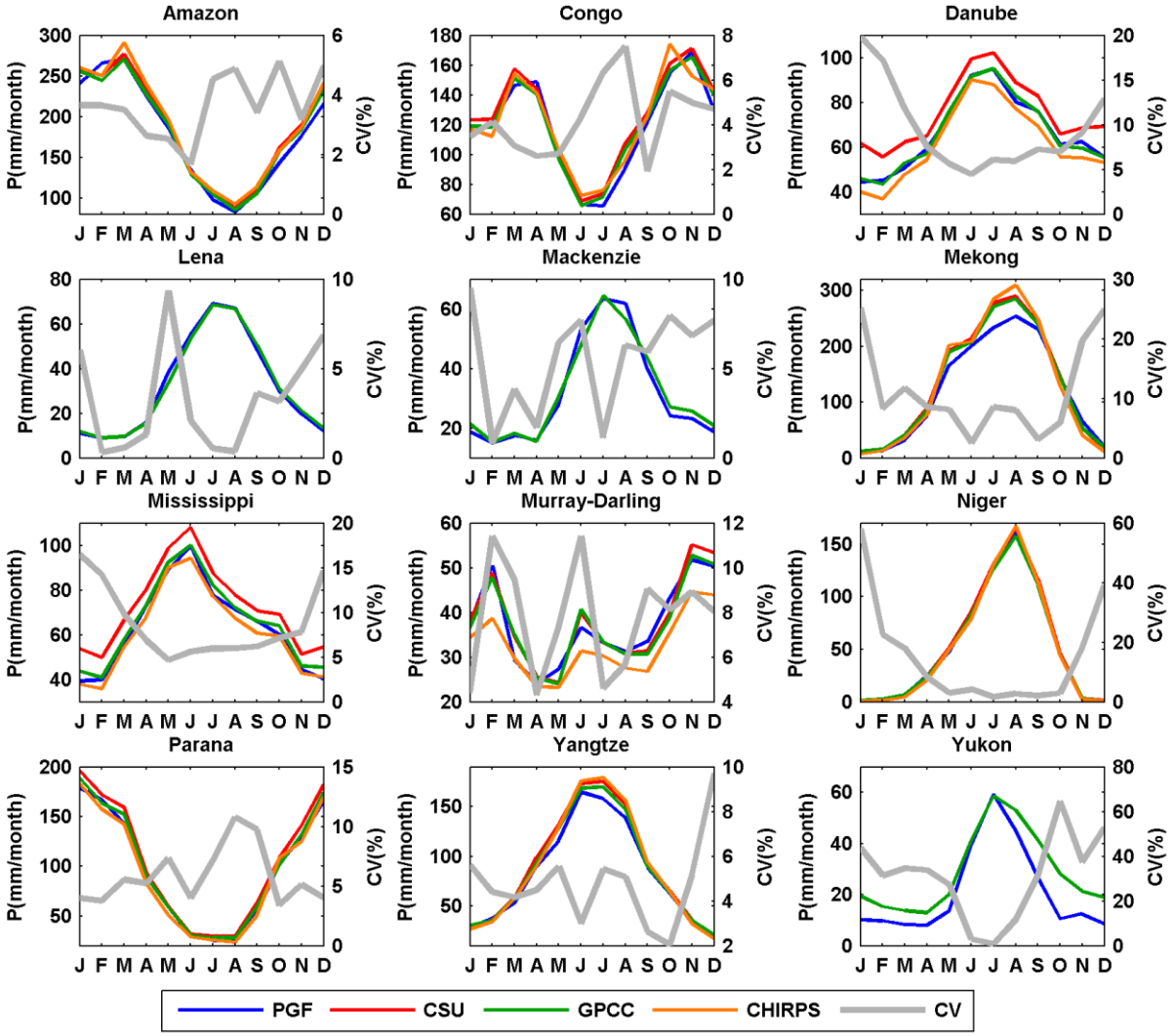

Figure 3. Seasonal cycles of precipitation from different products over twelve representative large basins for 1998-2010 (CHIRPS and CSU only cover the region between 50°N-50°S. For those basins either outside or across 50°N-50°S, only PGF and GPCC are visualized)

5

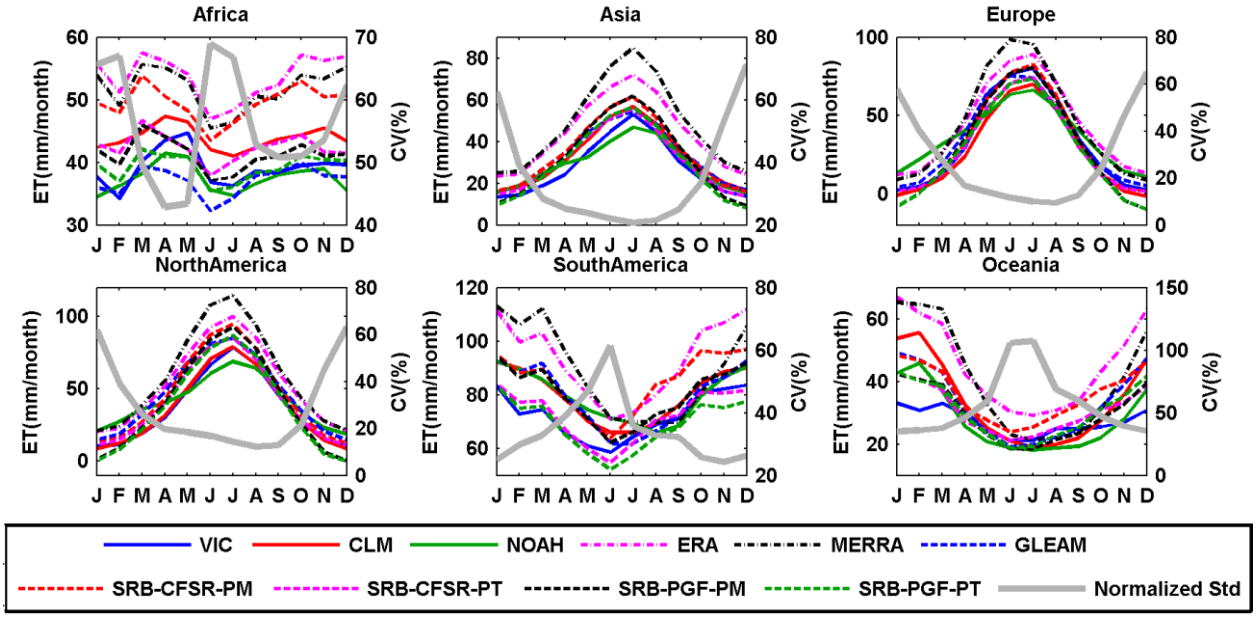

Figure 4. Seasonal cycles of evapotranspiration from different products over the six continents for 1984-2007 (Greenland is excluded for North America, the same for Figures 6 and 8)

5

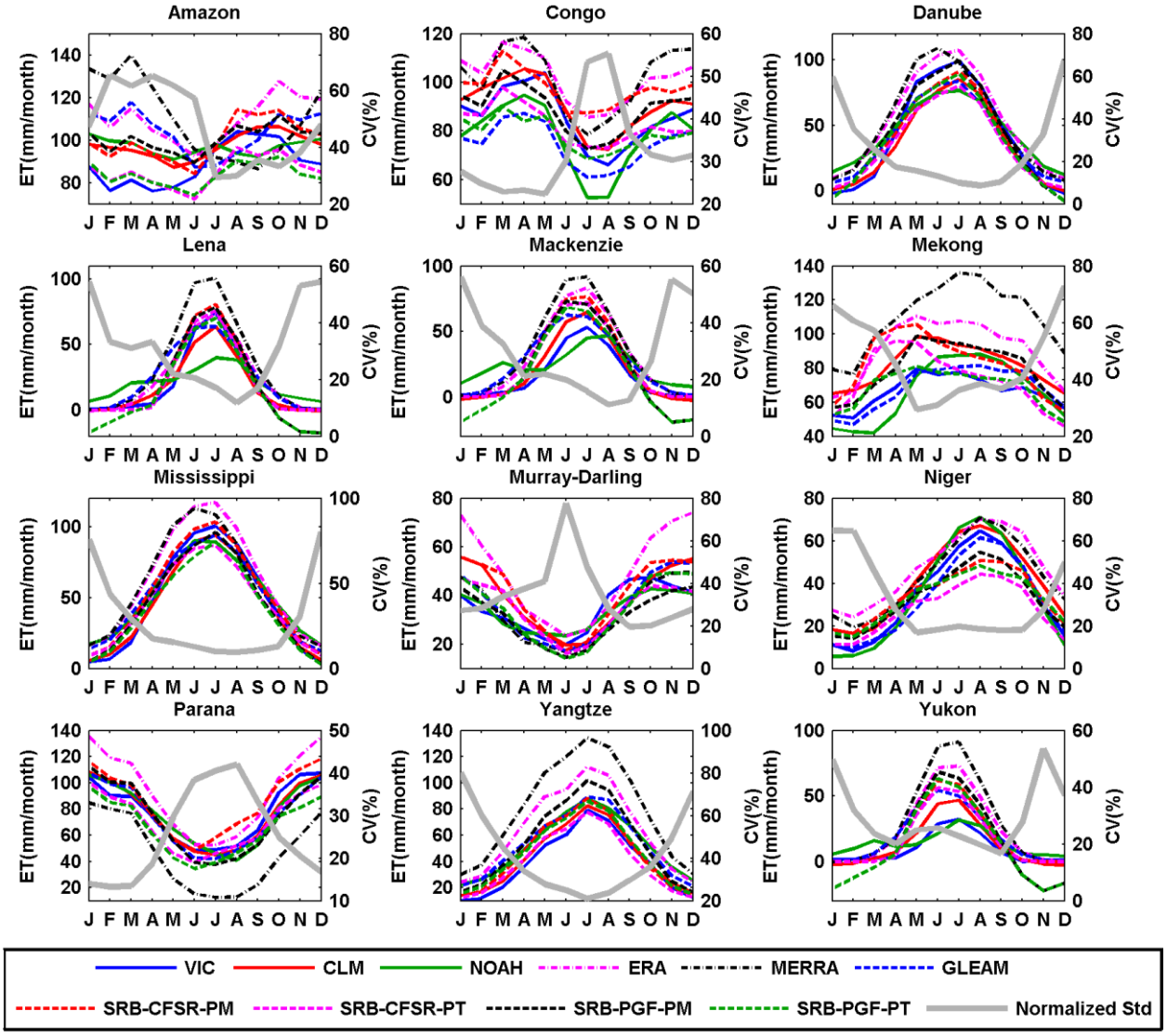

Figure 5. Seasonal cycles of evapotranspiration from different products over twelve representative large basins for 1984-2007

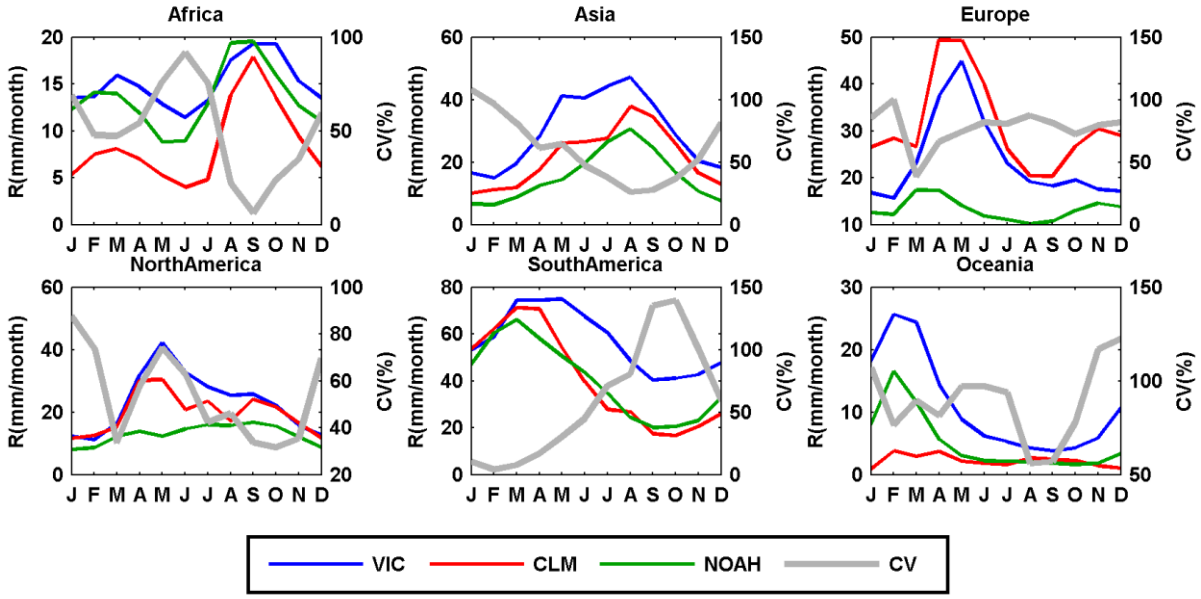

Figure 6. Seasonal cycles of runoff from different products over the six continents for 1984-2010

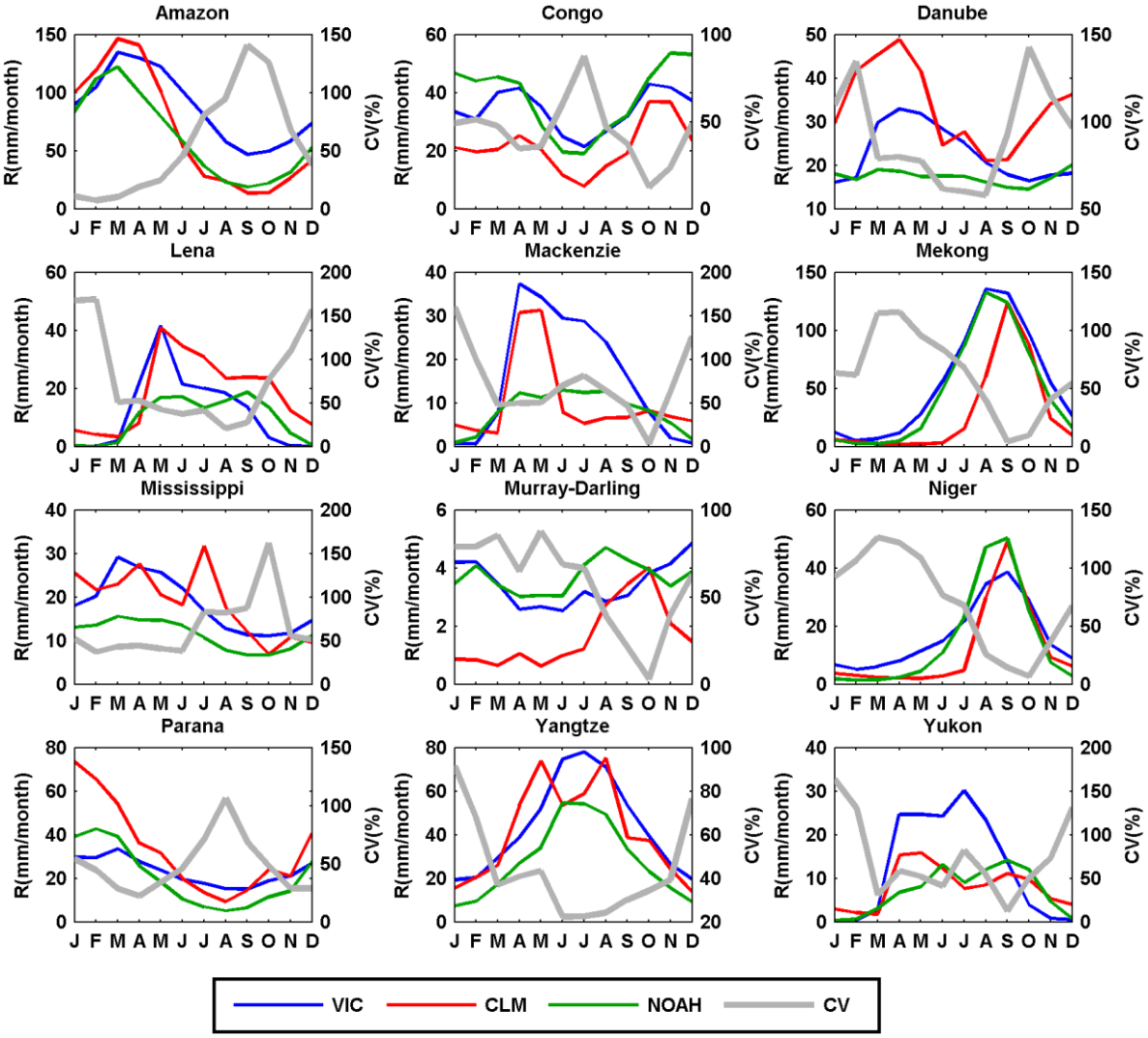

Figure 7. Seasonal cycles of runoff from different products over twelve representative large basins for 1984-2010

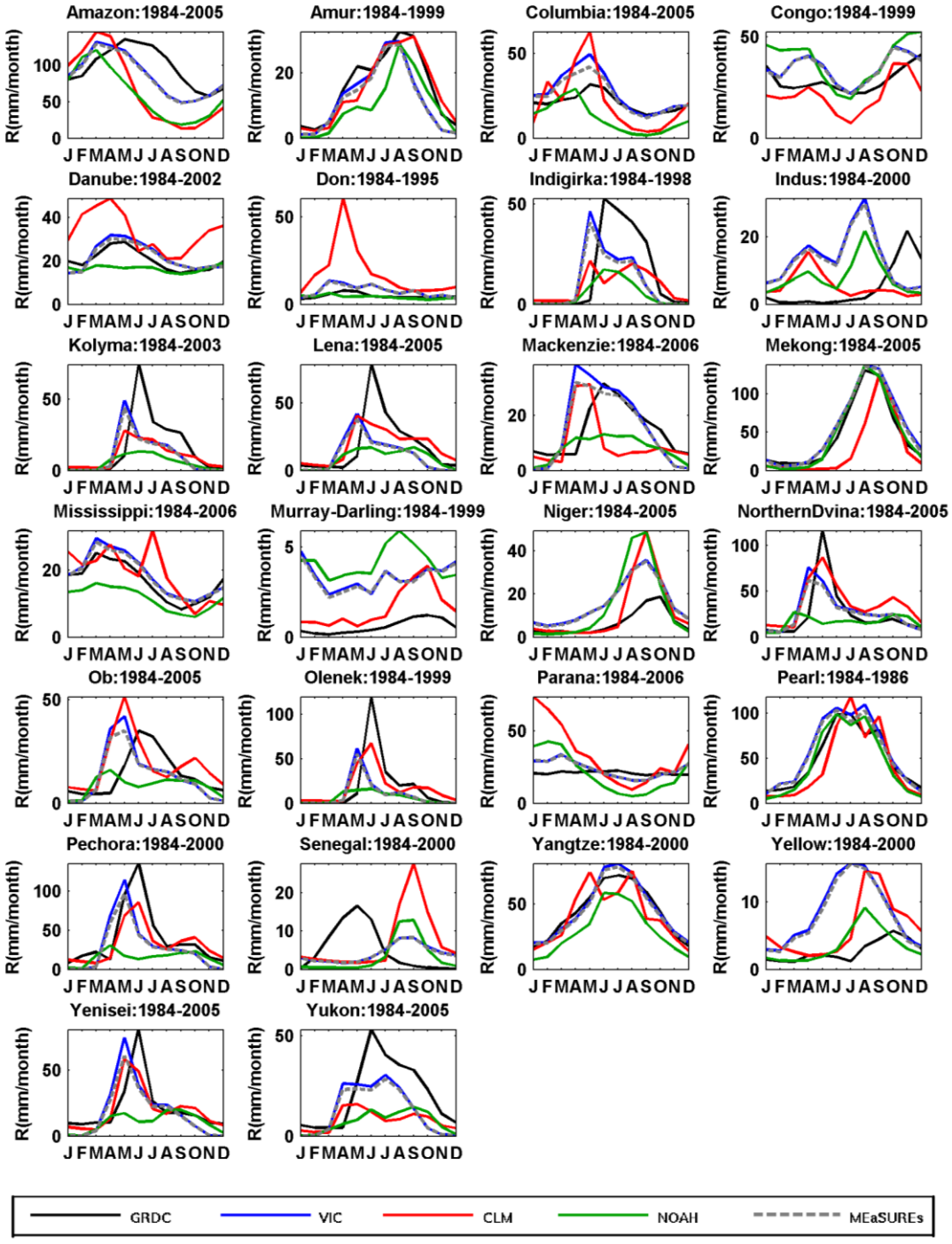

Figure 8 Seasonal Cycles of runoff from VIC, CLM, NOAH and MEaSUREs against GRDC runoff observation over 26 large basins for different periods according to in situ data availability

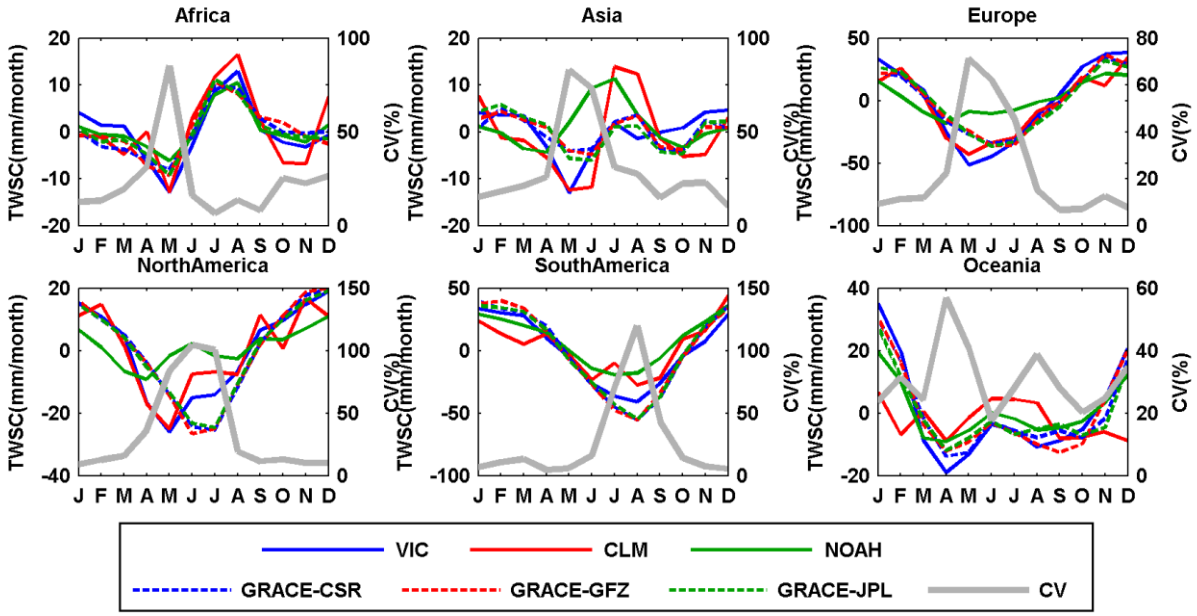

Figure 9. Seasonal cycles of total water storage change (TWSC) from different products over the six continents for 2003-2010 (*TWSC* is first normalized and then the CV (%) is calculated (The same for Figure 9))

5

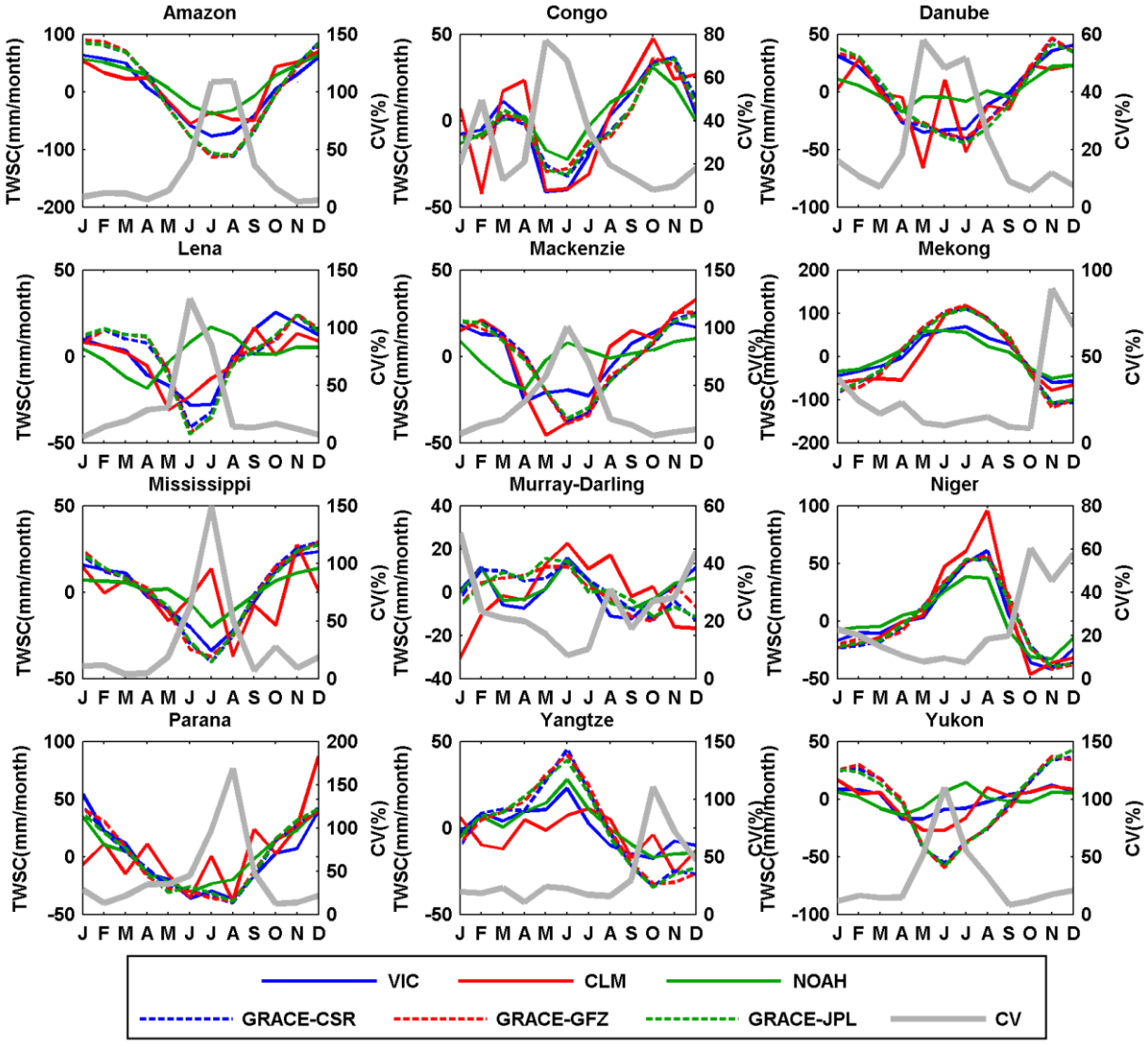

Figure 10. Seasonal cycle of total water storage change (TWSC) from different products over twelve representative large basins for 2003-2010

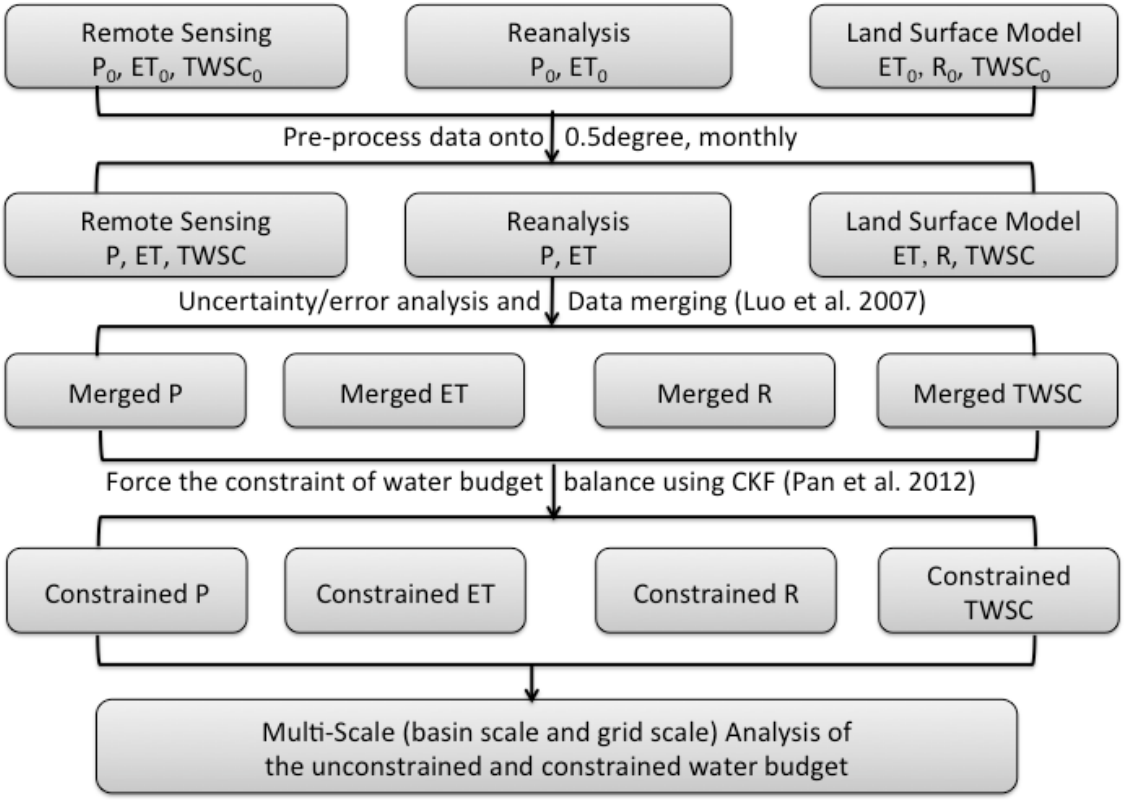

Figure 11. Flowchart

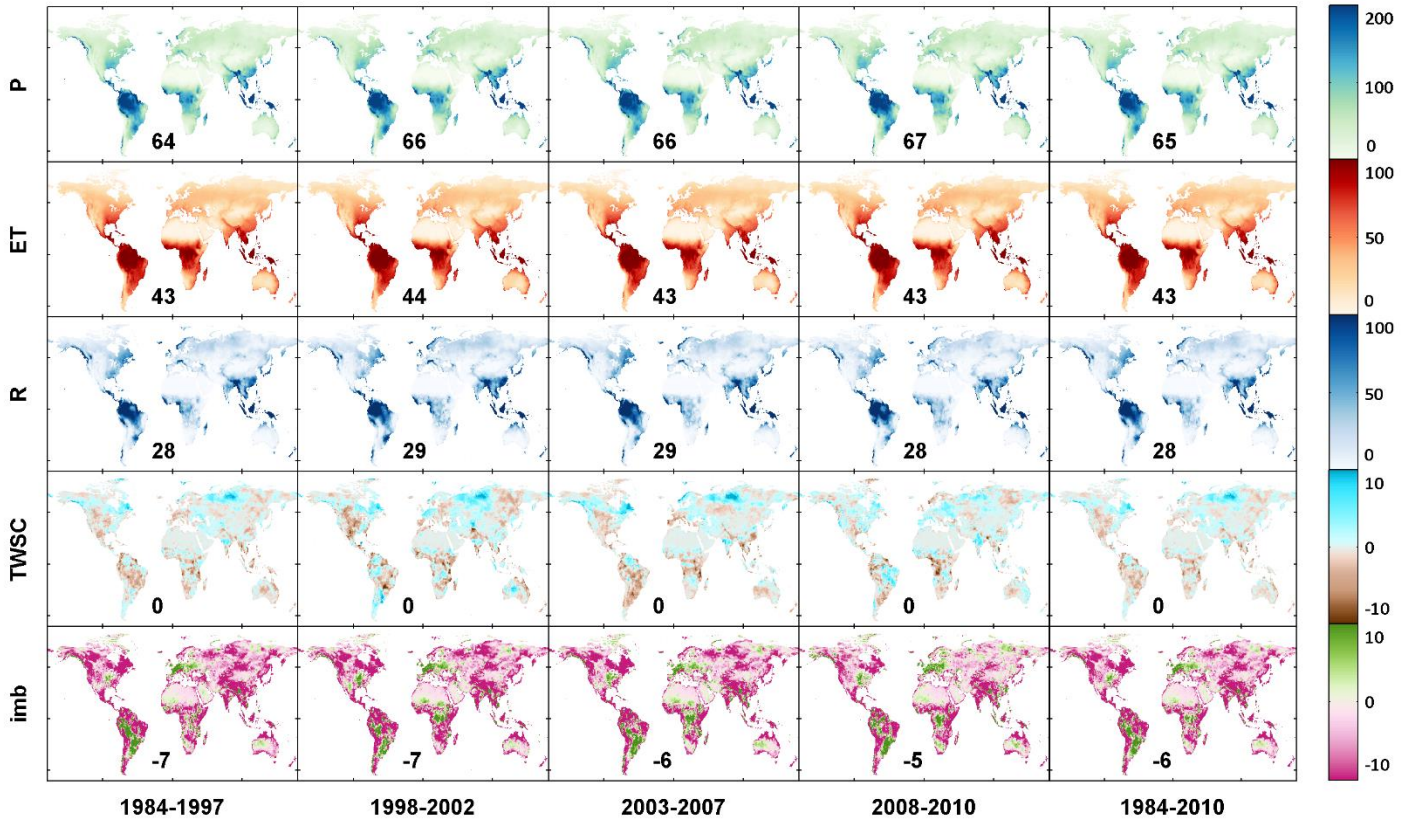

Figure 12. Monthly mean (mm/month) of different water budget terms (from the first row to the bottom: precipitation, evapotranspiration, runoff, total water storage change, imbalance) before water balance constraint throughout different periods (from the left to the right: 1984-1997, 1998-2002, 2003-2007, 2008-2010, and 1984-2010; the numbers listed on each sub-panel are the monthly mean value for each merged water budget variable before water balance constraint (mm/month) during different sub-periods (Greenland and Antarctica excluded). So as in Figures S4 and S5 in Supplementary II, but for the merged water budget variable after water balance constraint (mm/month), and the non-closure error attributions to each water budget variable (%))

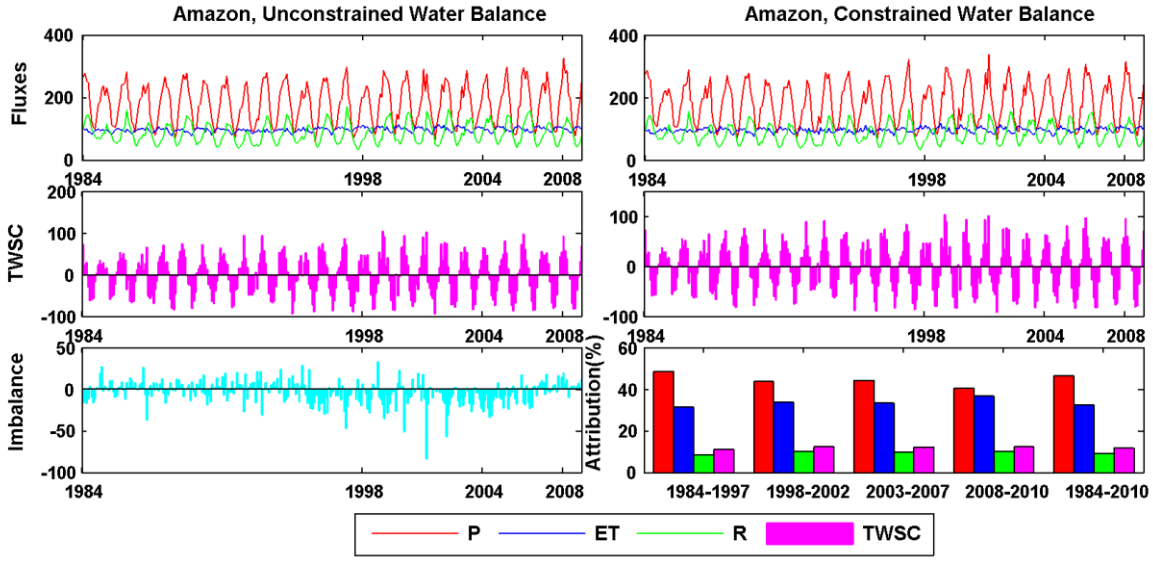

Figure 13. Unconstrained (left column) and constrained (right column) water budget estimates (mm/month) over Amazon River basin. The top, middle and bottom rows show the time series of water budget in terms of fluxes (precipitation, evapotranspiration and runoff), *TWSC* (Total Water Storage Change) and Imbalance. The imbalance/non-closure error after water budget constraint equals to zero and the imbalance/non-closure attributions to each water budget variables throughout different sub-periods are shown at the right bottom.

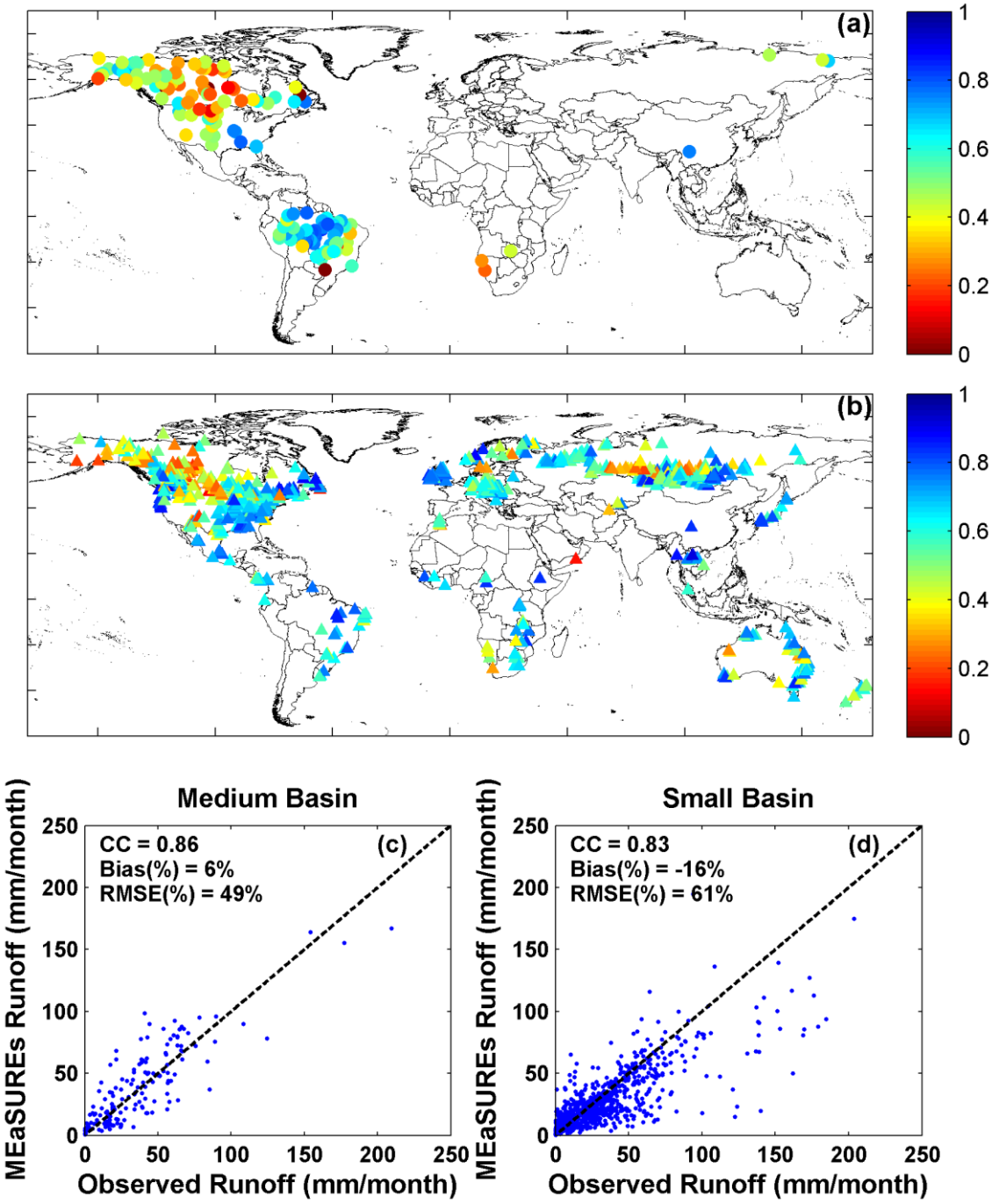

Figure 14 (a) Correlation Coefficient (CC) between monthly GRDC runoff observations and MEaSUREs runoff estimates for 165 medium basins; (b) same as (a), but for 862 small basins; (c) Monthly mean of MEaSUREs runoff estimates against GRDC runoff observations for medium basins; (d) same as (c), but for small basins

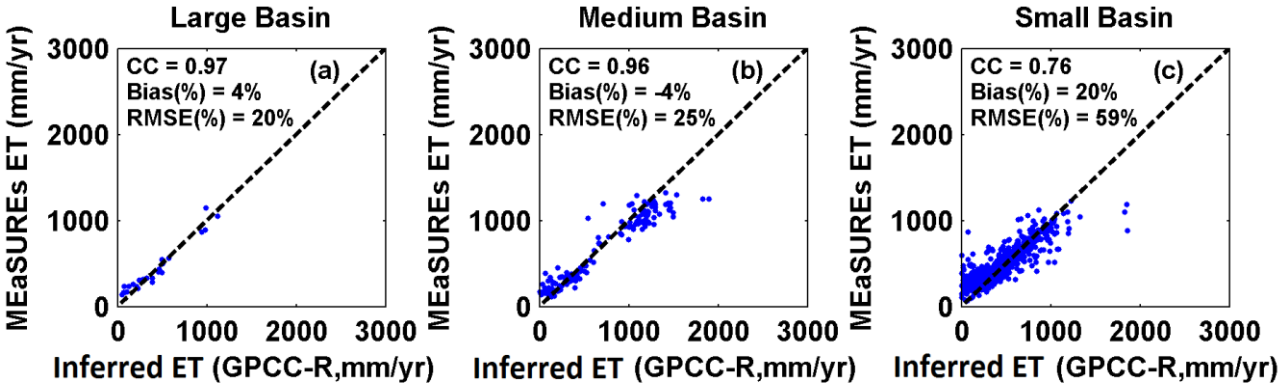

Figure 15. Validation of MEaSUREs ET estimates against Inferred ET ($P - R$) over large (25), medium (169) and small (813) basins

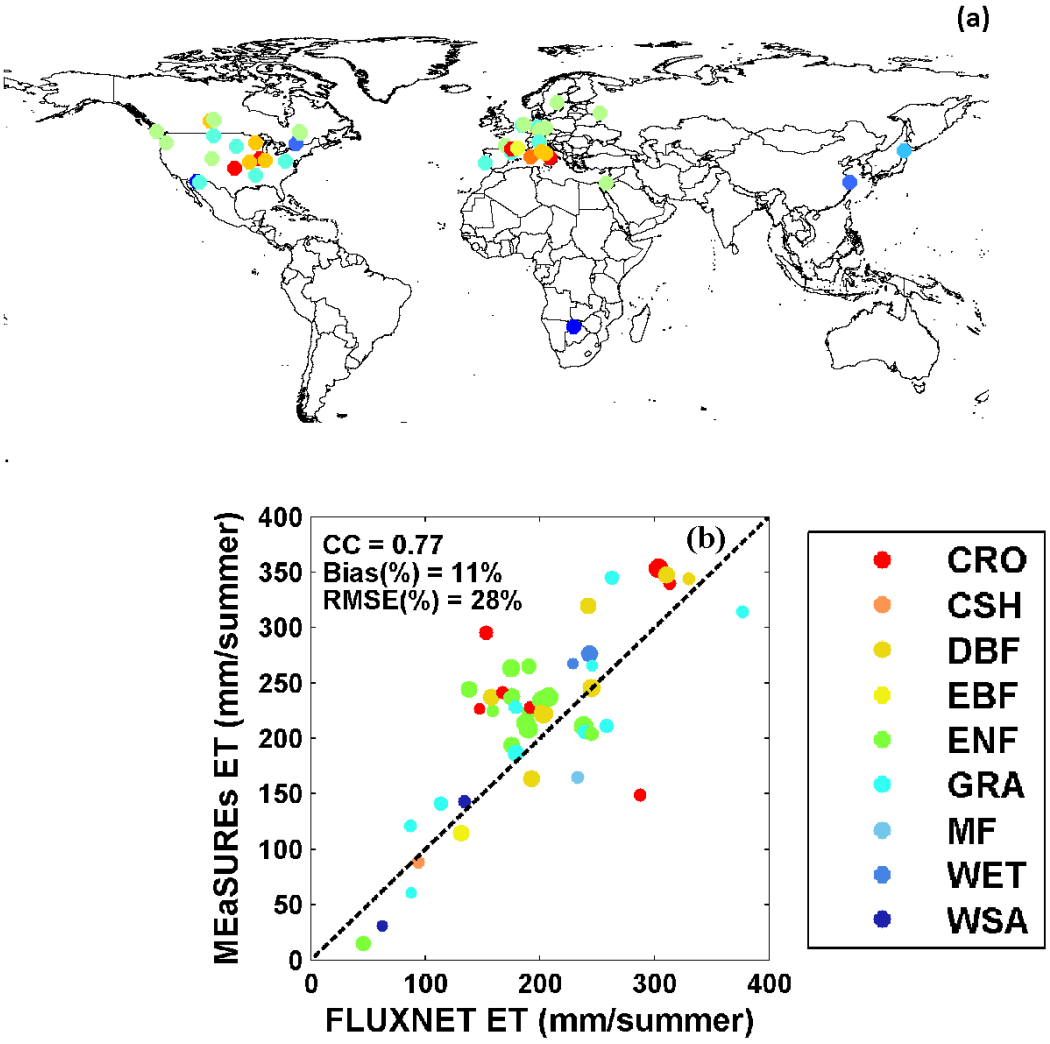

Figure 16. (a) Distributions of 47 flux towers over the globe (b) Validation of MEaSUREs ET estimates against FLUXNET observations (Different colors in Figure 16(b) represent different International Geosphere-Biosphere International Program (IGBP) land cover types (Loveland et al. 2000); the sizes of dots represents the data record length (ranging from 1 to 10 years) from the FluxNet; IGBP land cover types: (1)CRO: Cropland; (2) CSH: Closed Shrublands; (3) DBF: Deciduous Broadleaf Forests; (4) EBF: Evergreen Broadleaf Forest; (5) ENF: Evergreen Needleleaf Forest; (6) GRA: Grassland; (7) MF: Mixed Forest; (8) WET: Permanent Wetland; (9) WSA: Woody Savannas)