Remote Sensing of Environment 252 (2021) 112191

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

Remotely sensed ensembles of the terrestrial water budget over major global river basins: An assessment of three closure techniques

R. Abolafia-Rosenzweig [a,*], M. Pan [c], J.L. Zeng [a], B. Livneh [a,b]

[a] Department of Civil, Environmental, and Architectural Engineering, University of Colorado Boulder, Boulder, CO 80309, USA
[b] Cooperative Institute for Research in Environmental Science (CIRES), University of Colorado Boulder, Boulder, CO 80309, USA
[c] Department of Civil and Environmental Engineering, Princeton University, Princeton, NJ 08544, USA

ARTICLE INFO

Keywords:
Terrestrial water budget
Remote sensing
Global hydrology
Water cycle
Precipitation
Evapotranspiration
GRACE

ABSTRACT

Remote sensing is a useful tool for observing the water cycle. However, combining remote sensing products over any major river basin will result in a residual error in the overall water balance. Previous studies have either quantified this error without correcting it, or have merged observations together with land surface models (LSMs) to produce a single "best" estimate of the water balance. Here, we present a new approach in which combinations of remote sensing and in situ observations are constrained to enforce water balance closure. Rather than a single estimate, this produces an ensemble of unique water balance estimates intended to characterize uncertainty and to avoid biases implicit in LSMs. We evaluate three techniques of varying complexity to enforce water balance closure for individual ensemble members over 24 global basins from Oct. 2002 - Dec. 2014, resulting in as many as 60 realizations of the monthly water budget, contingent upon data availability. Compared with a published climate data record, the ensemble shows strong agreement for precipitation, evapotranspiration and changes in storage (R^2: 0.91–0.95), with less agreement for streamflow (R^2: 0.42–0.47), which may be indicative of LSM biases in the climate data record. Water balance residual errors resulting from combinations of raw products vary significantly ($p < 0.001$) with latitude, with a tendency for positive biases for low- and mid-latitude basins, and negative biases elsewhere. Overall, residual errors are equivalent to 15% of total precipitation when averaged across all data products and basins. This study shows that closure constraints provide additional value outside of closing the water budget, including reduction of uncertainty and transfer of closure constraints in time to provide skillful estimates of mean annual basin discharge. We also showed that a simpler closure technique, proportional redistribution, performed better than more complex ones in decreasing uncertainty and for transfer through time to estimate basin discharge when a rigorous analysis of errors for each data product is not accounted for. This observation-based dataset is distinct from modeled estimates and therefore has the potential to preserve important information of anthropogenic effects on the water balance.

1. Introduction

The adequacy of freshwater resources depends upon knowledge of fluctuations in the water cycle. Yet, the terrestrial water balance is increasingly impacted by human activities both indirectly, e.g. through climate change, and directly, e.g. through irrigation and development in support of economic, political and social systems (Vörösmarty and Sahagian 2000; Rosenzweig et al. 2008; Rodell et al. 2015). Thus, understanding human impacts on the water budget is necessary for strategic planning (NSIT 2007; Scanlon et al. 2012; Ranger et al. 2011).

The terrestrial water budget can be expressed as the balance between precipitation, evapotranspiration, streamflow and the change in terrestrial water storage. Prior assessments of the global historical water budget have combined remote sensing with land surface models (LSMs) that implicitly close the water budget (Zhang et al. 2018; Pan et al. 2012; Livneh and Lettenmaier 2012) to provide a single best estimate of the water cycle. However, LSM-based realizations of the water budget are reliant on parameters that require estimation (Sheffield and Wood 2007; Xia et al., 2012) and most of these models do not explicitly account for human activities. Conversely, combinations of products that independently observe components of the water budget are capable of detecting human impacts on the water budget, but result in an imbalance of the

* Corresponding author.
E-mail address: Ronnie.AbolafiaRosenzweig@colorado.edu (R. Abolafia-Rosenzweig).

https://doi.org/10.1016/j.rse.2020.112191
Received 20 April 2020; Received in revised form 5 November 2020; Accepted 9 November 2020
Available online 19 November 2020

water budget—or *water budget residual*—due to differing assumptions, algorithms, and signal processing uncertainties (Aires 2014; Long et al. 2014; Pan et al. 2012).

Existing efforts to apply exclusively remotely sensed retrievals to close the water budget have done so by inferring Q as the residual of remotely sensed products, "$Q = P\text{-}ET\text{-}\Delta S$", and then evaluating this estimate against observed and modeled Q (Sheffield et al. 2009; Gao et al. 2010). These analyses concluded that although some combinations of remote sensing satellites can nearly close the water budget, these combinations of products are inconsistent in space and time. Stated alternatively, there is no single-best combination of satellite products that can reliably force water budget closure. Another important conclusion has been that the magnitude of uncertainties across combinations of remotely sensed budget components are generally larger than the magnitude of Q itself (Sheffield et al. 2009; Gao et al. 2010). For example, Sheffield et al. (2009) found the uncertainty of inferred Q ranged between 0.4 and 2.9 mm/day, while the commensurate observed and model-simulated Q were generally around 1 mm/day.

Acknowledging the limitations of water budget imbalance, several attempts have been made to create a single best-estimate of a closed water budget (Aires 2014; Sahoo et al. 2011; Pan et al. 2012; Rodell et al. 2015; Zhang et al. 2018). These studies calculated best-estimates for each water budget component by merging multiple remotely sensed, gaged and modeled estimates through a weighted average. Weights were assigned proportional to the error variance of each product, such that products with higher error variance received less weight. After deriving a best estimate for P, ET, Q and ΔS, the total water budget residual was redistributed across the water budget components using a Kalman-type closure constraint—based on the Kalman filter data assimilation technique (Pan and Wood 2006; Evensen 1994)— to ensure mass balance. Following this approach, combining remotely sensed P, ET, ΔS with gaged Q on a monthly time scale required redistributing residual errors of 4.3% of mean annual precipitation at global scales (Rodell et al. 2015) or as high as 5%–25% for selected major river basins (Sahoo et al. 2011) with differences between results from these studies attributable to spatial and temporal resolutions. Pan et al. (2012) applied this methodology over 32 large river basins, combining remotely sensed retrievals, in situ observations, global re-analyses, and LSM simulations to create water budget estimates from 1984 to 2006. Zhang et al. (2018) extended the Pan et al. (2012) methodology over the entire global land surface, creating a climate data record (CDR) of the terrestrial water budget from 1984 to 2010.

This study is motivated by the idea that every remote sensing product contains unique and valuable information about the global terrestrial water budget and may be useful in accounting for human impacts. We build on previous work on estimating the large-scale terrestrial water cycle from satellite remote sensing (Gao et al. 2010; Sheffield et al. 2009; Sahoo et al. 2011; Rodell et al. 2015) by presenting a new REmotEly Sensed ENsemble of the water cycle (REESEN) that probabilistically resolves the water budget, instead of deterministically as done in prior studies that solely reported best-estimates. This approach follows the ensemble philosophy that no-single "best-estimate" exists (Epstein 1969; Leith 1974; Evensen 1994), and instead, seeks to learn about uncertainties through constructing multiple realizations that group different combinations of remote sensing and in situ observations together. The objectives of this manuscript are to: (i) quantify residual errors from remotely sensed realizations of the terrestrial water budget (Sections 3.1, 4.1 and 5.1); (ii) develop and describe the REESEN methodology (Section 3.2); (iii) evaluate three closure techniques used to create REESEN ensembles (Sections 3.3–3.5, 4.2–4.4, and 5.2); and (iv) summarize closure constraints applied to a suite of satellite-based data products (Sections 3.6, 4.5 and 5.3).

2. Study area and materials

Remote sensing of P, ET, S and gaged Q are applied to close the water budget over 24 globally distributed basins that represent a range of climates and anthropogenic influences (Fig. 1). Watershed boundaries, sourced from the USGS HydroSHEDS (Hydrological Data and Maps Based on Shuttle Elevation Derivatives at Multiple Scales; Lehner et al., 2006) and the USGS watershed boundary dataset (https://water.usgs.gov/GIS/huc.html), were chosen on the basis of size, availability of Q observations, and inclusion in prior water budget assessments (Sahoo et al. 2011; Pan et al. 2012; Zhang et al. 2018).

REESEN ensembles are built from widely used products derived entirely or partially from remote sensing products (Table 1) that have been evaluated in prior research through inter-comparison (Guo et al. 2015; Zhang et al. 2018; Khan et al., 2018; Sakumura et al., 2014; Miralles et al., 2015) and ground-truthing (Moazami et al. 2013; Beck et al. 2017b; Chen et al. 2014; Fisher et al. 2008; Miralles et al. 2011; Mu et al. 2011). For consistency, all remote sensing-based datasets are remapped to a 0.5° spatial grid following a conservative spatial aggregation (Schulzweida 2019) and a monthly time step. The space/time resolution of the REESEN analysis is contingent upon availability of key satellite products, like the GRACE product used for ΔS which is available at a monthly timescale with relatively coarse resolution appropriate only for large river basin systems. Moreover, in situ Q data are available at basin scale and the monthly time scale limits the impact of Q travel time. The list of basins and stream gage locations are provided in Table S1. The analysis is constrained to places and times where data products are available for overlapping time intervals within Oct. 2002-Dec. 2014.

There is no spatial coverage beyond 50°N to 50°S for the CHIRPS or TRMM-3B43 products, or beyond 60°N to 60°S for the PERSIANN product. Therefore, basins extending outside these limits do not use respective data products and result in fewer REESEN combinations. Datasets are described in more detail below with a complete set of figures of seasonal cycles for all 24 study basins for P, ET, ΔS, and Q products provided in Figs. S1-S4, respectively.

2.1. Precipitation (P)

P products used in this analysis predominately rely on microwave (Microwave imager on TRMM, Special Sensor Microwave Imager on Defense Meteorological Satellites, Advance Microwave Scanning Radiometer-Earth on *Aqua*, Advance Microwave Sounding Unit-B on NOAA satellites, Advanced Scatterometer) and infrared (GMS-5, GOES 8–15, Meteosat-5, Meteosat-7) satellite sensors and ground-based observations (Global Historical Climate Network version 2, Global Summary Of the Day dataset, Global Precipitation Climatology Project, Global Telecommunications Network, CRU TS2.0, Climate Prediction Center Unified Version 1.0) (Peterson and Vose 1997; Janowiak et al. 2001; Knapp et al. 2011; Durre et al. 2010; Rudolf 1993; Xie et al. 1996; Huffman et al. 1997; Huffman et al. 2001; Iguchi et al. 2009; Brocca et al., 2016, Xie et al. 2007; Chen et al. 2008; Brocca et al. 2016). However, several products also use information from reanalysis and coupled models. Specifically, the PGF relies largely on NCEP/NCAR reanalysis, CHIRPS uses atmospheric rainfall fields from the NOAA Climate Forecast System (CFS) reanalysis product, and MSWEP includes several reanalysis products.

2.2. Evapotranspiration (ET)

ET is primarily controlled by available energy, leaf area, soil moisture conditions, and evaporative demand of the atmosphere. The MOD16A2 product is based on the Penman-Monteith (PM) equation (Monteith 1965); whereas GLEAM and PT-JPL ET are based on the Priestley and Taylor (PT) formulation. Both PM and PT estimate available energy using remotely sensed net radiation and air temperature along with auxiliary variables that differ by product (e.g. albedo and skin temperature). GLEAM uses microwave vegetation optical depth as a proxy for vegetation density, whereas PT-JPL and MOD16A2 both

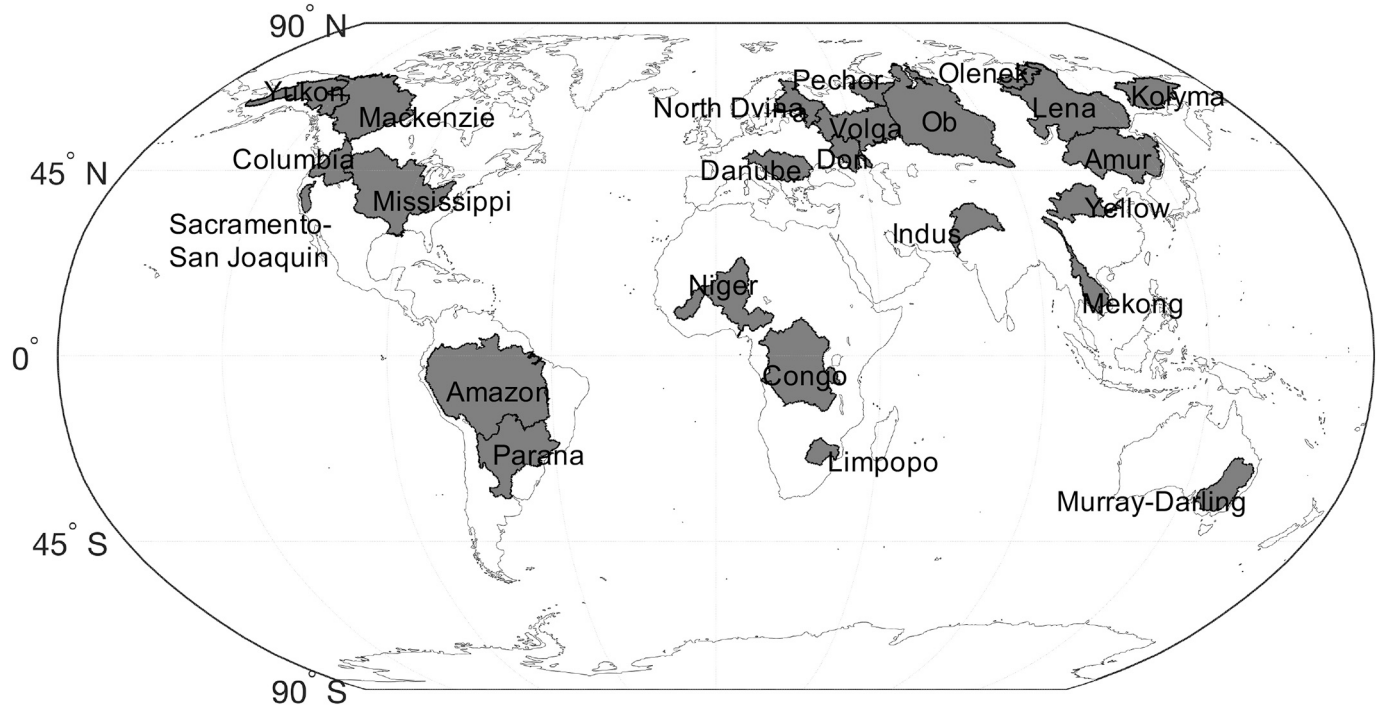

Fig. 1. Twenty-four basins selected for the water budget analysis. Basins are chosen to represent a range of climates (mean annual precipitation ranging from 378 to 2142 mm/year and mean temperatures ranging from −12 °C to 28 °C) and a range of anthropogenic impacts (area equipped for irrigation ranging from 0 to 33% of total basin area; Siebert et al. 2013; Siebert et al. 2015).

explicitly calculate vegetation cover fraction. All products calculate *ET* as the summation of soil evaporation, interception evaporation, and transpiration and assume a tradeoff between canopy evaporation and transpiration where canopy evaporation increases, and transpiration decreases as a function of canopy ponding. GLEAM explicitly calculates interception with an interception model, while PT-JPL and MOD16A2 use the fraction of saturated soil based on relative humidity to estimate interception. GLEAM employs a soil water module to quantify soil moisture controls on *ET*, whereas PT-JPL and MOD16A2 rely on vapor pressure deficit and relative humidity as proxies for soil moisture conditions.

2.3. Terrestrial water storage change (ΔS)

GRACE is the only remote sensing platform that observes ΔS and is derived from observations of Earth's time-varying gravity field (Tapley 2004). The estimated total error in a single monthly solution is 2.5 cm for an 800 km radius (Rodell 2004; Wahr et al. 2004). Errors in estimated ΔS include instrument and signal retrieval errors, errors from the removal of atmospheric mass variations using model analyses, and leakage errors from surrounding areas (Mu et al. 2017). Variance among different GRACE products (Table 1) all lie within the error of the GRACE data and show that no significant biases exist between solutions (Sakumura et al. 2014). GRACE ΔS comprises all water on and near the surface of the land, including groundwater, soil moisture, surface water, snow and ice, and biological water (Rodell et al. 2015). GRACE provides monthly mean anomalies of *S*. We use corresponding GRACE multiplicative scaling coefficients aimed at restoring some of the signal loss due to filtering and truncation of the original GRACE spherical harmonics used in the derivation of GRACE *S* anomaly observations (Landerer and Swenson 2012; Kumar et al. 2016). Following the approach of Rodell et al. (2015), daily *S* anomalies are estimated by linearly interpolating the monthly data to daily timesteps which are used to compute monthly ΔS by differencing *S* estimates from the start and end of a given month.

2.4. Streamflow (Q)

Among the observational datasets used in this study, *Q* is the only in situ observation. Stream gage observations were sourced from: (i) the Global Streamflow Data Center (GRDC https://www.bafg.de/GRDC/EN/Home/homepage_node.html), (ii) the Dai and Trenberth (DT) Global River Flow and Continental Discharge Dataset (Dai 2017; https://rda.ucar.edu/datasets/ds551.0/), (iii) the Dartmouth Flood Observatory (https://www.dartmouth.edu/~floods/), and (iv) the California Data Exchange Center (https://cdec.water.ca.gov/). *Q* time series were constructed using the following criteria:

1. Data are first sourced from the GRDC, as this was used in prior studies (Zhang et al. 2018; Sahoo et al. 2011).
2. Any gaps in GRDC data are filled with DT data during missing intervals.
3. For basins without GRDC or DT data (Yellow, Indus, Limpopo), time series were created from Dartmouth Flood Observatory data.
4. The California Data Exchange center provided data for the Sacramento-San Joaquin basin system.

An estimate of *Q* uncertainty is needed for two of the water balance residual redistribution algorithms (Constrained Kalman Filter and Multiple Collocation described in Section 3.2). However, *Q* has only a single observation, from an in situ gage, unlike the remotely sensed water balance components which have observations from multiple products. Gage uncertainty for *Q* has been published as a percent error for some of the study basins, ranging from 2.3%–28.8% (Clarke 1999; Carter and Anderson 1963; Shiklomanov et al. 2006; Mueller, 2003). However, the required uncertainty information such as the stage-discharge relationship, or gage type (e.g. velocity or stage measurements), is not available for most gages used in this study. Sahoo et al. (2011) addressed this issue by assigning a 7% RMS error to all in-situ measurements (Dingman 2002). Here, we assign uncertainty on the basis of *Q* magnitude following a large-sample observational study

Table 1
Datasets used in this study derived entirely or partially from remote sensing products, including spatial and temporal resolutions, as well as the period of record.

Dataset	Spatial resolution	Temporal resolution	Period of record	Reference and data source
Precipitation (P)				
CHIRPS-2.0	0.05°	monthly	1981-present	Funk et al. (2014) (ftp://ftp.chg.ucsb.edu/pub/org/chg/products/CHIRPS-2.0/global_monthly/)
TRMM-3B43	0.25°	monthly	1998-present	Huffman et al. (2007) (https://disc2.gesdisc.eosdis.nasa.gov/opendap/TRMM_L3/TRMM_3B43.7/)
PGF	0.5°	monthly	1948–2014	Sheffield et al. (2006) (http://hydrology.princeton.edu/data/pgf/)
PERSIANN	0.25°	daily	1995-present	Hsu et al. (1997) (https://www.ncei.noaa.gov/data/precipitation-persiann/access/)
MSWEP V2	0.1°	3-h	1979–2017	Beck et al. (2017a); Beck et al. (2019) (https://platform.princetonclimate.com/PCA_Platform/mswepRetro.html)
Evapotranspiration (ET)				
GLEAM	0.25°	daily	1980–2016	Miralles et al. (2011) & Martens et al. (2017) (https://www.gleam.eu/)
PT-JPL	36 km	monthly	2002–2017	Fisher et al. (2008) & Purdy et al. (2018) (http://josh.yosh.org/datamodels.htm)
MOD16A2	1 km	monthly	2000-Present	Mu et al. (2007); Mu et al. (2009); Mu et al. (2011) (http://files.ntsg.umt.edu/data/NTSG_Products/MOD16/MOD16A2_MONTHLY.MERRA_GMAO_1kmALB/)
Change in terrestrial water storage (ΔS)				
GRACE Mascons	0.5°	monthly	2002–2017	Wiese et al. (2016); Wiese et al. (2018) & Watkins et al. (2015) (https://grace.jpl.nasa.gov/data/get-data/jpl_global_mascons/)
GRACE CSR	1.0°	monthly	2002–2016	Swenson and Wahr (2006); Swenson (2012) and Landerer and Swenson (2012) (https://grace.jpl.nasa.gov/data/get-data/monthly-mass-grids-land/)
GRACE GFZ	1.0°	monthly	2002–2016	Swenson and Wahr (2006); Swenson (2012) and Landerer and Swenson (2012) (https://grace.jpl.nasa.gov/data/get-data/monthly-mass-grids-land/)
GRACE JPL	1.0°	monthly	2002–2016	Swenson and Wahr (2006); Swenson (2012) and Landerer and Swenson (2012) (https://grace.jpl.nasa.gov/data/get-data/monthly-mass-grids-land/)

(Shiklomanov et al. 2006), which reported decreases in uncertainty (percent-error) with increasing Q magnitude, i.e. lower flows generally have higher uncertainty. We assign the lowest flow in a basin time series a mean uncertainty of 28.8% and linearly interpolate the uncertainty downwards to 2.3% for highest flow in a time series. We acknowledge that this assumption does not universally hold across all sites, but represents a starting-point, with additional inquiry into specific-basin uncertainty representing an avenue for further investigation beyond the scope this study. Finally, Q-uncertainty is constrained to be smaller than remotely sensed component uncertainties, because we assume that in situ, observations will necessarily have smaller uncertainties than remotely sensed retrievals.

3. Methods

3.1. Residual error from remotely sensed terrestrial water budget realizations

The water budget is commonly characterized by Eq. 1, which states that changes in terrestrial water storage (ΔS) are equal to the difference between incoming precipitation (P) with outgoing evapotranspiration (ET) and streamflow (Q). Note that P includes water in both liquid and solid states, and ΔS includes all stores of water at or near the land surface, as soil water, groundwater, in lakes, reservoirs, wetlands, stream channels, and in glaciers and ice caps (Gao et al. 2010).

$$\Delta S = P - ET - Q \quad (1)$$

Combinations of products that independently observe components of the water budget result in an imbalance of the water budget—or *water budget residual* (r from Eq. 2)—due to differing assumptions, algorithms, and signal processing uncertainties (Long et al. 2014; Pan et al. 2012).

$$r = P - ET - Q - \Delta S \quad (2)$$

This study quantifies residual errors from unique combinations of observation-based P, ET, ΔS, and Q over the study area. Spatial variability of residual errors in the remotely sensed water budget are examined relative to six quantities: (i) enhanced vegetation index (EVI) derived from the Global MODIS MOD13A2 product (Didan et al. 2015), (ii) near surface air temperature derived from the Global Land Data Assimilation System Version 2 (GLDAS-2) database (Beaudoing et al., 2020; Rodell et al. 2004), (iii) precipitation, (iv) basin size, (v) percentage of irrigated area derived from the Global Map of Irrigation Areas (GMIA) (Siebert et al. 2013; Siebert et al. 2015), and (vi) geographical latitude, which reflects the interaction of several of the aforementioned phenomena. Precipitation used in this evaluation (Section 4.1) is the mean time series from the five precipitation products in Table 1, noting that these are the same products used to construct REESEN ensembles and evaluate uncertainty.

3.2. REESEN: Remotely sensed ensemble of the terrestrial water budget

REESEN ensembles of the terrestrial water budget are generated through different combinations of products in Table 1. Unique combinations of 5 P datasets, 3 ET datasets, 4 ΔS, and 1 Q dataset results in 60 unique realizations of the water budget for each basin between 50°S-50°N, with less realizations for basins outside of these bounds as a result of products' coverage.

Combining different observation datasets with random and systematic errors results in water budget imbalance (i.e. Eq. 2). To resolve the

imbalance, residual errors are redistributed to produce 'corrected' water budget components through additive closure constraints (ε in Eq. 3 and Fig. 2), enforcing water budget closure. Please note that Eq. 3 is applied across all ensemble members within REESEN to generate an ensemble of the closed water budget.

$$r' = 0 = (P + \varepsilon_P) - (ET + \varepsilon_{ET}) - (\Delta S + \varepsilon_{\Delta S}) - (Q + \varepsilon_Q) \quad (3)$$

Where r' is the corrected water budget residual and ε is the error from water budget components denoted by respective subscripts. Note that closure constraints sum to equal the r from Eq. 2. This approach allows Eq. 3 to estimate an ensemble of error time series for each product in Table 1 and Table S1.

The balance can be characterized a linear constraint with the complete set of water budget variables expressed as a column vector:

$$\mathbf{x} = [P\ ET\ Q\ \Delta S]^{\mathrm{T}} \quad (4)$$

and the balance constraint can be expressed as a linear function of the vector,

$$r' = \mathbf{G}\mathbf{x} = 0 \quad (5)$$

where r' is the 'corrected' water budget residual and $\mathbf{G}$ is expressed as:

$$\mathbf{G} = [1 - 1 - 1 - 1] \quad (6)$$

The problem of enforcing the water budget constraint can be defined by the following. Given an unclosed set of water budget estimates: $\hat{x} = \left[\hat{P}\ \widehat{ET}\ \hat{Q}\ \widehat{\Delta S}\right]^T$ and their error estimates, where $r = \mathbf{G}\hat{x} \neq 0$, we look for a new set of estimates, $\hat{x}' = \left[\hat{P}\ \widehat{ET}'\ \hat{Q}'\ \widehat{\Delta S}'\right]^T$ such that $\mathbf{G}\hat{x}' = 0$ and the expected errors in the new estimate are minimized in a root mean square error (RMSE) sense.

Closure constraints (ε in Eq. 3) are calculated from three techniques ranging from simple redistributions of residual to more complex Constrained Kalman filtering (CKF) (Pan and Wood 2006; Sahoo et al. 2011; Pan et al. 2012; Rodell et al. 2015; Zhang et al. 2018) and multiple-collocation (MCL) (Pan et al. 2015), to understand the impact of algorithm choice on resulting water budget partitioning. The CKF and MCL algorithms assume products estimating the same water budget component are independent. However, for cases where products are not independent, the result of the assumption will be that product errors and closure constraints are likely to be underestimated. This occurs because the CKF and MCL implement a kind of "voting" process where one estimate gets one vote (Pan et al. 2015). If estimates are not independent, then product agreement is not necessarily an indicator of accuracy, but rather may be a biproduct of the non-independence. This effect is discussed in context of ΔS estimates in Section 4.5. A non-negativity constraint is applied to ensure 'corrected' values of P, Q and ET are strictly positive. The absolute value of negative components is redistributed among the remaining budget components using the same redistribution technique.

In this study we choose only to apply bias correction at the water budget level, by forcing closure of the water balance, rather than adjusting individual components of the water budget. For example, previous studies (Pan et al. 2012; Sahoo et al. 2011) bias corrected individual products towards the climatology of a reference dataset (e.g. LSM or in situ) before applying closure constraints. Correcting towards an LSM has been shown to remove human signatures (i.e. irrigation) that are present in native observations (Kumar et al. 2015). Further, most of the basins in this study lack the dense in situ observing networks that would be needed to bias correct these water budget components. In previous studies where a priori bias correction is applied to individual satellite products, residual errors in the overall water balance persisted, suggesting that bias correction towards a zero residual in the overall

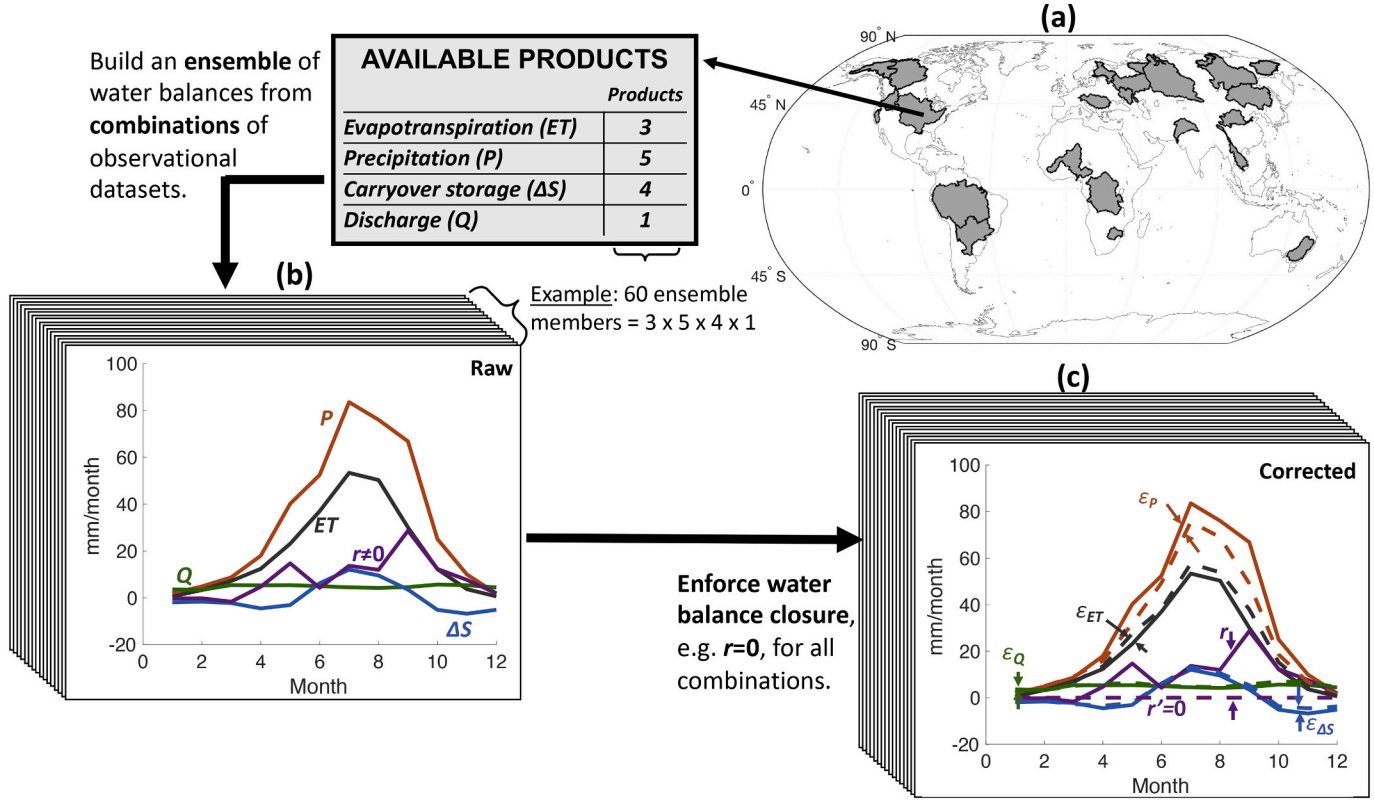

Fig. 2. REESEN Remotely Sensed Ensemble Water Budget Development: (a) For each major river basin (grey areas) **(b)** an ensemble of remotely sensed water balance combinations (example has 60, but this number may be smaller depending on data availability) are **(c)** corrected by enforcing mass balance closure and removing the residual term to be redistributed across components using constraints (Eq. 3).

water balance is a necessary target (Sahoo et al. 2011; Sheffield et al. 2009). Sahoo et al. (2011) found no significant difference between water budget realizations generated by first bias correcting individual budget estimates and then applying closure constraints relative to realizations that withheld a priori bias correction and only corrected towards the zero residual target. Users of REESEN data should be aware of this decision to only correct components towards a zero residual in the overall water balance, which is discussed in context of REESEN results in Section 5.3.

Proportional redistribution (PR)

PR is the most straightforward approach and assumes all components have the same relative error, such that redistribution of the water budget residual (Eq. 3) is proportional to component magnitude. This follows a simple assumption that products' errors are only a function of product magnitude and corresponding residual error in the water budget. Other techniques (CKF and MCL) build on complexity by considering variability among retrievals. Mathematically, $\hat{\mathbf{x}}'$ is calculated as:

$$\hat{\boldsymbol{x}}'^{i} = \hat{\boldsymbol{x}}^{i} - r\left(\boldsymbol{G}^{i}\right)\left(\frac{\widehat{|\boldsymbol{x}^{i}|}}{\sum_{j=1}^{4}\left|\widehat{\boldsymbol{x}^{j}}\right|}\right) \tag{7}$$

and i represents the i$^{\text{th}}$ element in the vector.

Constrained Kalman filter (CKF)

The CKF is implemented here as described in Pan and Wood (2006) and Pan et al. (2012). The error covariances of the state estimate ($\hat{\boldsymbol{x}}$) are referred to by $\boldsymbol{\delta}_{\mathbf{xx}}$ and defined as

$$\delta_{xx} = \overline{(\hat{\boldsymbol{x}} - \boldsymbol{x})(\hat{\boldsymbol{x}} - \boldsymbol{x})^{T}} \tag{8}$$

The symbol without a caret, for example, $\mathbf{x}$, refers to the "true value" of the vector, and the bar over an expression means "expectation" or mean value. Here, $\mathbf{x}$ is calculated as the mean of all products estimating the water balance component in question (Zhang et al. 2018). Then $\boldsymbol{\delta}_{\mathbf{xx}}$ has dimension 4 × 4 which breaks down to error covariances between each individual water budget term:

$$\delta_{\mathbf{xx}} = \begin{bmatrix} \delta_{P-P} & \delta_{P-ET} & \delta_{P-Q} & \delta_{P-\Delta S} \\ \delta_{ET-P} & \delta_{ET-ET} & \delta_{ET-Q} & \delta_{ET-\Delta S} \\ \delta_{Q-P} & \delta_{Q-ET} & \delta_{Q-Q} & \delta_{Q-\Delta S} \\ \delta_{\Delta S-P} & \delta_{\Delta S-ET} & \delta_{\Delta S-Q} & \delta_{\Delta S-\Delta S} \end{bmatrix} \tag{9}$$

Because of the lack of information regarding the correlation of the error of different components/storage terms, all off-diagonal covariance elements are assumed to be zero (Rodell et al. 2015), so $\hat{\boldsymbol{x}}'$ is calculated as:

$$\hat{\boldsymbol{x}}' = \hat{\boldsymbol{x}} + \boldsymbol{K}(0 - \mathbf{G}\hat{\boldsymbol{x}}) \tag{10}$$

where $\boldsymbol{K} = \delta_{\mathbf{xx}}\boldsymbol{G}^{T}(\boldsymbol{G}\delta_{\mathbf{xx}}\boldsymbol{G}^{T})^{-1}$ is called the *Kalman gain*. Given $r = \mathbf{G}\hat{\boldsymbol{x}}$, we have:

$$\hat{\boldsymbol{x}}' = \hat{\boldsymbol{x}} - \delta_{\mathbf{xx}}\,\mathbf{G}^{T}\left(\mathbf{G}\delta_{\mathbf{rr}}\mathbf{G}^{T}\right)^{-1} r \tag{11}$$

where $\boldsymbol{\delta}_{\mathbf{rr}}$ is calculated entry by entry according to Eq. 9.

Multiple collocation (MCL)

The MCL is an extension of the triple collocation method (Stoffelen 1998). We refer the reader to Pan et al. (2015) for a comprehensive treatment of MCL. MCL does not require knowledge of the errors in the input products and instead calculates their error levels from the mutual distance (mean squared distance) among them under the assumption that their errors are uncorrelated with each other. Briefly, the MCL algorithm solves a potentially over-constrained system:

$$\mathrm{A}_{(N)}\mathrm{y}_{(\mathrm{N})} = \mathrm{b}_{(\mathrm{N})} \tag{12}$$

In the triple collocation algorithm $N = 3$, but the MCL algorithm allows N to be any positive integer. For example, when $N = 5$, we have:

$$\boldsymbol{A}_{(5)} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}, \boldsymbol{y}_{(5)} = \begin{bmatrix} d^{2}_{1t} \\ d^{2}_{2t} \\ d^{2}_{3t} \\ d^{2}_{4t} \\ d^{2}_{5t} \end{bmatrix}, \boldsymbol{b}_{(5)} = \begin{bmatrix} d^{2}_{12} \\ d^{2}_{13} \\ d^{2}_{14} \\ d^{2}_{15} \\ d^{2}_{23} \\ d^{2}_{24} \\ d^{2}_{25} \\ d^{2}_{34} \\ d^{2}_{35} \\ d^{2}_{45} \end{bmatrix} \tag{13}$$

Because there is no exact solution for an over-constrained system, MCL seeks to find a best "compromise" using a least squares solution, $\mathrm{y}_{(\mathrm{N})} = [d^{2}_{1t} \ldots d^{2}_{Nt}]^{\mathrm{T}}$ which minimizes Pythagorean distances of all constraints. Here d^{2}_{xy} is the mean squared distance between 2 estimates x and y. For example, d^{2}_{12} is the mean squared distance between product *1* and product *2*. d^{2}_{1t} is the mean squared distance between product *1* and the "true" value *1* is attempting to estimate. Here, the true value is assumed to be the mean of all products estimating a given water budget component (Zhang et al. 2018). Note, the true value may never be known (Pan et al. 2015).

The total water budget residual, r, is redistributed to products proportional to their relative distance from the true value, as calculated through the MCL. For example, a single realization of the water budget at a given time-step can be expressed by Eq. 2, where an estimate of the distances of each component from the true estimate is calculated as $d_{\hat{x}x}$. Then $d_{\hat{x}x}$ has dimension 4 × 1 and is normalized to sum to 1.0 by:

$$d^{i}_{\hat{x}x-norm} = \frac{\left|d^{i}_{\hat{x}x}\right|}{\sum_{j=1}^{4}\left|d^{j}_{\hat{x}x}\right|} \tag{14}$$

$\hat{\boldsymbol{x}}'$ is calculated as:

$$\hat{\boldsymbol{x}}'^{i} = \hat{\boldsymbol{x}}^{i} - r\left(\boldsymbol{G}^{i}\right)\left(d^{i}_{\hat{x}x-norm}\right) \tag{15}$$

Note that the MCL equation (Eq. 15) looks similar to PR (Eq. 7) but it relies on mutual distances among products instead of their magnitudes.

3.3. Comparing REESEN with climate data record (CDR)

The first evaluation step is a direct comparison between REESEN and the CDR (Zhang et al. 2018). The statistical performance measures computed between REESEN and CDR water budget estimates include the percent bias (PBIAS) to identify persistent errors (overprediction vs. underprediction), and the Pearson correlation coefficient to quantify the degree of collinearity between estimated and established water budget estimates (Yue et al. 2002). We also evaluate the 'relative range', calculated by dividing the range of REESEN estimates for a given water budget variable by the reported CDR uncertainty. This analysis follows prior studies (Bosshard et al. 2013; Zhang et al. 2018; Evensen 1994) by using ensemble spread as a proxy of uncertainty for water budget estimates; however, we note that our estimates of uncertainty are sensitive to the products used in this study (Table 1) and thus are based on limited

samples that provide an underestimate of the total uncertainty for the remotely sensed water budget.

3.4. Uncertainty in water budget estimates

The effect of enforcing water balance closure on uncertainty is evaluated using the coefficient of variation (CV), calculated as the REESEN ensemble standard deviation divided by the ensemble mean for each timestep. First, we quantify the mean magnitude of CV for each water budget component while exploring seasonality in the context of CV's relationship with component magnitude. Next, we quantify the effect of closure constraints on uncertainty using the ratio of CV between corrected estimates and raw estimates, herein termed 'relative CV.' A ratio less than unity indicates that REESEN has smaller uncertainty for a given water budget component relative to the raw retrievals. The application of CV to characterize uncertainty follows from the approach taken by Zhang et al. (2018). Since the same CV calculation is applied to both corrected and raw water balance components, we expect the change in CV to be informative of the directionality of changes in uncertainty, essentially assuming that a change in 'ensemble spread' is reflective of uncertainty. However, we acknowledge that relatively small sample sizes will make the CV calculation sensitive to outliers that will overstate the magnitude of changes in uncertainty. Therefore, we primarily focus on the directionality of changes in uncertainty, either increasing or decreasing, rather than their magnitudes. Significance testing for CV's correspondence with product magnitude and changes in CV are computed from nonparametric bootstrap hypothesis tests which are explained in detail in Section S2 of Supplementary Materials.

3.5. Temporal transferability of closure constraints

The temporal transferability of closure constraints is assessed using a jack-knifing cross validation through time following Zelelew and Alfredsen (2014). A complete time series of Q is generated through a drop 1-year transfer approach. The monthly climatology of closure constraints from non-dropped years for P, ET and ΔS are applied as closure constraints to the months in the 'dropped' year. Then the transferability is evaluated through comparing estimated Q, "$Q_{est} = P\text{-}ET\text{-}\Delta S$", relative to observed flows for all 'dropped' years. Results are analyzed relative to the raw residual, i.e. Q_{est} derived from raw retrievals of P, ET and ΔS, as was done by Sheffield et al. (2009) and Gao et al. (2010).

3.6. Summary of closure constraints applied to each data product

Knowledge of whether individual products generally receive larger or smaller closure constraints could potentially provide information on product accuracy. However, since 'true' values are not observable, it is impossible to discern whether inconsistencies in one product relative to others are truly indicative of data quality issues. A summary of closure constraints applied to data products is presented based on REESEN ensembles for the 12 basins covered by all remotely sensed products (south of 50°N). The summary includes the mean closure constraint applied to each product along with the interquartile range of closure constraints, based on the ensemble spread, which represents the sensitivity of a product's corrections to its combinations with other products in a REESEN.

4. Results

4.1. Residual error from remotely sensed terrestrial water budget realizations

Mean water budget residuals across the three REESEN ensembles for each study basin are presented in Fig. 3. Residual errors sum to 15% of observed precipitation when averaged across all data products. These

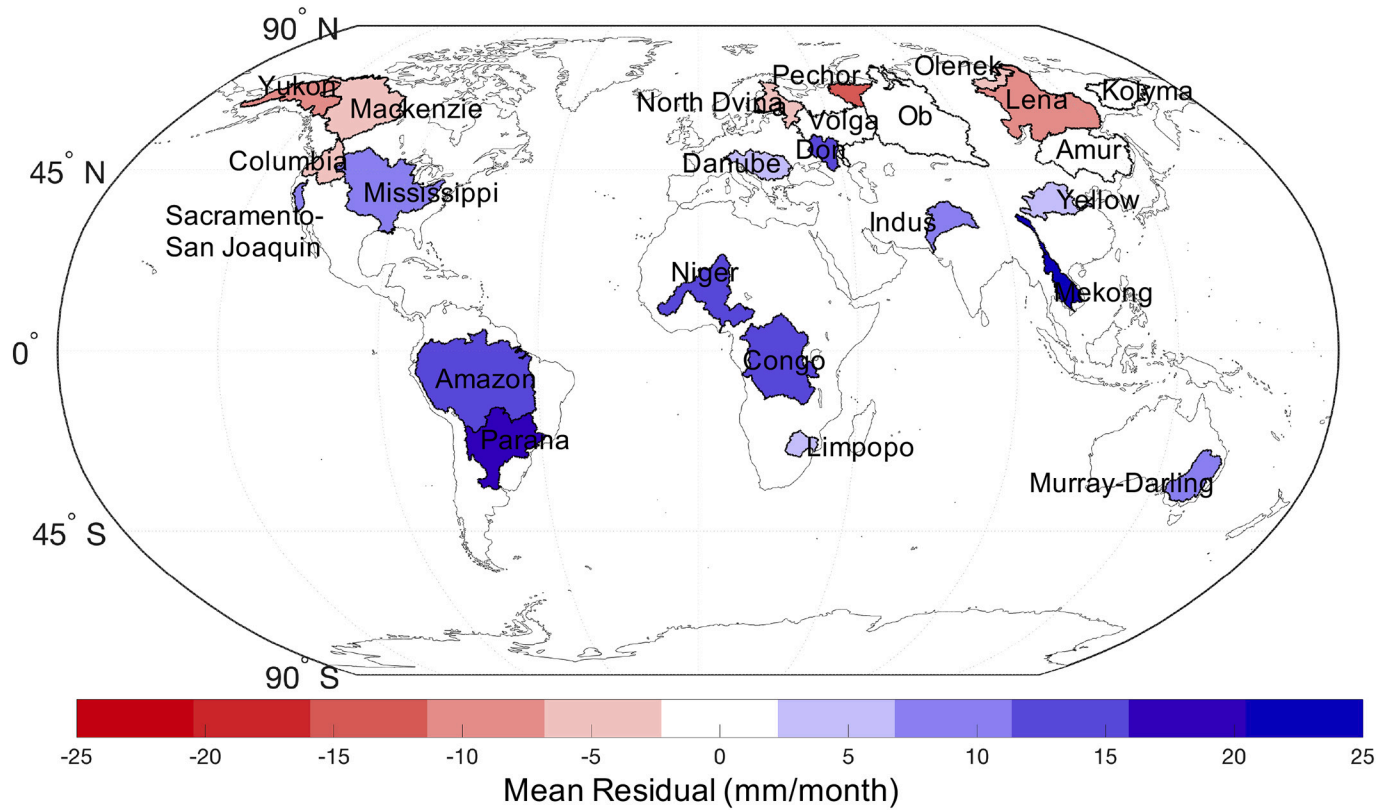

Fig. 3. Water balance residual for 24 major global river basins: The residual error in the water budget (r from Eq. 2) calculated as the mean residual across all combinations of available observational products for each basin at a monthly time step.

vary between 0.7% and 30% by basin, which is comparable to 4.3% globally and 5–25% by basin reported by Rodell et al. (2015) and Sahoo et al. (2011), respectively.

Fig. 4 shows the spatial variability of the water balance residual, r (Eq. 2 and Fig. 3), in the context of meteorological and geophysical variables. Consistent with the spatial patterns in Fig. 3, we find mean basin r decreases significantly with mean latitude ($R^2 = 0.55$, $p < 0.001$) and increases significantly with mean basin temperature ($R^2 = 0.52$, $p < 0.003$) in Fig. 4 below. We also explore the relationship of r with EVI and P, hypothesizing that vegetation may interfere with satellite retrievals. There is a weak correlation between mean basin r and mean basin EVI ($R^2 = 0.23$, $p < 0.02$) and a similar relationship between mean basin r and P ($R^2 = 0.21$, $p = 0.03$). The similarity between these is at least partially attributable to the relationship between P and EVI ($R^2 = 0.69$, $p < 0.001$). We also explored the collinearity between residuals with basin area, finding a weak relationship ($R^2 = 0.04$; not shown) that is not statistically significant. We further explored residual errors in the context of human modifications by calculating the collinearity between r with the percent of irrigated area for each basement, also finding a weak relationship ($R^2 = 0.07$; not shown) that is not statistically significant. This lack of relationship may be attributable to most basins having less than 2% irrigated area.

4.2. Comparing REESEN with climate data record (CDR)

To limit the potentially large number of figures associated with the full set of 24 basins, Fig. 5 and Fig. 6 show only a sub-set of three wet basins (Amazon, Mekong, Congo; blue in Fig. S13), and three dry basins (Yellow, Mackenzie, Olenek; red in Fig. S13), respectively. Nevertheless, summary statistics reflect all 24 study basins and the complete set of figures for all 24 basins are presented in supplementary Figs. S1-S11.

REESEN realizations show closest agreement with CDR for P, with less agreement in ET and ΔS, and then much less agreement for Q in Fig. 5 and Fig. 6 (summary statistics in Table 2). The range of REESEN ensemble spread is smaller than the reported CDR uncertainty for P and ET, whereas the REESEN ΔS and Q have a larger ensemble spread than CDR's uncertainty (Table 2). It is worth noting that although biases between REESEN ensembles and CDR ΔS are small (0.11–0.43 mm/month, varying by closure technique) we report large percent biases for ΔS in Table 2 as a result of a low mean ΔS which is used in the denominator of the percent bias calculation (Moriasi et al. 2015).

There are substantial differences in the seasonal cycle between REESEN and CDR Q (Fig. 5 and Fig. 6) that are partially attributable to error sources in the models used to construct CDR including: random or systematic errors in the model inputs, uncertainty in model parameterization and errors due to incomplete or biased model structure (Butts et al. 2004). Human impacts on the water cycle, which are expected to be captured in REESEN but not CDR, may also contribute to differences between the two data sources, especially in basin systems where humans significantly alter the water cycle (e.g. Sacramento-San Joaquin Basin System where approximately 32% of the land area is equipped for irrigation; Siebert et al. 2013; Siebert et al. 2015). Attribution of these error sources would require further analysis that is outside the scope of this study. In a system without human impacts, Q directly responds to P, and so it is noteworthy that the CDR estimates of Q have higher correlations with P ($R^2 = 0.50$) than in the REESEN realizations ($R^2 = 0.30$). Since the water budget estimates from CDR are driven by LSMs that neglect human activities, the weaker linear relationship in REESEN may reflect the impact of human activities, assuming human activities alter correlations between Q and P, for example through surface water withdrawals and reservoir management. One such indication over the Sacramento-San Joaquin basin system is that remotely sensed storage depletion and ET are greater in REESEN than in the CDR estimates during the peak irrigation months of June, July, August and September (Fig. S6 and S7).

4.3. Uncertainty in water budget estimates

The magnitude of mean monthly CV in raw retrievals is largest for ET (54.4%) compared to P (18.1%) and ΔS (0%), noting the extremely low CV for ΔS is attributable to all ΔS products used in this study being derived from GRACE observations. CV tends to follow a similar seasonal cycle to respective component magnitudes with significant correlations ($R^2 = 0.41$–0.50; $p < 0.05$) between observed component magnitude and the CV of that component, which supports results from prior literature (Tian and Peters-Lidard 2010).

The frequency of co-occurrences between months of low P (P in bottom quartile) and high uncertainty (CV in top quartile) for raw retrievals of Q, ET and ΔS is significantly greater than by chance ($p < 0.01$;

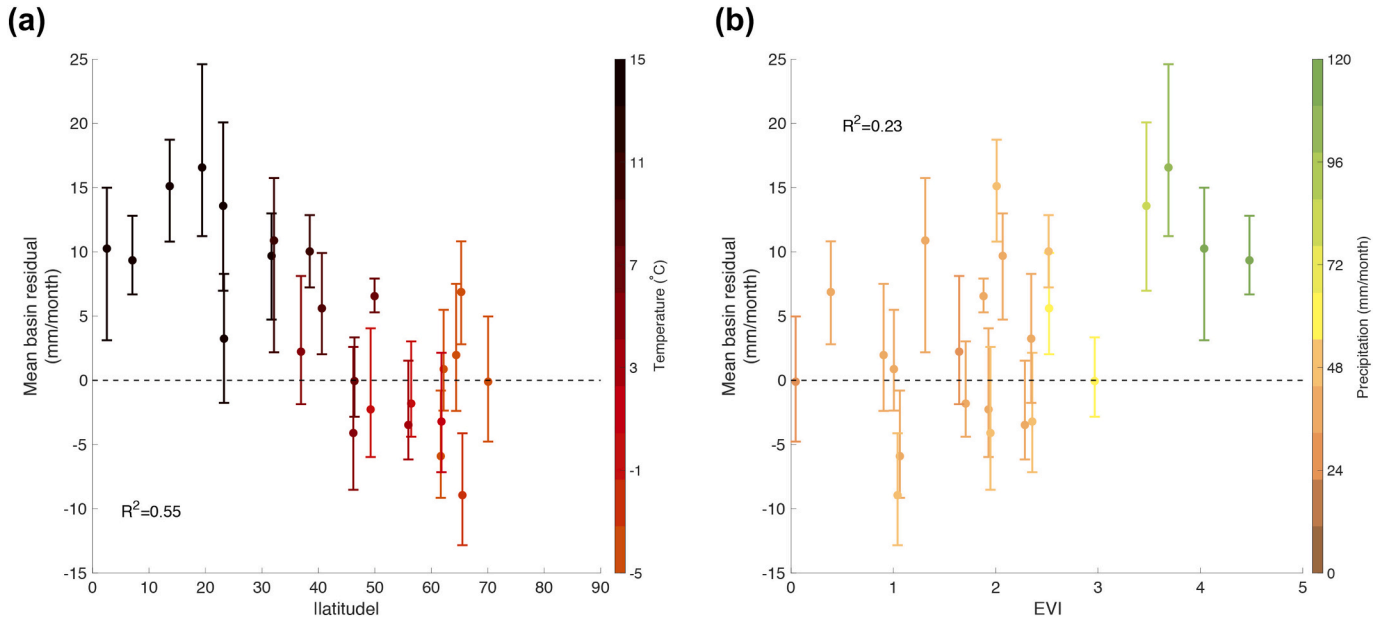

Fig. 4. Collinearity of water budget residuals with meteorological and geophysical variables for each of the 24 study basins: (a) Comparison of mean basin residual and mean basin absolute latitude, symbols colored by mean basin temperature. **(b)** Comparison of mean basin residual and mean basin enhanced vegetation index (EVI), symbols colored by mean basin precipitation; error bars represent the interquartile range of basin residuals across al REESEN realizations.

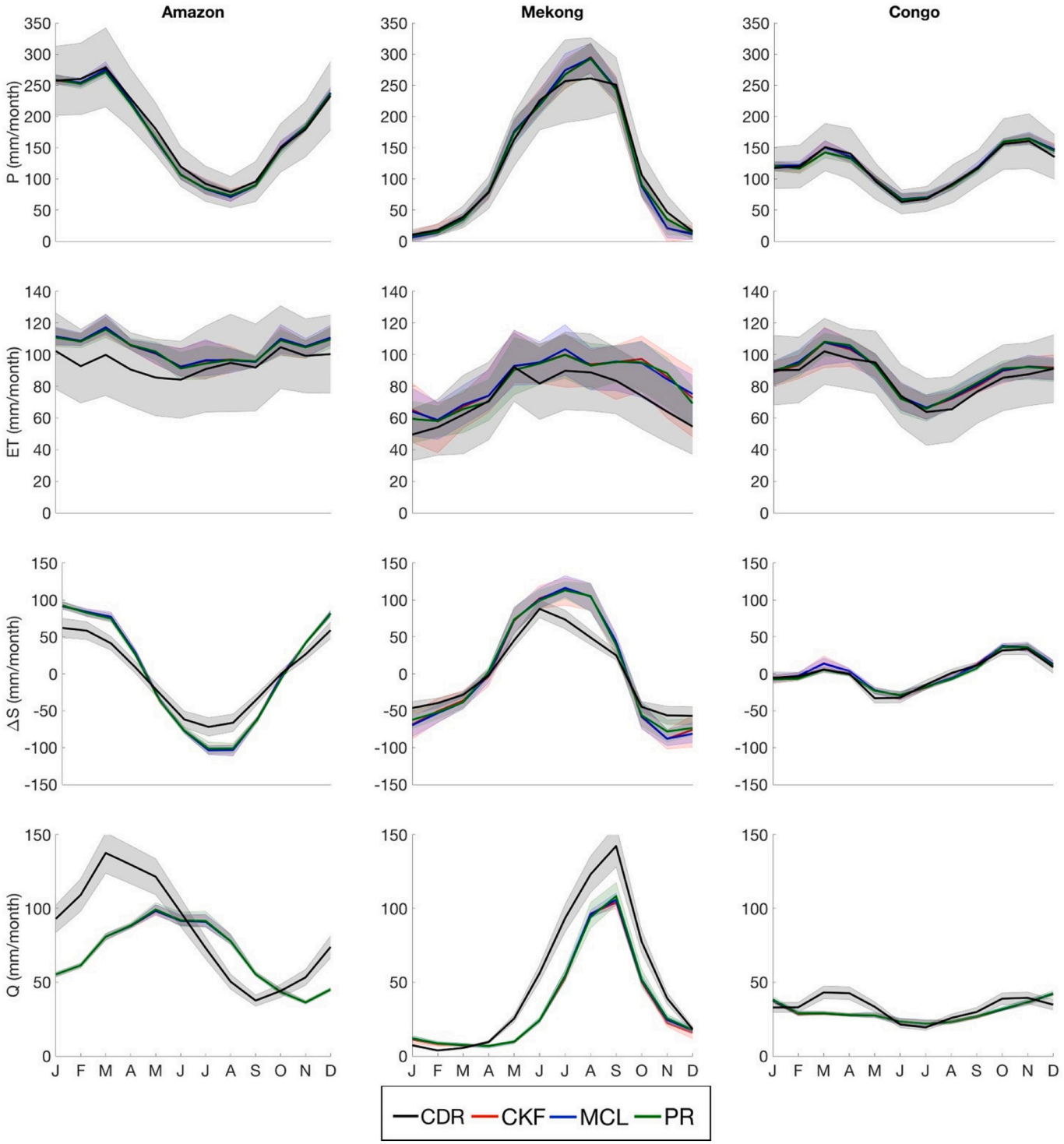

Fig. 5. Comparison of REESEN with the CDR for three representative wet basins: Seasonal cycles of balance-corrected REESEN water budget components compared to the Zhang et al. (2018) CDR (black) over 3 wet basins. REESEN includes the CKF (red), MCL (blue) and PR (green) closure methods, with uncertainty expressed as the ensemble range, whereas CDR uncertainty reflects the values reported in Zhang et al. (2018).

Table 3). Also, the frequency for co-occurrences between bottom quartiles, IQRs and top quartiles of uncertainty across raw retrievals of P, ET, and ΔS is also statistically significant ($p < 0.01$; Table 3). This result is also found when using corrected estimates of P, ET, and ΔS, rather than raw retrievals (not shown).

We note consistent decreases in uncertainty of water budget components after applying closure constraints, relative to raw remotely sensed retrievals (Table 4; Fig. 7). All three closure constraint algorithms significantly decrease uncertainty in ET and ΔS ($p < 0.01$), but only the PR technique results in significant reductions of P uncertainty. There is a high correspondence in relative CV for the CKF and MCL closure techniques ($R^2 > 0.6$ for all budget components), whereas relative CV from PR behaves differently ($R^2 < 0.3$ for all budget components). This is attributable to a key difference between closure techniques: PR assumes error variance is proportional to product magnitude, whereas CKF and MCL techniques assumes error variance is quantified based on ensemble

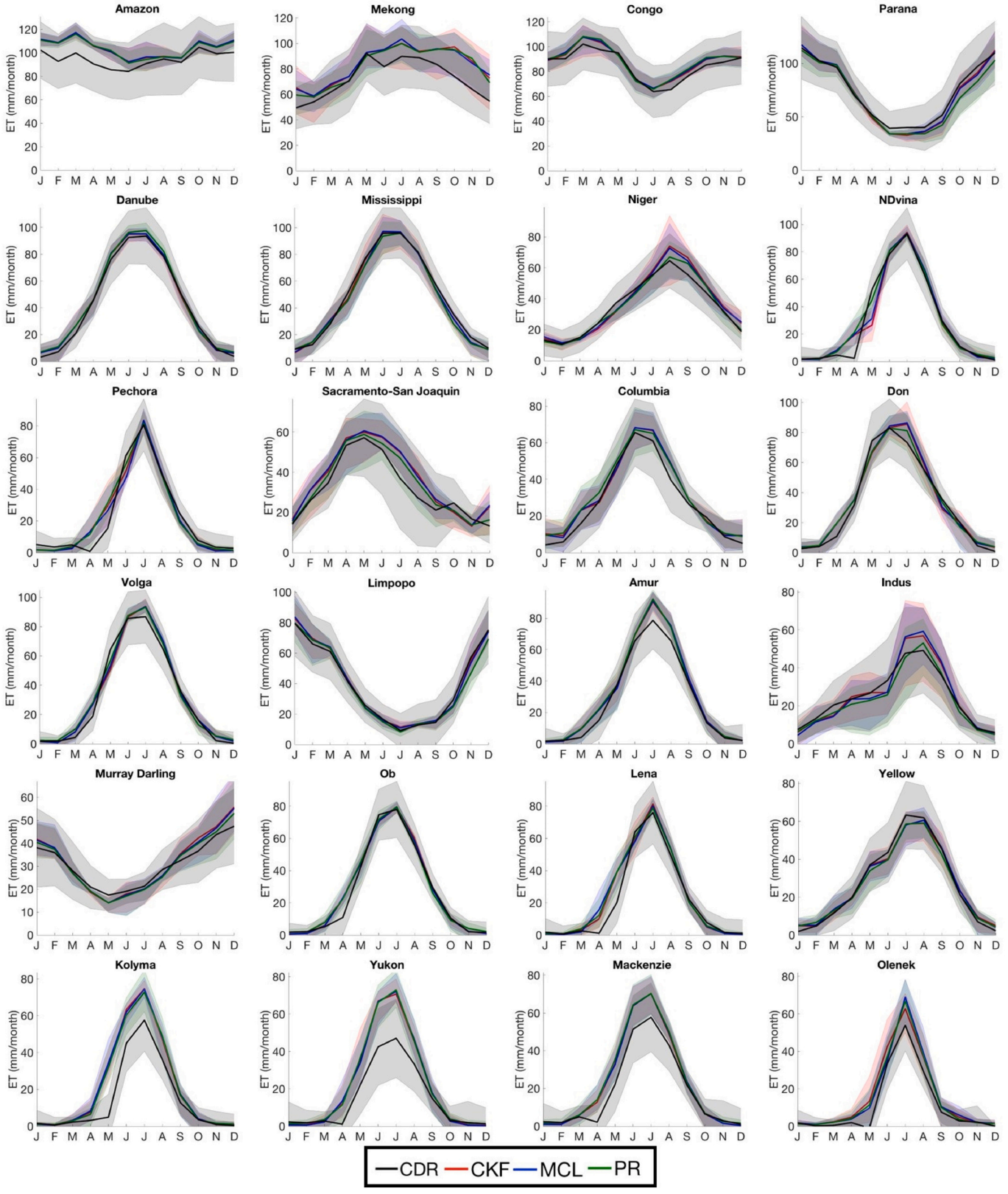

Fig. 6. Comparison of REESEN with the CDR for three representative dry basins: Seasonal cycles of balance-corrected REESEN water budget components compared to the Zhang et al. (2018) CDR (black) over 3 dry basins. REESEN includes the CKF (red), MCL (blue) and PR (green) closure methods, with uncertainty expressed as the ensemble range, whereas CDR uncertainty reflects the values reported in Zhang et al. (2018).

Table 2
R^2 and percent bias between the CDR and the REESEN ensemble mean across three closure algorithms for each water budget component. The relative range reflects the range of estimates for a water budget variable across all REESEN realizations divided by CDR uncertainty.

Closure method	R^2	Percent bias	Relative range
PRECIPITATION (P)			
CKF	0.94	6.6%	0.63
MCL	0.95	6.6%	
PR	0.93	4.0%	
EVAPOTRANSPIRATION (ET)			
CKF	0.94	10.9%	0.65
MCL	0.94	10.9%	
PR	0.94	9.7%	
CHANGE IN TERRESTRIAL WATER STORAGE (ΔS)			
CKF	0.92	49.2%	1.92
MCL	0.92	55.6%	
PR	0.91	24.6%	
STREAMFLOW (Q)			
CKF	0.42	−7.2%	1.4
MCL	0.45	−7.1%	
PR	0.47	−9.0%	

Table 3
Co-occurrences of low precipitation (*P*) and high uncertainty (CV): Frequency of instances of low precipitation (bottom quartile) and high uncertainty (top quartile) for the three remotely sensed water budget components. **Co-occurrences of CV quartile categories across variables**: Co-occurrences of uncertainty expressed by the frequency of overlapping bottom quartiles, interquartile ranges, and top quartiles of CV between combinations of remotely sensed water budget components, with *p-values* reported for statistical significance.

Description of components and quartiles	Frequency of co-occurrences (%)	Significance (*p-value*)
Co-occurrences of low precipitation (*P*) and high uncertainty (CV)		
Bottom quartile *P* & top quartile CV of *P*	46%	$p < 0.01$
Bottom quartile *P* & top quartile CV of *ET*	46%	$p < 0.01$
Bottom quartile *P* & top quartile CV of *ΔS*	39%	$p < 0.01$
Co-occurrences of CV quartile categories across variables		
Bottom quartiles of *P* CV & *ET* CV	32%	$p < 0.01$
Bottom quartiles of *P* CV & *ΔS* CV	22%	$p < 0.01$
Bottom quartiles of *ET* CV & *ΔS* CV	35%	$p < 0.01$
IQRs of *P* CV & *ET* CV	53%	$p < 0.01$
IQRs of *P* CV & *ΔS* CV	48%	$p < 0.01$
IQRs of *ET* CV & *ΔS* CV	45%	$p < 0.01$
Top quartiles of *P* CV & *ET* CV	40%	$p < 0.01$
Top quartiles of *P* CV & *ΔS* CV	23%	$p < 0.01$
Top quartiles of *ET* CV & *ΔS* CV	27%	$p < 0.01$

variability. As a result, there is a strong linear relationship between relative CV and product magnitude in PR compared to CKF and MCL solutions. Overall, the PR technique, relative to CKF and MCL, most frequently calculates REESEN water budget estimates with reduced uncertainty relative to the ensemble of raw retrievals.

4.4. Temporal transferability of closure constraints

The closure constraints from one time period are transferred and used to estimate Q (Q_{est}) in another time period, which is evaluated against gaged Q (Q_{obs}) in Fig. 8. The R^2 between raw Q_{est}, i.e. using raw retrievals, and Q_{obs} is 0.64. This improves to 0.92, 0.76, and 0.75 for REESEN applications, i.e. when using corrected budget estimates using PR, CKF, and MCL techniques, respectively. Similarly, the percent bias for the raw Q_{est} relative to Q_{obs} is 20%, which improves to 5%, −5%, and

Table 4
Frequency of uncertainty reductions for each component and each closure constraint algorithm in REESEN ensembles relative to raw retrievals, with *p-values* reported for statistical significance.

Reductions in uncertainty (CV) induced by closure constraints		
Closure method	Reductions in CV (% of timesteps)	Significance (*p-value*)
Precipitation		
CKF	90%	$p > 0.99$
MCL	94%	$p > 0.99$
PR	99%	$p < 0.01$
Evapotranspiration		
CKF	91%	$p < 0.01$
MCL	91%	$p < 0.01$
PR	99%	$p < 0.01$
Change in terrestrial water storage		
CKF	88%	$p < 0.01$
MCL	94%	$p < 0.01$
PR	100%	$p < 0.01$

− 1%, for REESEN PR, CKF, and MCL techniques, respectively. Therefore, we can say that closure constraint transferability improves estimates of Q for all three correction techniques on the basis of increasing correlation and decreasing the magnitude of the bias relative to gage observations for all basins simultaneously. However, when summary statistics are computed for each basin individually, the median correlation between raw Q_{est} and Q_{obs} ($R^2 = 0.49$) is greater than the correlations between REESEN Q_{est} and Q_{obs}: $R^2 = 0.44, 0.15, 0.14$ for PR, CKF and MCL respectively. Both raw and REESEN cases can be categorized as "unsatisfactory" according to Moriasi et al. (2015). Notwithstanding, closure constraints consistently decrease the PBIAS at the single basin scale, with a median PBIAS for raw Q_{est} of 18%, decreasing to 5%, −3% and − 1% for the REESEN Q_{est} from PR, CKF and MCL algorithms, respectively. The PBIAS is "satisfactory" for the raw case, improving to "good" or "very good" in the corrected cases (Moriasi et al. 2015). Overall, temporal transferability of closure constraints is a viable technique for reducing bias, although it does not consistently improve the coefficient of determination for individual basins.

4.5. Summary of closure constraints applied to each data product

Overall, *P* closure constraints tend to be negative (Fig. 9), meaning the corrected REESEN *P* values are generally smaller than their raw values. Conversely, closure constraints for *ET*, *Q*, and *ΔS* are positive, which results from the tendency for positive residuals in lower latitude basins (Fig. 4). These effects of closure constraints on product magnitude were also consistent for the 12 basins excluded from Fig. 9 (Fig. 3 and Fig. S12).

The mean magnitude of *ΔS* across the 12 basins and ensemble members for all products is approximately 0 mm/month, attributable to the seasonally oscillating positive and negative storage changes. Additionally, the tight grouping of *ΔS* magnitudes and closure constraints for all techniques is attributable to a single source (GRACE) for all *ΔS*-products. For all closure techniques, the GLEAM *ET* receives the lowest closure constraints compared to other *ET* products. This is because water budget realizations with GLEAM *ET* yielded relatively low residuals, which indicates that the magnitude of GLEAM *ET* has higher correspondence with other components from a water budget perspective. The interquartile range of mean closure constraints is largest for *P* products, indicating that *P* product closure constraints are most sensitive to the combinations of other products used in the REESEN ensemble member creation process.

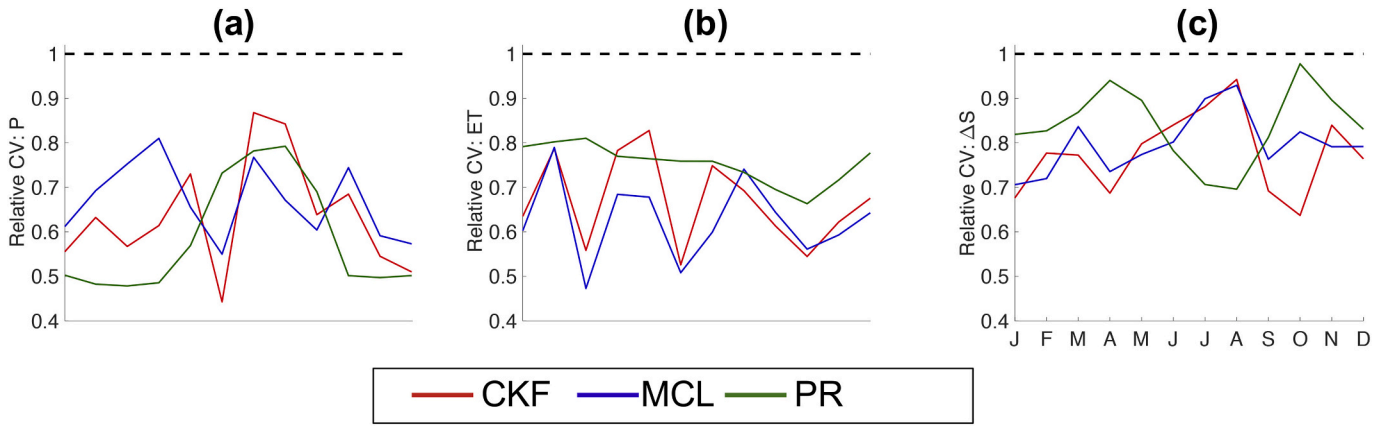

Fig. 7. Example of uncertainty reduction for each water budget component over the Amazon basin: Seasonal cycle of the relative coefficient of variation (relative CV) for water budget components from REESEN when CKF (red), MCL (blue) and PR (green) closure techniques are applied over the Amazon basin as an example. Values exceeding 1 (dashed black line) denote greater CV for corrected REESEN water budget components relative to the ensemble of raw retrievals. A similar plot for the other basins is provided in Figs. S9-S11.

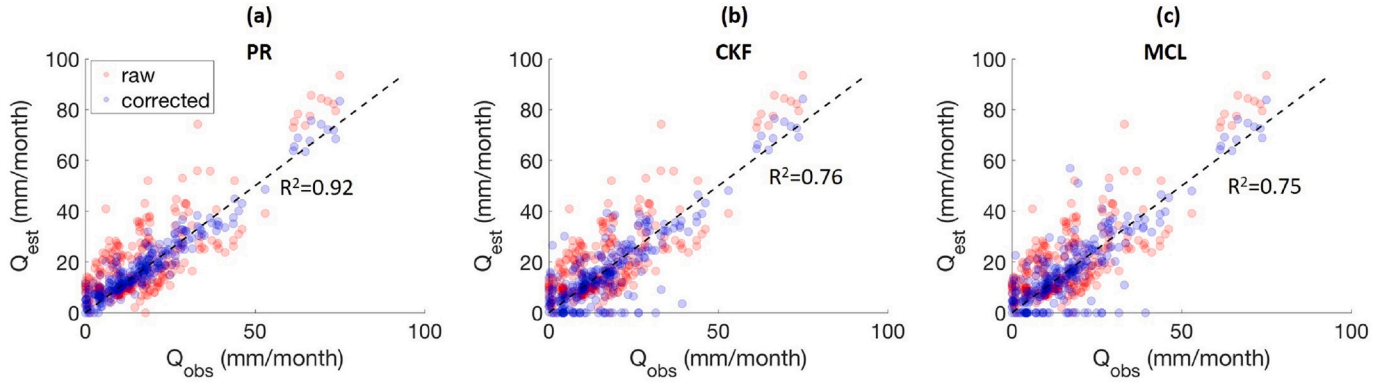

Fig. 8. Temporal transferability of closure constraints: Observed streamflow (horizontal axis) compared to estimated streamflow (vertical axis) at an annual time interval for all 24 study basins. Q_{est} is estimated by a drop 1-year jackknifing approach (Sec. 3.5). Red dots denote raw Q_{est}, while blue dots are from corrected REESEN Q_{est} using each closure technique: **(a)** proportional redistribution (PR), **(b)** constrained Kalman filter (CKF), **(c)** multiple collocation (MCL).

5. Discussion

5.1. Residual error from remotely sensed terrestrial water budget realizations

We show that basin Q cannot be reliably estimated as the residual from raw remotely sensed water budget retrievals because water budget realizations result in residual errors that are, on average, 1.09 times larger than the magnitude of Q. Rather than being random in nature, these residual errors are significantly correlated with geophysical characteristics (Figs. 3 and 4). There is a strong relationship between water budget residual errors with latitude and temperature, which we expect to be related to representation of liquid water versus snow in products' retrieval algorithms and satellite sampling. For example, there are differences in uncertainty associated with *ET* and sublimation calculations (Miralles et al. 2011). Fisher et al. (2008) notes the largest uncertainty in PT-JPL *ET* results in high northern latitudinal regions during winter months due to image obstruction that resulted in a zonal striping pattern. Also, because all precipitation products used in this study rely on gaged observations either directly or indirectly, underestimates in P magnitude may be expected because of gauge "under catch" errors. Fig. 4 also highlights a significant correlation between EVI and residual errors, which is consistent with prior literature that reports uncertainties in *ET* estimates over forested regions that are at least partially attributable to uncertainty in *ET* partitioning between soil evaporation, transpiration and canopy evaporation over these regions (Fisher et al. 2008; Miralles et al. 2011; Stewart 1977; Shuttleworth and Calder, 1979). Rigorous attribution of residual errors to individual sources is a topic for future work that falls outside the scope of this paper.

5.2. Evaluation of remotely sensed ensembles of the terrestrial water budget

Substantial differences in the seasonal cycle between REESEN and CDR Q are most likely attributable to error sources in the models used to construct CDR, but we also suspect that human impacts on the water cycle may significantly contribute to these differences in heavily irrigated basin systems. For example, biases over the Sacramento-San Joaquin basin system between REESEN, which contains human signatures, and CDR, which does not, follows a pattern consistent with irrigation: (i) REESEN shows higher *ET* during low-precipitation months when irrigation is most intense, (ii) REESEN shows larger storage depletion during low-precipitation months when irrigation is expect to heavily rely on groundwater withdrawals, and (iii) observed Q in REESEN is lower than simulated Q in CDR during high-precipitation months when irrigation extractions are expected to rely heavily on streamflow (Figs. S5–8).

Overall, this study showed that closure constraints provide additional value outside of closing the water budget, including reduction of

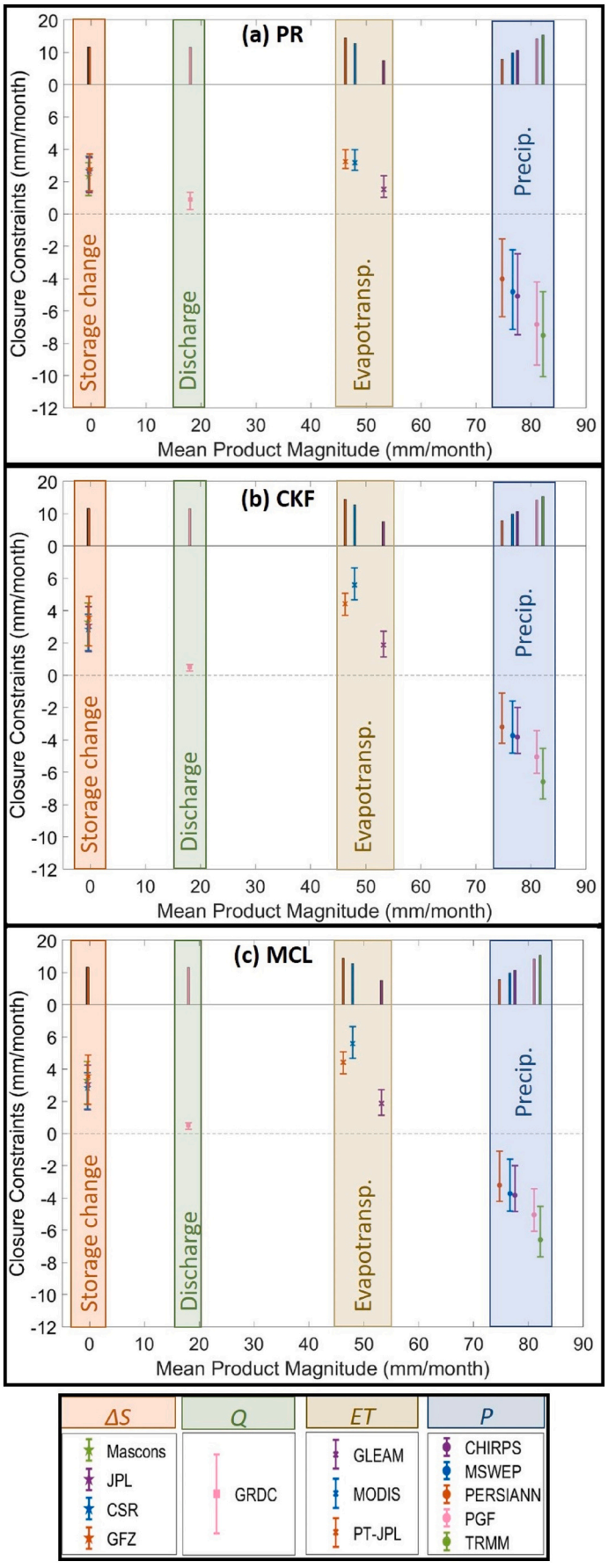

Fig. 9. Closure constraints and budget residuals for each data product: Comparison of mean product corrections (lower left axis) and mean residuals (upper left axis) relative to raw product magnitudes (horizontal axis) across the 12 basins with full coverage from all remotely sensed products used in the REESEN analysis. These basins are situated within 50°N to 50°S and include the Amazon, Congo, Danube, Indus, Limpopo, Mekong, Mississippi, Murray Darling, Niger, Sacramento-San Joaquin, Parana, and Yellow basins. Error bars represent the interquartile range across all ensemble members and shapes represent medians. The interquartile range represents the sensitivity of a product's corrections to its combinations with other products in the REESEN. Bars on the top panels represent the mean residual from ensemble members using respective products. Panels **a**, **b** and **c** show results from REESEN realizations using PR, CKF and MCL closure algorithms, respectively.

uncertainty and transfer of closure constraints in time to provide skillful estimates of mean annual *Q*. We also showed that a simpler closure technique, PR, performed better than more complex ones in decreasing uncertainty and for transfer through time to estimate *Q*. This may be attributable to CKF and MCL requiring more complex information about errors than PR. Overall, the challenges in characterizing these error structures may outweigh benefits that the more complex techniques offer relative to the simpler technique with more straightforward implementation.

5.3. Summary of closure constraints applied to each data product

Average residual errors associated with all products are positive for middle and low latitude basins (Fig. 9), leading to consistently negative closure constraints applied to *P* and positive closure constraints applied to *ET*, *Q* and ΔS. Thus, closure algorithms estimate a high bias in *P* and low biases in *ET*, *Q* and ΔS. However, these assigned biases may be incorrect. Namely, a bias in one component can dictate the sign of closure constraints applied to other components. For example, if observations of *ET*, ΔS, and *Q* are unbiased and *P* products have a positive bias relative to the "true" magnitude of precipitation, closure algorithms will erroneously apply positive closure constraints to *ET*, ΔS and *Q*. Similarly, combinations of *ET*, ΔS or *Q* may have negative biases that could systematically cause negative closure constraints to be applied to *P*. A rigorous assessment of product biases is outside the scope of this study. However, such an assessment would require relatively dense networks of observations of each water budget component, which are presently lacking for most components in most basins. Beck et al. (2020) bias corrected *P* using observations of *Q* within a Budyko framework that relates aridity to runoff production. However, such an approach lacks treatment of seasonal storage changes for which 'true' values are essentially 'unknowable' with current observational capabilities.

6. Conclusion

This study presented a new approach to quantifying the components of the terrestrial water budget using three remotely sensed ensembles called REESEN. Combinations of observation-based products produced residual errors equal to 15% of the total precipitation on average, varying between 0.7%–30% by basin. The magnitude of these errors varies significantly with geophysical and meteorological characteristics and are, on average, positive. Given the tendency for *r* to be positive, we conclude that either remotely sensed *P* has a wet bias or combinations of *ET*, *Q*, and ΔS have a dry bias. REESEN realizations were generally consistent with the Zhang et al. (2018) Climate Data Record for *P*, *ET* and ΔS but had large disagreements in *Q*. Overall, this study showed that closure constraints provide additional value beyond closing the water budget. These benefits include reducing water balance uncertainty and the potential to transfer closure constraints across time to provide skillful estimates of mean annual *Q*. We also showed that a simpler closure technique, PR, was superior to more complex ones in the context of decreasing uncertainty and for temporal transfer to estimate *Q*. REESEN realizations may be used in future studies that attempt to quantify human impacts on the water budget. For example, data assimilation of REESEN components using LSMs could provide models with information about unmodeled but observable human impacts such as irrigation (Abolafia-Rosenzweig et al., 2019). Further, differencing REESEN with data that lack human signatures, e.g. 'naturalized streamflow', has the potential to leave the human signal as an anomaly. This analysis provided evidence that the water budget, as observed by remote sensing and in situ measurements, is biased and errors are not strictly random. Thus, future analyses should consider incorporating biases and other detailed information in the process of calculating closure constraints.

Data availability

REESEN realizations developed in this study are archived on Mendeley Data and available at https://doi.org/10.17632/r24rdxt73j.3 (Abolafia-Rosenzweig and Livneh, 2020).

Credit author statement

Ronnie Abolafia-Rosenzweig: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Validation; Visualization; Writing. **Ben Livneh:** Conceptualization; Investigation; Methodology; Visualization; Funding acquisition; Project administration; Supervision; Writing; review & editing. **Ming Pan:** Conceptualization; Funding acquisition; Methodology; review & editing. **Jun Lin Zeng:** Formal analysis; Visualization; reviewing & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

- The authors declare no conflict of interest
- This research was funded by the National Aeronautics and Space Administration: NASA Grant # 80NSSC18K0951: A Remotely Sensed Ensemble to Understand Human Impacts on the Water Cycle and NASA Grant # NNX16AQ46G: Monitoring soil evaporation using SMAP surface soil moisture in a water balance framework.
- This work utilized the RMACC Summit supercomputer, which is supported by the National Science Foundation (awards ACI-1532235 and ACI-1532236), the University of Colorado Boulder, and Colorado State University. The Summit supercomputer is a joint effort of the University of Colorado Boulder and Colorado State University.
- The study and its presentation greatly benefit from the review comments from the Associate Editor, Dr. Tim R. McVicar, and the three anonymous reviewers.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.rse.2020.112191.

References

Abolafia-Rosenzweig, R., Livneh, B., 2020. Remotely sensed ensemble of the water cycle. Mendeley Data 2. https://doi.org/10.17632/r24rdxt73j.3.

Abolafia-Rosenzweig, R., Livneh, B., Small, E.E., Kumar, S.V., 2019. Soil moisture data assimilation to estimate irrigation water use. J. Adv. Model. Earth Syst 2019MS001797. https://doi.org/10.1029/2019MS001797.

Aires, F., 2014. Combining datasets of satellite-retrieved products. Part I: Methodology and Water Budget Closure. J. Hydrometeorol. 15, 1677–1691.

Beaudoing, H., Rodell, M., NASA/GSFC/HSL, 2020. GLDAS Noah Land Surface Model L4 3 hourly 0.25 x 0.25 degree V2.1, Greenbelt, Maryland, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC), Accessed: [07/2019], 10.5067/E7TYRXPJKWOQ.

Beck, H.E., van Dijk, A.I.J.M., Levizzani, V., Schellekens, J., Miralles, D.G., Martens, B., de Roo, A., 2017a. MSWEP: 3-hourly 0.25° global gridded precipitation (1979–2015) by merging gauge, satellite, and reanalysis data. Hydrol. Earth Syst. Sci. 21, 589–615. https://doi.org/10.5194/hess-21-589-2017.

Beck, H.E., Vergopolan, N., Pan, M., Levizzani, V., van Dijk, A.I.J.M., Weedon, G.P., Brocca, L., Pappenberger, F., Huffman, G.J., Wood, E.F., 2017b. Global-scale evaluation of 22 precipitation datasets using gauge observations and hydrological modeling. Hydrol. Earth Syst. Sci. 21, 6201–6217. https://doi.org/10.5194/hess-21-6201-2017.

Beck, H.E., Wood, E.F., Pan, M., Fisher, C.K., Miralles, D.G., van Dijk, A.I.J.M., McVicar, T.R., Adler, R.F., 2019. MSWEP V2 global 3-hourly 0.1° precipitation: methodology and quantitative assessment. Bull. Am. Meteorol. Soc. 100, 473–500. https://doi.org/10.1175/BAMS-D-17-0138.1.

Beck, H.E., Wood, E.F., Mcvicar, T.R., Zambrano-Bigiarini, M., Alvarez-Garreton, C., Baez-Villanueva, O.M., Sheffield, J., Karger, D.N., 2020. Bias correction of global high-resolution precipitation climatologies using streamflow observations from 9372 catchments. J. Clim. 33, 17.

Bosshard, T., Carambia, M., Goergen, K., Kotlarski, S., Krahe, P., Zappa, M., Schär, C., 2013. Quantifying uncertainty sources in an ensemble of hydrological climate-impact projections: UNCERTAINTY SOURCES IN CLIMATE-IMPACT PROJECTIONS. Water Resour. Res. 49, 1523–1536. https://doi.org/10.1029/2011WR011533.

Brocca, L., Ciabatta, L., Massari, C., Hahn, S., Wagner, W., 2016. SM2RAIN-ASCAT (1 Jan 2007–30 June 2015) global daily rainfall dataset, ResearchGate. https://doi.org/10.13140/RG.2.1.4434.8563.

Butts, M.B., Payne, J.T., Kristensen, M., Madsen, H., 2004. An evaluation of the impact of model structure on hydrological modelling uncertainty for streamflow simulation. J. Hydrol. 298, 242–266. https://doi.org/10.1016/j.jhydrol.2004.03.042.

Carter, R.W., Anderson, I.E., 1963. Accuracy of current meter measurements. American Society Civil Engineers Proc., Jour. Hydraulics Div., Julv 1963, pp. 105–115.

Chen, M., Shi, W., Xie, P., Silva, V.B.S., Kousky, V.E., Higgins, R.W., Janowiak, J.E., 2008. Assessing objective techniques for gauge-based analyses of global daily precipitation. J. Geophys. Res. 113, D04110 https://doi.org/10.1029/2007JD009132.

Chen, Y., Xia, J., Liang, S., Feng, J., Fisher, J.B., Li, Xin, Li, Xianglan, Liu, S., Ma, Z., Miyata, A., Mu, Q., Sun, L., Tang, J., Wang, K., Wen, J., Xue, Y., Yu, G., Zha, T., Zhang, L., Zhang, Q., Zhao, T., Zhao, L., Yuan, W., 2014. Comparison of satellite-based evapotranspiration models over terrestrial ecosystems in China. Remote Sens. Environ. 140, 279–293. https://doi.org/10.1016/j.rse.2013.08.045.

Clarke, R.T., 1999. Uncertainty in the estimation of mean annual flood due to rating-curve indefinition. J. Hydrol. 222 (1–4), 185–190.

Dai, A., 2017. Dai and Trenberth Global River flow and continental discharge dataset. Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory. https://doi.org/10.5065/D6V69H1T.

Didan, K., Barreto Munoz, A., Solano, R., Huete, A., 2015. MODIS Vegetation Index User's Guide; Collection 6. Washington, DC, USA, NASA.

Dingman, S.L., 2002. Physical hydrology. New York: Macmillan Publishing Company 646 pp.

Durre, I., Menne, M.J., Gleason, B.E., Houston, T.G., Vose, R.S., 2010. Comprehensive automated quality assurance of daily surface observations. J. Appl. Meteorol. Climatol. 49, 1615–1633. https://doi.org/10.1175/2010JAMC2375.1.

Epstein, E.S., 1969. Stochastic dynamic prediction. Tellus 21, 739–759. https://doi.org/10.1111/j.2153-3490.1969.tb00483.x.

Evensen, G., 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. J. Geophys. Res. 99, 10143. https://doi.org/10.1029/94JC00572.

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

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., Verdin, A.P., 2014. A quasi-global precipitation time series for drought monitoring: U.S. Geological Survey Data Series, 832. https://doi.org/10.3133/ds832, 4 p.

Gao, H., Tang, Q., Ferguson, C.R., Wood, E.F., Lettenmaier, D.P., 2010. Estimating the water budget of major US river basins via remote sensing. Int. J. Remote Sens. 31, 3955–3978. https://doi.org/10.1080/01431161.2010.483488.

Guo, H., Chen, S., Bao, A., Hu, J., Gebregiorgis, A., Xue, X., Zhang, X., 2015. Inter-comparison of high-resolution satellite precipitation products over Central Asia. Remote Sens. 7, 7181–7211. https://doi.org/10.3390/rs70607181.

Hsu, K.-L., Gao, X., Sorooshian, S., Gupta, H.V., 1997. Precipitation estimation from remotely sensed information using artificial neural networks. J. Appl. Meteorol. 36, 15.

Huffman, G.J., Adler, R.F., Arkin, P., Chang, A., Ferraro, R., Gruber, A., Janowiak, J., McNab, A., Rudolf, B., Schneider, U., 1997. The global precipitation climatology project (GPCP) combined precipitation dataset. Bull. Amer. Meteor. Soc. 78, 5–20. https://doi.org/10.1175/1520-0477(1997)078<0005:TGPCPG>2.0.CO;2.

Huffman, G.J., Adler, R.F., Morrissey, M.M., Bolvin, D.T., Curtis, S., Joyce, R., Mcgavock, B., Susskind, J., 2001. Global precipitation at one-degree daily resolution from multisatellite observations. J. Hydrometeorol. 2, 15.

Huffman, G.J., Bolvin, D.T., Nelkin, E.J., Wolff, D.B., Adler, R.F., Gu, G., Hong, Y., Bowman, K.P., Stocker, E.F., 2007. The TRMM multisatellite precipitation analysis (TMPA): quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. J. Hydrometeorol. 8, 38–55. https://doi.org/10.1175/JHM560.1.

Iguchi, T., Kozu, T., Kwiatkowski, J., Meneghini, R., Awaka, J., Okamoto, K., 2009. A Kalman filter approach to the global satellite mapping of precipitation (GSMaP) from combined passive microwave and infrared radiometric data. J. Meteorol. Soc. Jpn. 87A, 137–151.

Janowiak, J.E., Joyce, R.J., Yarosh, Y., 2001. A real-time global half-hourly pixel-resolution infrared dataset and its applications. Bull. Am. Meteorol. Soc. 82 (2), 205–217.

Khan, M.S., Liaqat, U.W., Baik, J., Choi, M., 2018. Stand-alone uncertainty characterization of GLEAM, GLDAS and MOD16 evapotranspiration products using an extended triple collocation approach. Agric. For. Meteorol. 252, 256–268. https://doi.org/10.1016/j.agrformet.2018.01.022.

Knapp, K.R., Ansari, S., Bain, C.L., Bourassa, M.A., Dickinson, M.J., Funk, C., Helms, C. N., Hennon, C.C., Holmes, C.D., Huffman, G.J., Kossin, J.P., Lee, H.-T., Loew, A., Magnusdottir, G., 2011. Globally gridded satellite observations for climate studies. Bull. Amer. Meteor. Soc. 92, 893–907. https://doi.org/10.1175/2011BAMS3039.1.

Kumar, S.V., Peters-Lidard, C.D., Santanello, J.A., Reichle, R.H., Draper, C.S., Koster, R. D., Nearing, G., Jasinski, M.F., 2015. Evaluating the utility of satellite soil moisture retrievals over irrigated areas and the ability of land data assimilation methods to correct for unmodeled processes. Hydrol. Earth Syst. Sci. 19, 4463–4478. https://doi.org/10.5194/hess-19-4463-2015.

Kumar, S.V., Zaitchik, B.F., Peters-Lidard, C.D., Rodell, M., Reichle, R., Li, B., Jasinski, M., Mocko, D., Getirana, A., De Lannoy, G., Cosh, M.H., Hain, C.R., Anderson, M., Arsenault, K.R., Xia, Y., Ek, M., 2016. Assimilation of gridded GRACE terrestrial water storage estimates in the north American land data assimilation system. J. Hydrometeorol. 17, 1951–1972. https://doi.org/10.1175/JHM-D-15-0157.1.

Landerer, F.W., Swenson, S.C., 2012. Accuracy of scaled GRACE terrestrial water storage estimates. Water Resour. Res. 48, W04531, 11 PP. https://doi.org/10.1029/2011WR011453.

Lehner, B., Verdin, K., Jarvis, A., 2006. HydroSHEDS Technical Documentation. World Wildlife Fund US, Washington, DC. Available from: https://hydrosheds.cr.usgs.gov/.

Leith, C.E., 1974. Theoretical skill of Monte Carlo forecasts. Mon. Weather Rev. 102, 409–418.

Livneh, B., Lettenmaier, D.P., 2012. Multi-criteria parameter estimation for the unified land model. Hydrol. Earth Syst. Sci. Discuss. 9, 4417–4463. https://doi.org/10.5194/hessd-9-4417-2012.

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

Martens, B., Miralles, D.G., Lievens, H., van der Schalie, R., de Jeu, R.A.M., Fernández-Prieto, D., Beck, H.E., Dorigo, W.A., Verhoest, N.E.C., 2017. GLEAM v3: satellite-based land evaporation and root-zone soil moisture. Geosci. Model Dev. 10, 1903–1925. https://doi.org/10.5194/gmd-10-1903-2017.

Miralles, D.G., Holmes, T.R.H., De Jeu, R.A.M., Gash, J.H., Meesters, A.G.C.A., Dolman, A.J., 2011. Global land-surface evaporation estimated from satellite-based observations. Hydrol. Earth Syst. Sci. 15, 453–469. https://doi.org/10.5194/hess-15-453-2011.

Miralles, D.G., Jiménez, C., Jung, M., Michel, D., Ershadi, A., McCabe, M.F., Hirschi, M., Martens, B., Dolman, A.J., Fisher, J.B., Mu, Q., Seneviratne, S.I., Wood, E.F., Fernaìndez-Prieto, D., 2015. The WACMOS-ET project – part 2: evaluation of global terrestrial evaporation data sets. Hydrol. Earth Syst. Sci. Discuss. 12, 10651–10700. https://doi.org/10.5194/hessd-12-10651-2015.

Moazami, S., Golian, S., Kavianpour, M.R., Hong, Y., 2013. Comparison of PERSIANN and V7 TRMM multi-satellite precipitation analysis (TMPA) products with rain gauge data over Iran. Int. J. Remote Sens. 34, 8156–8171. https://doi.org/10.1080/01431161.2013.833360.

Monteith, J.L., 1965. Evaporation and environment. Symposium of the society of experimental biology 19, 205–224.

Moriasi, D.N., Gitau, M.W., Pai, N., Daggupati, P., 2015. Hydrologic and water qualitymodels: performance measures and evaluation criteria. Trans. ASABE 58 (6), 1763–1785. https://doi.org/10.13031/trans.58.10715.

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. https://doi.org/10.1016/j.rse.2007.04.015.

Mu, Q., Jones, L.A., Kimball, J.S., McDonald, K.C., Running, S.W., 2009. Satellite assessment of land surface evapotranspiration for the pan-Arctic domain: SATELLITE ASSESSMENT OF LAND SURFACE ET. Water Resour. Res. 45 https://doi.org/10.1029/2008WR007189.

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

Mu, D., Yan, H., Feng, W., Peng, P., 2017. GRACE leakage error correction with regularization technique: Case studies in Greenland and Antarctica. Geophys. J. Int. https://doi.org/10.1093/gji/ggw494 ggw494.

Mueller, D.S., 2003. Field evaluation of boat-mounted acoustic doppler instruments used to measure streamflow. In: Proceedings of the IEEE/OES Seventh Working Conference on Current Measurement Technology, 2003. Presented at the IEEE/OES Seventh Working Conference on Current Measurement Technology, 2003., IEEE, San Diego, CA, USA, pp. 30–34. https://doi.org/10.1109/CCM.2003.1194278.

NSIT, 2007. A NASA earth science implementation plan for energy and water cycle research: predicting energy and water cycle consequences of earth system variability and change. NASA energy and water cycle study (NEWS) science integration team (NSIT), 89 pp. [available online a]t. http://news.cisc.gmu.edu/doc/NEWS_implementation.pdf.

Pan, M., Wood, E.F., 2006. Data assimilation for estimating the terrestrial water budget using a constrained ensemble Kalman filter. J. Hydrometeorol. 7, 534–547. https://doi.org/10.1175/JHM495.1.

Pan, M., Sahoo, A.K., Troy, T.J., Vinukollu, R.K., Sheffield, J., Wood, E.F., 2012. Multisource estimation of long-term terrestrial water budget for major Global River basins. J. Clim. 25, 3191–3206. https://doi.org/10.1175/JCLI-D-11-00300.1.

Pan, M., Fisher, C.K., Chaney, N.W., Zhan, W., Crow, W.T., Aires, F., Entekhabi, D., Wood, E.F., 2015. Triple collocation: beyond three estimates and separation of structural/non-structural errors. Remote Sens. Environ. 171, 299–310. https://doi.org/10.1016/j.rse.2015.10.028.

Peterson, T.C., Vose, R.S., 1997. An overview of the Global Historical Climatology Network temperature database. Bull. Am. Meteorol. Soc. 78 (12), 2,837–2,849.

Purdy, A.J., Fisher, J.B., Goulden, M.L., Colliander, A., Halverson, G., Tu, K., Famiglietti, J.S., 2018. SMAP soil moisture improves global evapotranspiration. Remote Sens. Environ. 219, 1–14. https://doi.org/10.1016/j.rse.2018.09.023.

Ranger, N., Hallegatte, S., Bhattacharya, S., Bachu, M., Priya, S., Dhore, K., Rafique, F., Mathur, P., Naville, N., Henriet, F., Herweijer, C., Pohit, S., Corfee-Morlot, J., 2011.

An assessment of the potential impact of climate change on flood risk in Mumbai. Clim. Chang. 104, 139–167. https://doi.org/10.1007/s10584-010-9979-2.
Rodell, M., 2004. Basin scale estimates of evapotranspiration using GRACE and other observations. Geophys. Res. Lett. 31, L20504 https://doi.org/10.1029/2004GL020873.
Rodell, M., Houser, P.R., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C.-J., Arsenault, K., Cosgrove, B., Radakovich, J., Bosilovich, M., Entin, J.K., Walker, J.P., Lohmann, D., Toll, D., 2004. The global land data assimilation system. Bull. Amer. Meteor. Soc. 85, 381–394. https://doi.org/10.1175/BAMS-85-3-381.
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., Chambers, D., Clark, E., Fetzer, E.J., Gao, X., Gu, G., Hilburn, K., Huffman, G.J., Lettenmaier, D.P., Liu, W.T., Robertson, F.R., Schlosser, C.A., Sheffield, J., Wood, E.F., 2015. The observed state of the water cycle in the early twenty-first century. J. Clim. 28, 8289–8318. https://doi.org/10.1175/JCLI-D-14-00555.1.
Rosenzweig, C., Karoly, D., Vicarelli, M., Neofotis, P., Wu, Q., Casassa, G., Menzel, A., Root, T.L., Estrella, N., Seguin, B., Tryjanowski, P., Liu, C., Rawlins, S., Imeson, A., 2008. Attributing physical and biological impacts to anthropogenic climate change. Nature 453, 353–357. https://doi.org/10.1038/nature06937.
Rudolf, B., 1993. Management and analysis of precipitation data on a routine basis. *Proceedings of International Symposium on Precipitation and Evaporation,* Vol. 1, B. Sevruk and M. Lapin, Eds., Slovak Hydrometeorology institution, 69–76.
Sahoo, A.K., Pan, M., Troy, T.J., Vinukollu, R.K., Sheffield, J., Wood, E.F., 2011. Reconciling the global terrestrial water budget using satellite remote sensing. Remote Sens. Environ. 115, 1850–1865. https://doi.org/10.1016/j.rse.2011.03.009.
Sakumura, C., Bettadpur, S., Bruinsma, S., 2014. Ensemble prediction and intercomparison analysis of GRACE time-variable gravity field models. Geophys. Res. Lett. 41, 1389–1397. https://doi.org/10.1002/2013GL058632.
Scanlon, B.R., Faunt, C.C., Longuevergne, L., Reedy, R.C., Alley, W.M., McGuire, V.L., McMahon, P.B., 2012. Groundwater depletion and sustainability of irrigation in the US High Plains and Central Valley. Proc. Natl. Acad. Sci. 109, 9320–9325. https://doi.org/10.1073/pnas.1200311109.
Schulzweida, Uwe, 2019. CDO User Guide (Version 1.9.8). https://doi.org/10.5281/zenodo.3539275.
Sheffield, J., Wood, E.F., 2007. Characteristics of global and regional drought, 1950–2000: analysis of soil moisture data from off-line simulation of the terrestrial hydrologic cycle. J. Geophys. Res. 112, D17115 https://doi.org/10.1029/2006JD008288.
Sheffield, J., Goteti, G., Wood, E.F., 2006. Development of a 50-year high-resolution global dataset of meteorological Forcings for land surface modeling. J. Clim. 19, 3088–3111. https://doi.org/10.1175/JCLI3790.1.
Sheffield, J., Ferguson, C.R., Troy, T.J., Wood, E.F., McCabe, M.F., 2009. Closing the terrestrial water budget from satellite remote sensing: WATER BUDGET FROM REMOTE SENSING. Geophys. Res. Lett. 36 https://doi.org/10.1029/2009GL037338 n/a-n/a.
Shiklomanov, A.I., Yakovleva, T.I., Lammers, R.B., Karasev, I.Ph., Vörösmarty, Charles, J., Linder, E., 2006. Cold region river discharge uncertainty—estimates from large Russian rivers. J. Hydrol. 326, 231–256. https://doi.org/10.1016/j.jhydrol.2005.10.037.
Shuttleworth, W.J., Calder, I.R., 1979. Has the Priestley-Taylor equation any relevance to forest evaporation? J. Appl. Meteorol. 18, 639–646.
Siebert, S., Verena, H., Karen, F., Jacob, B., 2013. Update of the Digital Global Map of Irrigation Areas to Version 5, October, 1–171.
Siebert, S., Kummu, M., Porkka, M., Döll, P., Ramankutty, N., Scanlon, B.R., 2015. A global data set of the extent of irrigated land from 1900 to 2005. Hydrol. Earth Syst. Sci. 19 (3), 1521–1545. https://doi.org/10.5194/hess-19-1521-2015.
Stewart, J.B., 1977. Evaporation from the wet canopy of a pine forest. Water Resour. Res. 13, 915–921. https://doi.org/10.1029/WR013i006p00915.
Stoffelen, A., 1998. Toward the true near-surface wind speed: error modeling and calibration using triple collocation. J. Geophys. Res. 103, 7755–7766. https://doi.org/10.1029/97JC03180.
Swenson, S.C., 2012. GRACE monthly land water mass grids NETCDF RELEASE 5.0. Ver. 5.0. PO.DAAC, CA, USA. Dataset accessed [2018-03-01] at. https://doi.org/10.5067/TELND-NC005.
Swenson, S.C., Wahr, J., 2006. Post-processing removal of correlated errors in GRACE data. Geophys. Res. Lett. 33, L08402 https://doi.org/10.1029/2005GL025285.
Tapley, B.D., 2004. GRACE measurements of mass variability in the earth system. Science 305, 503–505. https://doi.org/10.1126/science.1099192.
Tian, Y., Peters-Lidard, C.D., 2010. A global map of uncertainties in satellite-based precipitation measurements: UNCERTAINTIES IN PRECIPITATION DATA. Geophys. Res. Lett. 37 https://doi.org/10.1029/2010GL046008 n/a-n/a.
Vörösmarty, C.J., Sahagian, D., 2000. Anthropogenic disturbance of the terrestrial water cycle. BioScience 50, 753. https://doi.org/10.1641/0006-3568(2000)050[0753:ADOTTW]2.0.CO;2.
Wahr, J., Swenson, S., Zlotnicki, V., Velicogna, I., 2004. Time-variable gravity from GRACE: First results: TIME-VARIABLE GRAVITY FROM GRACE. Geophys. Res. Lett. 31 https://doi.org/10.1029/2004GL019779 n/a-n/a.
Watkins, M.M., Wiese, D.N., Yuan, D.-N., Boening, C., Landerer, F.W., 2015. Improved methods for observing Earth's time variable mass distribution with GRACE using spherical cap mascons. J. Geophys. Res. Solid Earth 120. https://doi.org/10.1002/2014JB011547.
Wiese, D.N., Landerer, F.W., Watkins, M.M., 2016. Quantifying and reducing leakage errors in the JPL RL05M GRACE mascon solution. Water Resour. Res. 52, 7490–7502. https://doi.org/10.1002/2016WR019344.
Wiese, D.N., Yuan, C., Boening, F.W., Landerer, M.M., Watkins, 2018. JPL GRACE Mascon Ocean, Ice, and Hydrology Equivalent Water Height Release 06 Coastal Resolution Improvement (CRI) Filtered Version 1.0. Ver. 1.0. PO.DAAC, CA, USA. Dataset accessed [2018-03-01] at. https://doi.org/10.5067/TEMSC-3MJC6.
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., 2012. 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: WATER AND ENERGY FLUX ANALYSIS. J. Geophys. Res.-Atmos. 117 https://doi.org/10.1029/2011JD016048 n/a-n/a.
Xie, P., Rudolf, B., Schneider, U., Arkin, P.A., 1996. Gauge-based monthly analysis of global land precipitation from 1971 to 1994. J. Geophys. Res. 101, 19023–19034. https://doi.org/10.1029/96JD01553.
Xie, P., Yatagai, A., Chen, M., Hayasaka, T., Fukushima, Y., Liu, C., Yang, S., 2007. A gauge-based analysis of daily precipitation over East Asia. J. Hydrometeorol. 8, 607–626. https://doi.org/10.1175/JHM583.1.
Yue, S., Pilon, P., Cavadias, G., 2002. Power of the Mann–Kendall and Spearman's rho tests for detecting monotonic trends in hydrological series. J. Hydrol. 259, 254–271. https://doi.org/10.1016/S0022-1694(01)00594-7.
Zelelew, M.B., Alfredsen, K., 2014. Transferability of hydrological model parameter spaces in the estimation of runoff in ungauged catchments. Hydrol. Sci. J. 59, 1470–1490. https://doi.org/10.1080/02626667.2013.838003.
Zhang, Y., Pan, M., Sheffield, J., Siemann, A.L., Fisher, C.K., Liang, M., Beck, H.E., Wanders, N., MacCracken, R.F., Houser, P.R., Zhou, T., Lettenmaier, D.P., Pinker, R.T., Bytheway, J., Kummerow, C.D., Wood, E.F., 2018. A climate data record (CDR) for the global terrestrial water budget: 1984–2010. Hydrol. Earth Syst. Sci. 22, 241–263. https://doi.org/10.5194/hess-22-241-2018.

Update

Remote Sensing of Environment

Volume 254, Issue , 1 March 2021, Page

DOI: https://doi.org/10.1016/j.rse.2020.112255

Remote Sensing of Environment 254 (2021) 112255

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

Corrigendum

Corrigendum to "Remotely sensed ensembles of the terrestrial water budget over major global river basins: An assessment of three closure techniques" [Remote Sensing of Environment Volume 252 (2021)]

R. Abolafia-Rosenzweig [a,*], M. Pan [c], J.L. Zeng [a], B. Livneh [a,b]

[a] Department of Civil, Environmental, and Architectural Engineering, University of Colorado Boulder, Boulder, CO 80309, USA
[b] Cooperative Institute for Research in Environmental Science (CIRES), University of Colorado Boulder, Boulder, CO 80309, USA
[c] Department of Civil and Environmental Engineering, Princeton University, Princeton, NJ 08544, USA

The authors regret that Fig. S6 appears in both the place of Fig. 6 and Fig. S6 in the manuscript: "Remotely sensed ensembles of the terrestrial water budget over major global river basins: An assessment of three closure techniques". In consequence, the intended Fig. 6 schematic does not appear in the published manuscript. We include the Fig. 6 schematic below.

The authors would like to apologise for any inconvenience caused.

DOI of original article: https://doi.org/10.1016/j.rse.2020.112191.
* Corresponding author.
E-mail address: Ronnie.abolafiarosenzweig@colorado.edu (R. Abolafia-Rosenzweig).

https://doi.org/10.1016/j.rse.2020.112255

Available online 22 December 2020

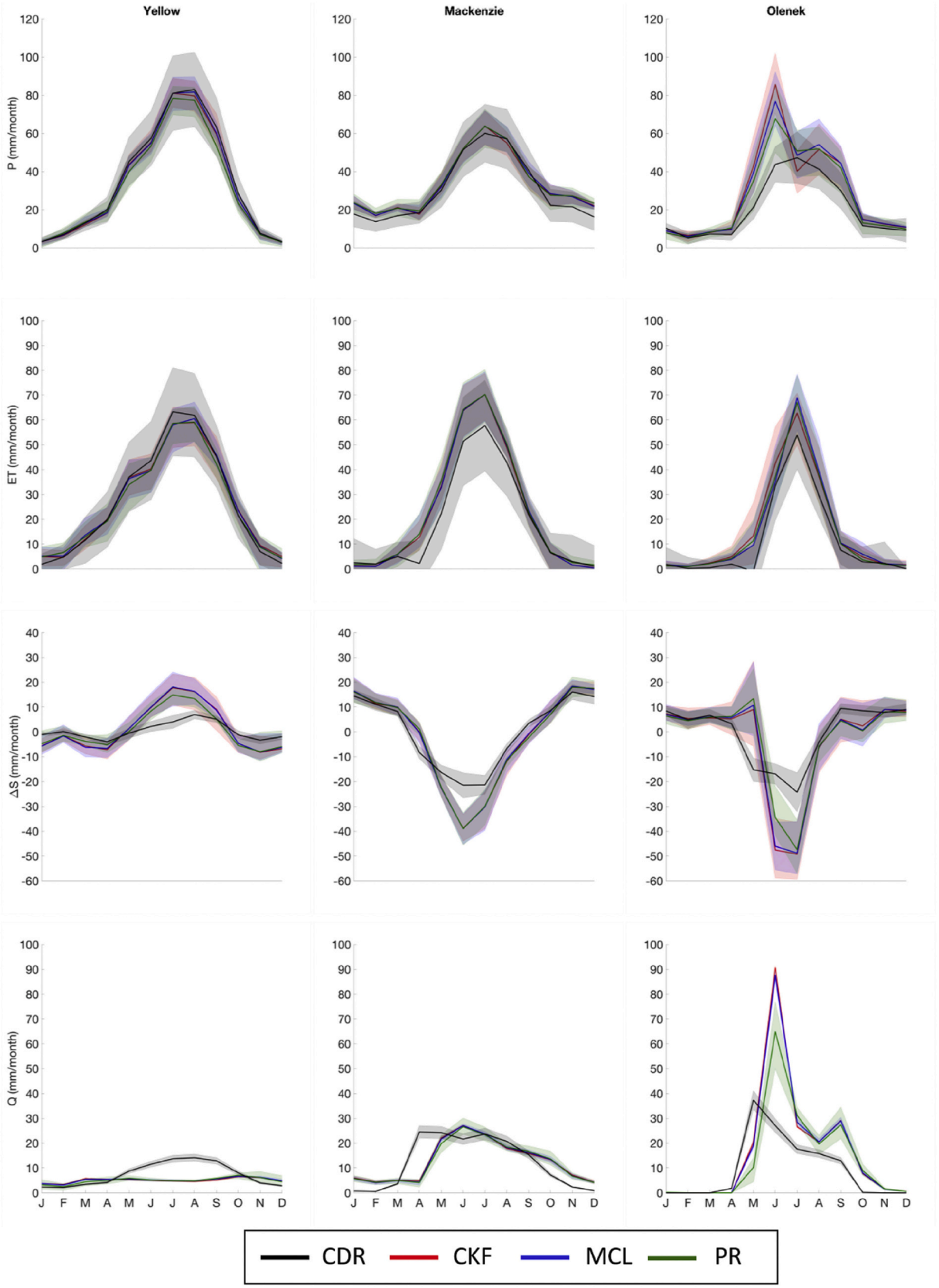

Fig. 6. Comparison of REESEN with the CDR for three representative dry basins: Seasonal cycles of balance-corrected REESEN water budget components compared to the Zhang et al. (2018) CDR (black) over 3 dry basins. REESEN includes the CKF (red), MCL (blue) and PR (green) closure methods, with uncertainty expressed as the ensemble range, whereas CDR uncertainty reflects the values reported in Zhang et al. (2018). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)