Conserving Land–Atmosphere Synthesis Suite (CLASS)

SANAA HOBEICHI, GAB ABRAMOWITZ, AND JASON EVANS

Climate Change Research Centre, and ARC Centre of Excellence for Climate Extremes, University of New South Wales, New South Wales, Sydney, Australia

(Manuscript received 24 January 2019, in final form 20 September 2019)

ABSTRACT

Accurate estimates of terrestrial water and energy cycle components are needed to better understand climate processes and improve models' ability to simulate future change. Various observational estimates are available for the individual budget terms; however, these typically show inconsistencies when combined in a budget. In this work, a Conserving Land–Atmosphere Synthesis Suite (CLASS) of estimates of simultaneously balanced surface water and energy budget components is developed. Individual CLASS variable datasets, where possible, 1) combine a range of existing variable product estimates, and hence overcome the limitations of estimates from a single source; 2) are observationally constrained with in situ measurements; 3) have uncertainty estimates that are consistent with their agreement with in situ observations; and 4) are consistent with each other by being able to solve the water and energy budgets simultaneously. First, available datasets of a budget variable are merged by implementing a weighting method that accounts both for the ability of datasets to match in situ measurements and the error covariance between datasets. Then, the budget terms are adjusted by applying an objective variational data assimilation technique (DAT) that enforces the simultaneous closure of the surface water and energy budgets linked through the equivalence of evapotranspiration and latent heat. Comparing component estimates before and after applying the DAT against in situ measurements of energy fluxes and streamflow showed that modified estimates agree better with in situ observations across various metrics, but also revealed some inconsistencies between water budget terms in June over the higher latitudes. CLASS variable estimates are freely available via <https://doi.org/10.25914/5c872258dc183>.

1. Introduction

A key role of the land surface in the climate system is to absorb solar and atmospheric radiation (surface net radiation R_n) and to partition this energy into latent heat flux (LH) and sensible heat flux H —both released into the atmosphere—and ground heat flux G transferred to the subsurface. The distribution of R_n over the three forms of energy (i.e., LH, G , and H) contributes to, and is partly affected by, precipitation over land P . Some precipitation infiltrates downward from the land surface and changes the soil moisture and the amount of groundwater stored in the aquifers or accumulates on the surface in the form of snow (denoted hereafter by ΔS).

A larger part of P is released into the air by evapotranspiration (ET) or flows across the land surface as runoff Q in a water body or a channel. The surface terrestrial water and energy budgets over a spatial domain (i.e., in situ, grid cell, river basin, continent, and global) are typically studied using the following equations for the water and energy balances, respectively:

$$P = ET + Q + \Delta S, \quad \text{and} \quad (1)$$

$$R_n = H + LH + G. \quad (2)$$

There have been many attempts to estimate each of the terms in Eqs. (1) and (2) that have resulted in multiple products for each term. There is also disagreement between the various estimates of each variable for many reasons, including their source (e.g., model, reanalysis, satellite or in situ measurement based), approach (e.g., physical, empirical, synthesis), assumptions, calibration, parameterization, imposed constraints, forcing, scale (e.g., in situ, regional, global), and

Supplemental information related to this paper is available at the Journals Online website: <https://doi.org/10.1175/JCLI-D-19-0036.s1>.

Corresponding author: Sanaa Hobeichi, s.hobeichi@unsw.edu.au

many other factors (Adler et al. 2001; Kato et al. 2007; Jiménez et al. 2011; Wang and Dickinson 2012; Best et al. 2015; Zhang et al. 2016; Skok et al. 2016; Beck et al. 2017; Liu et al. 2018).

Additionally, many studies evaluated the ability of the available observational estimates to close the budgets, and found that when the estimates of the different budget terms are combined in a single budget [i.e., Eq. (1) or (2)], they show inconsistency and generally do not solve the budget (Vörösmarty et al. 1989; Pan and Wood 2006; Sheffield et al. 2009; Trenberth et al. 2009; Azarderakhsh et al. 2011; Sahoo et al. 2011; Stephens et al. 2012; Vinukollu et al. 2012; Wild et al. 2013; Trenberth and Fasullo 2013; L'Ecuyer et al. 2015; Lorenz et al. 2015; Ferreira and Asiah 2015; Wild et al. 2015). On the other hand, hydrological, climate, and land surface models solve one or both budgets by design, and therefore provide coherent estimates of the budget terms (Su 2002; Betts et al. 2003; Mitchell 2004; Rodell et al. 2004; Sheffield and Wood 2007; Stackhouse et al. 2011; Reichle et al. 2011; Balsamo et al. 2015; Loew et al. 2016; Schellekens et al. 2017; Abatzoglou et al. 2018). However, these models are challenged by biases in the forcing datasets, physical assumptions, parameterization, or calibration (Sheffield et al. 2006; Alemohammad et al. 2015), which lead to error propagation in the components of the budgets and make some components less reliable than the others (Trenberth et al. 2007; Haddeland et al. 2011). This suggests that none of the existing estimates of budget terms is likely to be accurate.

Hybrid products have the potential to overcome the limitations of a single product and provide improved variable estimates. A wide range of techniques has been developed to create a hybrid product of one variable (Luo et al. 2007; Pan et al. 2012; Bishop and Abramowitz 2013; Aires 2014; Van Dijk et al. 2014; Lorenz et al. 2015; Yao et al. 2017; Jiménez et al. 2018; Herger et al. 2018; Munier and Aires 2018), typically by assigning weights to different products based on some measure of their performance. A lot of these techniques compute performance metrics as a function of uncertainties in the individual products. At the point scale, discrepancy with respect to the ground measurements gives an obvious avenue for defining uncertainty. However, at the gridded regional or global scale, this problem is much more difficult, and different techniques define uncertainties very differently. For instance, several studies assume that model outputs provide the “truth” for the budget terms at the regional or global gridded scale and use them as reference products to quantify the errors of nonmodel outputs (Pan et al. 2008; Sheffield et al. 2009; Sahoo et al. 2011). Conversely, Aires (2014) deems remote sensing (RS)-based estimates to be global observations

with predefined uncertainties. Alternatively, performance has been assessed by comparing various RS estimates of the same budget term (Munier et al. 2014). Rodell et al. (2015) for example use the deviation of a flux estimate from the mean of multiple estimates of the same budget term as a proxy to calculate its performance, while in the companion study, L'Ecuyer et al. (2015) use uncertainty estimates taken from the literature. Pan et al. (2012) use uncertainty estimates derived from previously developed formulas for relative uncertainties. More recently, Jiménez et al. (2018) use the available ground-truth observation to determine the uncertainty of gridded estimates.

A very brief description of the aforementioned studies is presented here since our purpose is only to highlight the differences in what these studies consider a “true observation” or “an ideal reference” to determine the uncertainty of the current estimates of the budget variables, with a view to determining the relative performance of each product. For this work, we define our performance metric of a single product by its ability to match direct in situ observations. Additionally, the merging approach that we adopt to derive a hybrid product or “best estimates” from multiple estimates of a single budget variable accounts for the error covariance between the individual products. Considering performance of the individual products as well as the dependence of errors in the merging process was shown to be more efficient in producing best estimates than relying on performance only (Herger et al. 2018). Our approach here is to assume that direct measurements provide the most reliable estimates of the budget terms and should be used to define the performance of model-based estimates. This of course requires an assumption that point scale in situ measurements can ultimately represent the grid scale, which, while not immediately obvious, has been demonstrated for ET (Hobeichi et al. 2018).

Despite the wide ranges in the uncertainty estimates of available flux products, there is consistency in how studies calculate the uncertainty in their derived hybrid product(s). This is almost always defined by computing the standard deviation of the constituent products (i.e., products that contributed to the final hybrid product). While this is understandable, it is not immediately clear what this range might represent. Without an attempt to define the independence of individual component products, it cannot be said to be the uncertainty in estimating the mean of constituent products, nor is this uncertainty defined by a discrepancy with observations. Here, we adopt the approach detailed in Hobeichi et al. (2018) and Hobeichi et al. (2019), which ensures that uncertainty estimates, while still global, accurately represent the hybrid product's agreement with point-scale observations where they exist. Below we argue that the

ability of uncertainty estimates obtained using this approach to successfully close the energy and water budgets is further evidence of its validity.

While hybrid products show a better ability to close budgets, they still do not achieve a zero residual. Realizing budget closure requires 1) the uncertainties of the budget terms to be well quantified, and 2) implementing a technique that takes into account these uncertainties to assimilate the individual budget terms in a complete budget (Stephens et al. 2012). To achieve requirement 2, two data assimilation techniques are commonly adopted to enforce the physical constraints. One is based on the ensemble Kalman filter (Kalman 1960; Pan and Wood 2006) where the residual of the budget is distributed back among all the budget terms based on the relative uncertainties of the individual terms and their correlation (Sahoo et al. 2011; Pan et al. 2012; Aires 2014; Munier et al. 2014; Lorenz et al. 2015). The other technique is based on variational data assimilation and the optimal estimation retrieval algorithm (Rodgers 2000; L'Ecuyer and Stephens 2002; Kalnay 2003). Similar to the previous method, this accounts for the relative uncertainties of the budget components, but also offers benefits over the previous method by allowing multiple linked budgets to be closed at the same time (Rodell et al. 2015; L'Ecuyer et al. 2015). In both cases, the uncertainty associated with each budget component will constrain the size of the adjustment used to close the budgets. Therefore, the success of any of these techniques in achieving coherent budget terms that are also physically realistic relies on the attribution of accurate uncertainties to each budget term.

Drawing inspiration from the work of L'Ecuyer et al. (2015) that achieves the closure of multiple budgets simultaneously and the work of Zhang et al. (2018) that solves the water budget at the global grid scale, this work aims at deriving global gridded monthly estimates of the components of both surface terrestrial water and energy budgets from existing estimates covering the period 2003–09. We offer improvement over the previous studies by

- using hybrid products, observationally constrained by ground observations, for each component variable where possible;
- ensuring that the uncertainty estimates of hybrid products reflect their deviation from observations and are therefore more defensible than those previously incorporated in algorithms that enforce physical constraints; and
- improving the derived hybrid products and their uncertainties through the additional constraint offered by solving water and energy balances simultaneously and at the grid scale.

To achieve our aim we employ our previously developed observationally constrained hybrid products Derived

Optimal Linear Combination Evapotranspiration, version 1.0 (DOLCE; Hobeichi 2017), and Linear Optimal Runoff Aggregate, version 1.0 (LORA; Hobeichi 2018), which provide flux and uncertainty estimates of ET and Q , respectively, and employ the same weighting technique as in DOLCE (Bishop and Abramowitz 2013; Abramowitz and Bishop 2015) to derive hybrid H , R_n , and G products. We argue that the computed uncertainties in the derived products are more justifiable and therefore address some of the shortcomings of the previous studies. These are merged using the data assimilation method detailed in L'Ecuyer et al. (2015) and Rodell et al. (2015) to further adjust the derived hybrid estimates of the budget components by enforcing the closure of the terrestrial water and energy budgets simultaneously. To our knowledge this is the first time that the water and energy budgets are solved simultaneously at the global monthly grid scale.

The paper is organized as follows: section 2 describes the participating global datasets and the observational networks. Section 3 presents the weighting method and details the assimilation technique. We then present our derived dataset and discuss our results in sections 4 and 5 before concluding.

2. Datasets

This study employs hybrid estimates of the components of the surface water and energy budgets where possible. This includes the previously developed synthesis ET and Q products, that is, DOLCE and LORA, respectively, and new hybrid estimates for R_n , H , and G that we develop in this study from a range of currently available global gridded estimates. We use Rainfall Estimates on a Gridded Network (REGEN; Contractor 2018), a newly developed gauge-based P dataset, and ΔS estimates derived from GRACE sensors. The employed datasets for both P and ΔS are considered single source products. We provide a brief description of the employed datasets. For more details about these datasets we refer the readers to the associated publications. A list of these datasets is displayed in Tables 1 and 2.

The current analysis spans the period 2003–09. We are restricted to this short period due to the availability of GRACE and DOLCE. All R_n , H , and G products are resampled on the same 0.5° grid using bilinear interpolation, while REGEN is regridded using a conservative interpolation method.

a. Deriving the components of the water budget

1) PRECIPITATION P

We compute monthly gridded precipitation estimates from REGEN, a new global daily precipitation product.

TABLE 1. Summary of the employed datasets for each of the water and energy budget terms. Items in bold represent observational networks that contributed to the employed datasets. See Table 2 for details on the employed datasets.

Budget variable	Source	Uncertainty	Budget variable	Source	Uncertainty
R_n	MERRA-2 Land surface diagnostics GLDAS-Noah NCEP-DOE ERA-Interim CERES-EBAF BSRN, FLUXNET2015-tier1, and LaThuile Free Fair Use	Observationally constrained	H	MERRA-2 Land surface diagnostics MERRA-2 Surface flux diagnostics GLDAS-Noah NCEP-DOE MPI-BGC, Princeton FLUXNET2015-tier1 and LaThuile Free Fair Use	Observationally constrained
G	MERRA-2 Land surface diagnostics MERRA-2 Surface flux diagnostics GLDAS-Noah NCEP-NCAR NCEP-DOE FLUXNET2015-tier1 and LaThuile Free Fair Use	Observationally constrained	P	REGEN GHCN and GPCC observational network	Observationally constrained
ET, LH	DOLCE derived from GLEAM-V2A GLEAM-V2B GLEAM-V3A MPI-BGC MOD16 PML FLUXNET2015-tier1 and LaThuile Free Fair Use	Observationally constrained	ΔS Q	GRACE Mascons-JPL LORA derived from HTESSEL JULES LISFLOOD PCR-GLOBWB SURFEX W3RA WaterGAP3 HBV-SIMREG USGS; GAGES-II, GRDB, and Dai (2016)	Quoted uncertainty Observationally constrained

REGEN has been derived by interpolating quality controlled in situ observations from multiple archives including the Global Historical Climatology Network (GHCN; Menne et al. 2012), the observational network hosted by the GPCC (Rudolf et al. 2003), and other smaller networks (Contractor et al. 2019). We used the provided Kriging error fields as uncertainty estimates for REGEN. Kriging error can be interpreted as the percentage of variance and depends on the availability of the contributing observation to a grid estimate and the size of the grid. In this dataset, gauge undercatch has not been corrected. However, during the process of deriving REGEN, the employed daily observation was compared with surrounding stations and historical data, and as a result of this, anomalous observations were flagged. This may have caught many but not all cases of gauge undercatch.

2) CHANGE IN WATER STORAGE ΔS

The change in water storage can be calculated from the monthly equivalent water thickness (EWT) of water storage anomalies (WSA) provided by GRACE sensors (Swenson and Wahr 2006; Swenson et al. 2006; Landerer and Swenson 2012; Watkins et al. 2015; Wiese et al. 2018). These account for the change in all water in a region, which could not be obtained completely from

another source (Van Dijk et al. 2014). The standard method for deriving WSA from the monthly changes in Earth's gravity field is by solving for gravity variations in terms of spherical harmonic coefficients. An alternative approach adopted by the Jet Propulsion Laboratory (JPL) is to use mass concentration blocks (mascons), which, unlike spherical harmonics, have known geophysical locations that can offer constraints during the data inversion process. Therefore, the mascon solution offers advantage over the spherical harmonics solution by being geophysically constrained in the sense that it does not require postprocessing to remove correlated error, free from latitudinal contamination and allows better separation of land and ocean signals. Details on the data processing for the spherical harmonic and mascon solutions can be found in Landerer and Swenson (2012) and Watkins et al. (2015), respectively. We use the latest GRACE mascon dataset (RL06M.MSCNv01).

Equations (3) and (4) are commonly used to calculate ΔS from WSA at any time step j . We choose to apply the backward difference equation [Eq. (4)] over the central difference equation [Eq. (3)] as the latter tends to smooth out the monthly changes. We fill the gap in WSA for June 2003 by applying Eq. (5), which calculates

TABLE 2. Source, resolution, and data access details of the employed datasets in this study.

Source	Temporal and spatial resolution	Temporal coverage	Data access
CERES-EBAF	Monthly 1°	March 2000– March 2018	https://eosweb.larc.nasa.gov/project/ceres/ebaf-surface_ed4.0_table
DOLCE	Monthly 0.5°	2000–09	http://dapds00.nci.org.au/thredds/catalog/ua8/ARCCSS_Data-9/v1-0/catalog.html
ERA-Interim	Monthly 0.5°	1979–present	http://apps.ecmwf.int/datasets/data/interim-mdfa/levtype=sfc/
GLDAS-Noah	Monthly 0.5°	2000–present	https://hydro1.gesdisc.eosdis.nasa.gov/data/GLDAS/GLDAS_NOAH025_M.2.1/
GRACE Tellus	Monthly 0.5°	March 2002–present	ftp://podaac-ftp.jpl.nasa.gov/allData/tellus/L3/mascon/RL06/JPL/
LORA	Monthly 0.5°	1980–2012	http://dapds00.nci.org.au/thredds/catalogs/ua8/LORA/LORA.html
MERRA-2 Land surface diagnostics	Monthly 0.5° × 0.625°	1980–present	https://goldsmr4.gesdisc.eosdis.nasa.gov/data/MERRA2_MONTHLY/M2TMNXLND.5.12.4/
MERRA-2 Surface flux diagnostics	Monthly 0.5° × 0.625°	1980–present	https://goldsmr4.gesdisc.eosdis.nasa.gov/data/MERRA2_MONTHLY/M2TMNXFLX.5.12.4/
MPI-BGC	Monthly 0.5°	December 1982– November 2011	https://vesgint-data.ipsl.upmc.fr/thredds/dodsC/IPSLFS/igcmg/IGCM/ReferenceDatasets/ts/land/mo/fluxsens/fluxsens_Lmon_EnsembleHcor_observations_198201-201112.nc.html
NCEP–NCAR	Monthly	1948–present	https://www.esrl.noaa.gov/psd/data/gridded/data.ncep.reanalysis.surfaceflux.html
NCEP–DOE	T62 Gaussian grid with 192 × 94 points	1979–present	https://www.esrl.noaa.gov/psd/data/gridded/data.ncep.reanalysis2.surface.html
Princeton- <i>H</i>	Daily 0.5°	1984–2007	Princeton University, New Jersey
REGEN	Daily 1°	1950–2018	https://doi.org/10.25914/5b9fa55a8298c
BSRN	Minutely In situ	Varies by station	https://dataportals.pangaea.de/bsrn/
FLUXNET 2015-tier1	Daily In situ	Varies by station	https://fluxnet.ornl.gov/
LaThuile Free Fair Use 2007	Daily In situ	Varies by station	http://www.fluxdata.org/DataDownload/default.aspx

$WSA^{June2003}$ as the product of WSA from the previous month (i.e., May) and the ratio of WSA climatology for May and June over 2002–16:

$$\Delta S^j = (WSA^{j+1} - WSA^{j-1})/2, \quad (3)$$

$$\Delta S^j = WSA^j - WSA^{j-1}, \quad \text{and} \quad (4)$$

$$WSA^{June2003} = \frac{WSA^{May2003} \times \overline{WSA^{June}}}{\Delta S^{May}}. \quad (5)$$

This guarantees temporal continuity in ΔS estimates. We note that EWT data from GRACE data are available from 2002; however, we exclude 2002 data due to the presence of many gaps.

The change in water storage computed following these steps is denoted hereafter as GRACE. We will use GRACE as the best estimate of ΔS in Eq. (1).

Since we have no independent verification data source, we have little choice to calculate the time-variant uncertainty associated with ΔS^j at every grid from the uncertainty of WSA provided by the data producer. To achieve this, we first calculate the relative uncertainties $|\text{unc}_{WSA^j}/WSA^j|$ and $|\text{unc}_{WSA^{j-1}}/WSA^{j-1}|$. Then we choose $\text{unc}_{\Delta S^j} = \min\{|\text{unc}_{WSA^j}/WSA^j|, |\text{unc}_{WSA^{j-1}}/WSA^{j-1}|\} \times \Delta S^j$ to represent the standard deviation uncertainty estimates of ΔS^j .

3) RUNOFF Q

We use LORA (Hobeichi et al. 2019) to provide global gridded monthly runoff. It is a hybrid product derived by merging 11 runoff estimates from eight hydrological models produced as part of the earth2Observe project (available via [ftp://wci.earth2observe.eu/](http://wci.earth2observe.eu/)), and constraining their weighting with observational streamflow records from around 600 downstream stations. Weights

account for both performance differences between models and error covariance between them. Uncertainty estimates are computed based on the discrepancy between the aggregated derived runoff and the associated observational streamflow dataset in each river basin. More details can be found in [Hobeichi et al. \(2019\)](#).

4) EVAPOTRANSPIRATION ET

The DOLCE product provides hybrid estimates for ET (and LH) and its associated uncertainty. It is derived from the combination of six global ET datasets including Global Land Evaporation Amsterdam Model (GLEAM), versions 2A and B ([Miralles et al. 2011](#)); GLEAM, version 3A ([Martens et al. 2017](#)); Max Planck Institute for Biogeochemistry (MPI-BGC; [Mueller et al. 2011](#)), MODIS Global Evapotranspiration Project (MOD16; [Mu et al. 2011](#)), and Penman–Monteith–Leuning (PML; [Zhang et al. 2015](#)). The uncertainty estimates are derived based on the discrepancy between the optimal hybrid ET flux estimate and observational data from 159 flux towers. We refer the reader to [Hobeichi et al. \(2018\)](#) for more detail.

b. Deriving the components of the energy budget

1) SURFACE NET RADIATION R_N

We derive a hybrid surface net radiation product from available estimates of surface net radiation using the same methodology as implemented in DOLCE ([Hobeichi et al. 2018](#)). The employed products include the Clouds and Earth's Radiant Energy Systems (CERES) Energy Balanced and Filled (EBAF) ([Loeb 2017](#)); ERA-Interim reanalysis ([Dee et al. 2011](#)); the Modern-Era Retrospective Analysis for Research and Applications, version 2, land surface diagnostics (MERRA-2; [GMAO 2015a](#)); the Global Land Data Assimilation System (GLDAS; [Beaudoin and Rodell 2016](#)) Noah land surface model; and the NOAA National Centers for Environmental Prediction (NCEP)–DOE AMIP II reanalysis 2 ([Kanamitsu et al. 2002](#)). We choose those datasets because of their availability and compatibility with the spatial and temporal coverage of this study. The observational dataset that constrains the merging approach is a composite of flux network (FLUXNET) 2015-Tier 1, LaThuile2007 Free Fair Use databases [[Baldocchi et al. 2001](#); [Baldocchi 2008](#); [Papale et al. 2012](#); Oak Ridge National Laboratory (ORNL) DAAC 2015] and Baseline Surface Radiation Network (BSRN; [Driemel et al. 2018](#)). We aggregate BSRN high-frequency observations, FLUXNET and LaThuile half-hourly records to daily values, and then to monthly means. We omit daily aggregates if less than 50% of records are available on that day and impose a minimum threshold of 15 daily observations to derive monthly

means. Also, we keep only the sites that fall within the temporal and spatial coverage of the study. As a result of this, only four BSRN sites and a total of 162 sites from FLUXNET2015 and LaThuile qualify for the analysis. A list of the contributing sites is available in the supplemental material (Table S1 in the online supplemental material).

2) SENSIBLE HEAT FLUX H

A similar approach is used to derive our hybrid sensible heat product and its associated uncertainty. We employ monthly sensible heat flux estimates from GLDAS and NCEP–DOE Reanalysis 2, MERRA-2 land surface diagnostics, MERRA-2 surface flux diagnostics ([GMAO 2015b](#)), MPI-BGC ([Jung et al. 2009](#)), and daily estimates developed by [Siemann et al. \(2018\)](#) that we will refer to as “Princeton- H .” The observational dataset that we use to constrain the merging approach consists of a total of 140 sites again from FLUXNET2015 and the LaThuile release (Table S1). In addition to the filtering applied to the FLUXNET2015 observations by the number of available records, we correct the observed sensible heat flux in LaThuile dataset for energy balance nonclosure based on the Bowen ratio method ([Bowen 1926](#)), as outlined in [Hobeichi et al. \(2018\)](#). Given that the Princeton- H dataset only covers up to 2007, this dataset does not contribute to the hybrid product over 2008–09. While we understand that this might have caused some temporal inconsistency, we expect that the assimilation technique will to some extent overcome this inconsistency in the final Conserving Land–Atmosphere Synthesis Suite (CLASS) product.

3) GROUND HEAT FLUX G

Again, the methodology in [Hobeichi et al. \(2018\)](#) is used to create the hybrid ground heat flux product. The same suite of land surface and reanalysis products as used as with sensible heat (i.e., excluding Princeton and MPI-BGC products) and with the addition of NCEP–NCAR ([Kalnay et al. 1996](#)) Reanalysis 1, and a total of 130 flux tower sites are employed to derive a hybrid ground heat flux and to compute its uncertainties. A map showing the locations of all the involved sites is illustrated in Fig. S1.

3. Methods

The purpose of this study is to derive monthly global gridded estimates of the complete surface water and energy budgets. The process for achieving this involves first implementing a merging approach to derive hybrid estimates for each of the budget terms. This results in new hybrid R_n , H , and G products that we derive in this

study and previously derived LH (or ET) and Q estimates. All these products also have uncertainty estimates associated with them that reflect their discrepancy with corresponding observational datasets.

The merging method was suggested by Bishop and Abramowitz (2013) and has already been implemented in this context by Hobeichi et al. (2018). It consists of building a linear combination of the participating datasets that minimizes the mean square difference with respect to a corresponding observational dataset. The linear combination for a given budget variable (e.g., R_n , G , and H) is expressed as

$$\boldsymbol{\mu}^j = \mathbf{w}^T \mathbf{x}^j = \sum_{k=1}^K w_k x_k^j, \quad (6)$$

where x_k^j is the j th time-space step of the k th bias-corrected product (i.e., after subtracting the mean error from the product). The term $\boldsymbol{\mu}$ is the derived hybrid product and the analytical solution that minimizes

$$\sum_{j=1}^J (\boldsymbol{\mu}^j - \mathbf{y}^j)^2, \quad (7)$$

where $j \in [1, J]$ are the time steps, $k \in [1, K]$ represent the participating datasets, and $\mathbf{y}^j$ is the j th observed time step of the budget variable. In Eq. (6), T denotes transpose. The weights $\mathbf{w}$ account both for performance differences as well as error covariance between products. More detail can be found in Bishop and Abramowitz (2013), Hobeichi et al. (2018), and Hobeichi et al. (2019).

An additional benefit of this technique is that it also allows computing the spatiotemporal uncertainty of $\boldsymbol{\mu}$. This is achieved by transforming the participating ensemble so that its variance about the best estimate $\boldsymbol{\mu}$ matches the error variance of $\boldsymbol{\mu}$ with respect to $\mathbf{y}$ for the time steps where observational data are available [following the ensemble transformation process detailed in Bishop and Abramowitz (2013)]. A validation of the computed uncertainty is provided in the supplemental material (Fig. S20).

As opposed to all the other budget terms (with the exception of ΔS), precipitation is the only single-source variable. The reason for this choice is that our hybrid approach requires weighting each participating dataset based on its performance with respect to in situ observations. That is, the improvement offered by the derived hybrid estimate over each of its constituent datasets results principally from the observationally constrained weighting. Given that we can access only a small subset of what contributes to deriving the REGEN product, we feel that REGEN will be more observationally

constrained than our merging approach. As an alternative to REGEN, other global gridded precipitation products such as GPCC (Schamm et al. 2016) and GPCP (Adler et al. 2018) could be used in the analysis. We plan to explore the sensitivity to these products in future analyses.

We then apply a data assimilation technique that is based on a theory detailed in L'Ecuyer and Stephens (2002). This technique modifies the budget term estimates (presented in previous section) based on their relative uncertainties to achieve budget closure while minimizing deviation from the initial estimates. We note that we apply a common time-variant spatial mask to all the employed products that we derive from the spatial intersection of all the constituent products (from Table 1, excluding the observational datasets).

In the following we do not intend to illustrate the theory that underpins the assimilation technique, but rather we demonstrate how we can apply it to achieve the budget closures, following the method adopted in L'Ecuyer et al. (2015) and Rodell et al. (2015).

Equations (1) and (2) can be rearranged so that the subject variables are R_n and ΔS . All the estimates have been converted to the same units, watts per square meter (W m^{-2}), which means that we can use ET and LH interchangeably. In this case, we can write

$$\Delta S = P - \text{LH} - Q, \quad \text{and} \quad (8)$$

$$R_n = H + \text{LH} + G. \quad (9)$$

Or, alternatively, $\mathbf{R} = \mathbf{A}\mathbf{F}$, where

$$\begin{aligned} \mathbf{R} &= \begin{pmatrix} \Delta S \\ R_n \end{pmatrix}, \\ \mathbf{A} &= \begin{pmatrix} 1 & -1 & -1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 \end{pmatrix}, \quad \text{and} \\ \mathbf{F} &= \begin{pmatrix} P \\ \text{LH} \\ Q \\ H \\ G \end{pmatrix}. \end{aligned}$$

Since the initial budget terms do not precisely balance, we denote our initial estimates of $\mathbf{F}$ and $\mathbf{R}$ by $\mathbf{F}_i$ and $\mathbf{R}_i$ (i refers to imbalance), where

$$\mathbf{F}_i = \begin{pmatrix} \text{REGEN} \\ \text{DOLCE} \\ \text{LORA} \\ \mu_H \\ \mu_G \end{pmatrix}, \quad \text{and}$$

$$\mathbf{R}_i = \begin{pmatrix} \overline{\text{GRACE}} \\ \mu_{R_n} \end{pmatrix}.$$

We aim to find new vectors, $\mathbf{F}_{\text{CLASS}}$ and $\mathbf{R}_{\text{CLASS}}$, respectively, that conserve the budget closure by adjusting each term in $\mathbf{F}_i$ based on their relative uncertainty in such a way that the deviation from $\mathbf{R}_i$ and $\mathbf{F}_i$ is minimized, when weighted by their uncertainty distributions. Mathematically, this situation can be represented by a joint probability, and $\mathbf{F}_{\text{CLASS}}$ and $\mathbf{R}_{\text{CLASS}}$ should maximize this joint probability:

$$\begin{aligned} \text{Probability}(\mathbf{F}_{\text{CLASS}}|\mathbf{F}_i, \mathbf{R}_i) &= \exp[-(\mathbf{F}_{\text{CLASS}} - \mathbf{F}_i)^T \\ &\times \mathbf{S}_{\mathbf{F}_i}^{-1}(\mathbf{F}_{\text{CLASS}} - \mathbf{F}_i)] \times \exp[-(\mathbf{R}_{\text{CLASS}} - \mathbf{R}_i)^T \\ &\times \mathbf{S}_{\mathbf{R}_i}^{-1}(\mathbf{R}_{\text{CLASS}} - \mathbf{R}_i)]. \end{aligned} \quad (10)$$

The terms $\mathbf{S}_{\mathbf{F}_i}$ and $\mathbf{S}_{\mathbf{R}_i}$ represent the error covariance of $\mathbf{F}_i$ and $\mathbf{R}_i$ derived from the uncertainty as described in the previous section.

The maximum of Eq. (10) can be achieved by maximizing the exponent term, or alternatively minimizing J_a , where

$$\begin{aligned} J_a &= (\mathbf{F}_{\text{CLASS}} - \mathbf{F}_i)^T \mathbf{S}_{\mathbf{F}_i}^{-1}(\mathbf{F}_{\text{CLASS}} - \mathbf{F}_i) \\ &+ (\mathbf{R}_{\text{CLASS}} - \mathbf{R}_i)^T \mathbf{S}_{\mathbf{R}_i}^{-1}(\mathbf{R}_{\text{CLASS}} - \mathbf{R}_i). \end{aligned} \quad (11)$$

The minimum in Eq. (11) satisfies

$$\frac{dJ_a}{d\mathbf{F}_{\text{CLASS}}} = 0. \quad (12)$$

Solving Eq. (12) gives coherent fluxes (i.e., $\mathbf{F}_{\text{CLASS}}$, $\mathbf{R}_{\text{CLASS}}$) that can close the water and energy budgets simultaneously, where

$$\mathbf{F}_{\text{CLASS}} = \mathbf{F}_i + \mathbf{S}_{\mathbf{F}_{\text{CLASS}}} \mathbf{K}^T \mathbf{S}_{\mathbf{R}_{\text{CLASS}}}^{-1} (\mathbf{R}_i - \mathbf{K}\mathbf{F}_i). \quad (13)$$

Here, $\mathbf{K}$ represents the Jacobian of $\mathbf{R}$ at $\mathbf{F}_i$ and $\mathbf{S}_{\mathbf{F}_{\text{CLASS}}} = (\mathbf{K}^T \mathbf{S}_{\mathbf{R}_i}^{-1} \mathbf{K} + \mathbf{S}_{\mathbf{F}_i}^{-1})^{-1}$ is the error covariance of $\mathbf{F}_{\text{CLASS}}$, while $\mathbf{S}_{\mathbf{R}_{\text{CLASS}}} = \mathbf{S}_{\mathbf{R}_i}$, indicating that the uncertainties of ΔS and R_n remain unchanged. This approach assumes that the uncertainty estimates are uncorrelated, unbiased, and are Gaussian. Under this assumption, the estimated $\mathbf{F}_{\text{CLASS}}$ and $\mathbf{R}_{\text{CLASS}}$ are the best possible fit to $\mathbf{F}_i$ and $\mathbf{R}_i$, respectively, given our uncertainty estimates $\mathbf{S}_{\mathbf{F}_i}$ and $\mathbf{S}_{\mathbf{R}_i}$.

The χ^2 test can then be employed to test these assumptions together with the appropriateness of the uncertainty estimates $\mathbf{S}_{\mathbf{F}_i}$ and $\mathbf{S}_{\mathbf{R}_i}$. This can be achieved by substituting the derived $\mathbf{F}_{\text{CLASS}}$ and $\mathbf{R}_{\text{CLASS}}$ in Eq. (11)

$$\begin{aligned} \chi^2 &= (\mathbf{F}_{\text{CLASS}} - \mathbf{F}_i)^T \mathbf{S}_{\mathbf{F}_i}^{-1}(\mathbf{F}_{\text{CLASS}} - \mathbf{F}_i) \\ &\times (\mathbf{R}_{\text{CLASS}} - \mathbf{R}_i)^T \mathbf{S}_{\mathbf{R}_i}^{-1}(\mathbf{R}_{\text{CLASS}} - \mathbf{R}_i). \end{aligned} \quad (14)$$

In this case, the quantity χ^2 measures the goodness of the fit. The value of χ^2 is ideally approximately equal to the degree of freedom, that is, the number of variables that define the two budgets, which in this case is seven. On the other hand values of χ^2 that are greater than seven indicate that the Gaussian distribution does not characterize the uncertainty estimates, and/or these uncertainties were too low, which might lead to derived estimates of $\mathbf{F}_{\text{CLASS}}$ or $\mathbf{R}_{\text{CLASS}}$ that lie at the extremes of the uncertainty distributions of $\mathbf{F}_i$ or $\mathbf{R}_i$, respectively. Values that are much smaller than seven indicate that the assumed uncertainties have been overestimated and correspondingly, the new derived uncertainty estimates $\mathbf{S}_{\mathbf{R}_{\text{CLASS}}}$ are also overestimated. We note that the term ‘‘uncertainty’’ is used in this study to represent the uncertainty standard deviation or error standard deviation. Under the assumption that the individual budget variables have a normal distribution and given that 68% of the probability under a normal curve lies between the mean and ± 1 standard deviation, the error of any budget variable is expected to be within the uncertainty bounds around 68% of the time. At any point in time and space and for any budget term, the uncertainty bounds or uncertainty range is defined by the interval [estimate – uncertainty, estimate + uncertainty].

4. Results

In this section, we present details of the best estimates of the individual budget terms (i.e., the derived μ_{R_n} , μ_H , and μ_G and the previously derived DOLCE and LORA along with REGEN and GRACE). We will refer to these products as the pre-data assimilation technique (pre-DAT) ensemble. We also present the adjusted budget terms that balance the two budgets. These constitute the CLASS ensemble and are denoted by CLASS- R_n , CLASS- H , CLASS-LH (or CLASS-ET), CLASS- G , CLASS- P , CLASS- Q , and CLASS- ΔS .

We first test the coherence of the pre-DAT ensemble by calculating the seasonal imbalances $P - ET - Q - \Delta S$ and $R_n - H - LH - G$ for the water and energy budgets, respectively, during the extreme seasons DJF and JJA. The seasonal plots of the individual budget terms (Figs. 1a–d and 2a–d) and the imbalance results (Figs. 1e and 2e) are displayed in Figs. 1 and 2 for the water and energy budgets, respectively. The uncertainties of each variable in the pre-DAT ensemble are illustrated in Figs. S2 and S3 for the water and energy budget terms, respectively. We note that all the calculated global averages are area weighted; that is, the contribution of each grid cell to the final average is relative to its area. This accounts for the differences in gridcell areas that are larger over the equator and smaller toward the poles.

FIG. 1. Seasonal averages of the water budget terms in the pre-DAT ensemble including (a) precipitation P , (b) evapotranspiration ET , (c) runoff Q , (d) change in water storage ΔS , and (e) the imbalance in the water budget $P - ET - Q - \Delta S$, calculated for the period 2003–09 during the extreme seasons prior to the application of the assimilation technique. DJF and JJA stand for December–February and June–August, respectively.

The plots in Fig. 1e show a small positive imbalance during DJF as the precipitation estimates in the pre-DAT ensemble exceed the sum of the remaining water budget terms over Argentina, the Sahel, South Asia, and tropical Australia. On the other hand, a negative imbalance caused by higher $ET + Q + \Delta S$ than P is seen over the Pacific mountains and valleys, tropical Africa, and the northern Andes. During JJA, a strong negative imbalance is dominant over Arctic Siberia and South

Asia, while a positive imbalance is seen over the Sahel, the Amazon, and the Central Siberian Plateau.

In Fig. 2e, a small negative imbalance is dominant over the Northern Hemisphere during DJF due to an excess of heat in R_n compared to $H + LH + G$, while a high positive imbalance is seen over South America excluding the Andes. During JJA, a positive imbalance is seen over most of the land particularly around the equator and in Central Asia.

FIG. 2. Seasonal averages of the energy budget terms in the pre-DAT ensemble including (a) net radiation R_n , (b) sensible heat H , (c) latent heat LH , (d) ground heat G , and (e) the imbalance in the energy budget $R_n - H - LH - G$, calculated for the period 2003–09, during the extreme seasons.

The data assimilation technique modifies the pre-DAT estimates to become CLASS datasets that satisfy the water and energy budgets. Figures 3 and 4 show seasonal plots of the CLASS datasets during DJF and JJA. Also, the DAT induces (by design) changes to the uncertainty estimates of H , G , LH (and ET), P , and Q . Figures S4 and S5 in the supplemental material display seasonal plots of the uncertainties of the CLASS datasets and show that the DAT has effectively reduced the uncertainties in the new budget estimates.

The χ^2 test provides a metric to identify the regions where budget closure was hard to achieve, in the sense that the induced adjustments to multiple components of the pre-DAT ensemble are toward the extremes of their uncertainty distributions due to a strong incoherence between the budget terms. Ideally, values should not exceed seven (i.e., the total number of budget terms, since ET and LH consist of a single variable). Small values indicate that small adjustments were required relative to each variable's initial uncertainty distribution,

FIG. 3. Seasonal averages of the water budget terms in the CLASS ensemble including (a) precipitation P , (b) evapotranspiration ET , (c) runoff Q , and (d) change in water storage ΔS , calculated for the period 2003–09 during the extreme seasons after the application of the assimilation technique.

while values greater than seven indicate that the budget terms underwent significant changes to achieve the closure. The χ^2 -test results that are bigger than seven will be referred to by “high χ^2 -test results” and indicate that at least two budget components were modified beyond one standard deviation of their uncertainty range, a result of incoherence between them. Figure 5 displays the results of the χ^2 test averaged for each month over 2003–09. The plots in Fig. 5 indicate that, in general, in July and December, budget closure is generally achieved with minimal change to the fluxes. In October, high χ^2 test results are achieved over Europe, while in May and June high χ^2 -test values are exhibited over Arctic Siberia. In general, high values are common over the Pacific valleys and mountains, the plains in South America and tropical Australia. In Fig. S1 we show the spatial distribution of the incoherence—the total number of times (out of 84 months) that high χ^2 -test values occur. This highlights

the places where the inconsistencies between the budget terms are persistent. Note that these difficulties are reflected in the updated uncertainties in the post-DAT CLASS product.

While the χ^2 test shows the places where the budget terms are incoherent, it does not indicate which components were responsible of this disagreement, nor the nature of the final adjustment. We therefore calculate the relative adjustment applied to each budget term as $\{[\text{budget term}(\text{CLASS}) - \text{budget term}(\text{pre-DAT})] / \text{uncertainty of budget term}(\text{pre-DAT})\} \times 100\%$. For any budget term and at any grid cell, a relative adjustment exceeding 100% or going below -100% indicates that the budget closure cannot be achieved within one standard deviation of the uncertainty estimate for that term. As mentioned earlier, in the ideal case, this is expected to occur at no more than 32% of the time, since the remaining 68% of values are expected to be within

FIG. 4. Seasonal area-weighted averages of the energy budget terms in the CLASS ensemble including (a) net radiation R_n , (b) sensible heat H , (c) latent heat LH , and (d) ground heat G , calculated for the period 2003–09 during the extreme seasons after the application of the assimilation technique.

one standard deviation of the mean (i.e., uncertainty estimate). We calculate the number of grid cells as a percentage of the total number of nonnull grid cells that underwent an adjustment exceeding their associated uncertainties. These are provided in Table 3 for each budget term and averaged for every month over the period of study. The results show that, for any budget term, the number of grid cells that underwent adjustments exceeding their associated uncertainties do not contribute to more than 24% of the total grid cells (after excluding null grid cells). These results are discussed in section 5. We present the relative adjustment maps of P , ET , Q , ΔS , R_n , H , and G averaged for every month over the period of study in Figs. S6–S12, respectively, and the relative adjustments applied to the uncertainties of P , Q , H , LH , and G in Figs. S13–S17, respectively. We note that the DAT does not modify the uncertainties of R_n and ΔS . The absolute differences in CLASS and pre-DAT fluxes

(i.e., CLASS – pre-DAT) are displayed in Figs. S18 and S19 for the water and energy budgets, respectively.

We also calculate four metrics of performance for R_n , H , LH , and G in both the pre-DAT ensemble and CLASS ensemble to test how well the budget terms agree with the corresponding in situ observations. These metrics are

(i) $RMSE = \sqrt{\text{mean}(\text{dataset} - \text{observation})^2}$, (ii) $SD \text{ difference} = \sigma_{\text{dataset}} - \sigma_{\text{observation}}$, (iii) $\text{correlation} = \text{corr}(\text{observation}, \text{dataset})$, and (iv) $\text{mean bias} = \text{mean}|\text{dataset} - \text{observation}|$. The pre-DAT and CLASS ensembles are compared to the in situ measurement from FLUXNET, LaThuile, and BSRN (for R_n only). The map in Fig. S1 shows the locations of these sites. Results are shown in Figs. 6a–d for R_n , H , LH , and G , respectively.

Given that most of the employed observational sites (more than 130) are located in grid cells where at least one high χ^2 -test result occurred, this comparison should

FIG. 5. Results of the χ^2 test averaged for each month across the period of study. Regions with values less or equal to seven indicate that the datasets of the pre-DAT ensemble are in good agreement with each other. The dark blue regions have values above seven and indicate that at least two budget terms highly disagree with the remaining budget terms.

give an indication of any decrease in the performance of budget terms in the CLASS datasets as a result of a significant adjustment of the pre-DAT ensemble. Figure 6 shows that the components of the energy budgets from the two ensembles are indeed very similar. In most of the cases, CLASS datasets perform similar to or better than the pre-DAT ensemble.

We also compare runoff estimates from pre-DAT and CLASS with streamflow observations at 11 river basins. The time series in Figs. 7 and 8 display the seasonal cycles of integrated runoff estimates over non-Siberian and Siberian basins, respectively, from pre-DAT Q and CLASS- Q as well as observed streamflow. These basins are selected in attempt to highlight regions where there are either significant differences or almost no differences in the dynamic cycle of pre-DAT Q and CLASS- Q as well as the impact of the differences on the agreement with observations.

It is notable from Fig. 7 that the modifications made to pre-DAT Q across five non-Siberian basins (i.e., Amazonas, Mississippi, Niger, Burdekin, and Maranon basins) has in some cases reduced the biases with respect

to observations and in other cases degraded the performance of pre-DAT Q .

Figure 8 shows that the drop in Q over the Siberian basins during spring is actually improved in CLASS- Q

TABLE 3. Percentage of grid cells that underwent an adjustment exceeding one standard deviation of the uncertainty estimate, calculated for each budget term and averaged for every month.

Month	Budget variable						
	P	ET	Q	ΔS	R_n	H	G
Jan	11	2	2	4	11	0	0
Feb	13	3	2	7	16	0	0
Mar	15	4	4	4	24	2	0
Apr	22	9	9	4	17	3	0
May	21	11	11	3	4	2	0
Jun	12	7	9	5	0	0	0
Jul	10	5	5	5	0	0	0
Aug	8	6	5	4	0	0	0
Sep	10	8	5	5	7	0	0
Oct	16	12	6	6	13	0	0
Nov	13	5	3	4	11	0	0
Dec	12	3	2	7	9	0	0

(a) Net radiation

(b) Sensible heat flux

(c) Latent heat flux

(d) Ground heat flux

FIG. 6. The results of the metric of performance for the estimates of budget terms (a) R_n , (b) H , (c) LH , and (d) G in the pre-DAT ensemble (red) and CLASS ensemble (purple) validated against in situ observations. The metrics of performance are $\text{RMSE} = \sqrt{\text{mean}(\text{dataset} - \text{observation})^2}$, $\text{SD difference} = \sigma_{\text{dataset}} - \sigma_{\text{observation}}$, $\text{correlation} = \text{corr}(\text{observation}, \text{dataset})$, and $\text{mean bias} = \text{mean}|\text{dataset} - \text{observation}|$.

FIG. 7. Monthly cycle of runoff aggregates from pre-DAT (i.e., LORA) and CLASS (i.e., CLASS- Q) compared with the observed streamflow over five non-Siberian basins.

relative to pre-DAT Q over all the basins. However, the subsequent June snowmelt peak is not well captured in the dynamics of CLASS- Q time series, as opposed to pre-DAT Q . This is discussed further in section 5.

5. Discussion

a. Enforcing the closure of two budgets versus one budget

Enforcing the closure of the coupled water and energy budgets offers additional constraint on the budget

variables compared to enforcing the closure of a single budget by allowing more variables to influence the scale of adjustments. In this part of the discussion, we refer to the relative adjustment plots in Figs. S6–S12 and examine regions where the adjustments made to one budget affect the nature of adjustments made to the components of the other budget. For example, over the Amazon, there are large water and energy imbalances in June and July, caused by an excess of $Q + ET + \Delta S$ and R_n , respectively. The energy balance enforcement is achieved after a decline in R_n accompanied by an increase in H and G . On

FIG. 8. Monthly cycle of runoff aggregates from pre-DAT (i.e., LORA) and CLASS (i.e., CLASS- Q) compared with the observed streamflow over six Siberian basins.

the other hand, the water balance is achieved after a significant increase in P accompanied by a decrease in Q and ΔS . From the energy balance perspective, LH might be expected to increase (similarly to H and G), while from the water budget perspective ET is expected to decrease (similarly to Q and ΔS). Interestingly, LH shows a variant behavior (drop and rise) over the region in order to maintain a balance in both budgets. If this study was solving a single budget only, we would have seen very different changes applied to ET than those achieved here.

In January, the energy components are in balance over Europe and North Australia, however, the water budget is initially imbalanced with an excess of P . To achieve a complete balance (i.e., in both budgets), the DAT reduces P and increases each of Q , ET, and ΔS . The energy balance is sustained by a slight decrease in H and G and an increase in R_n to neutralize the imbalance induced by the change in LH. This suggests that even when a single budget is in balance, enforcing additional constraints can further adjust the budget terms while maintaining the balance.

The adjustments applied in this study to the individual budget terms guarantee that the adjusted variables fulfil the physical constraints of both water and energy budgets and therefore are likely to be more reliable than those that solve a single budget only.

b. Disagreement between pre-DAT variables and induced adjustment

It follows from Table 3 that the pre-DAT datasets and their associated uncertainties are reasonable and allow enough flexibility to achieve the closure of the water and energy budgets over most of the land easily. This is concluded from the relatively low number of grid cells that require adjustments beyond one standard deviation of their uncertainty estimate—fewer than the $\sim 30\%$ we might expect. This suggests that the initial uncertainty estimates may have been conservative. The plots in Fig. 5 also highlight locations where the budget variables have to undergo adjustments of magnitudes that exceed one standard deviation, indicating that initially the budget terms disagree with each other. The plots in Figs. S6–S12 provide additional details concerning the nature of the adjustments [i.e., a significant increase ($>100\%$) or a significant decrease ($<100\%$)], and to which budget terms such significant adjustments are applied. For instance, in the high latitudes, P is the most significantly modified variable throughout the year and hence is initially the most incoherent with the other budget terms. This is perhaps because the rain gauge networks that contribute to developing the pre-DAT P (i.e., REGEN) lack density in these regions and because gauge undercatch has not been accounted for. The

second most modified variable is R_n , particularly over the mid- and high latitudes of the Northern Hemisphere where R_n exhibits considerable increase in many regions. On the other hand, H exhibits a significant increase over most of the Southern Hemisphere during March to May, indicating that the original hybrid H estimates have been underestimated. In October, LH exhibits a significant increase over Europe and the Siberian plains. This coincides with a drop in P in these regions. In many areas over the Amazon, both LH and P exhibit significant changes and contribute to high χ^2 values. The inconsistency of these two variables with the other budget terms is possibly owing to the lack of direct flux observations in these particular ecosystems.

c. CLASS- Q negative bias in June over the high latitudes

Hobeichi et al. (2019) note that LORA (i.e., pre-DAT Q) is likely to overestimate runoff in the high latitudes. It is therefore expected that enforcing the closure of the water and energy budgets will remove excess water from pre-DAT Q and will provide new runoff estimates that are in better agreement with the observed streamflow. Over the high latitudes, this holds true in spring only, when CLASS- Q is in better agreement with the observed streamflow compared to pre-DAT Q . However, the drop, seen in Lena, Olenek, Ob, Yana, and Yenisei (Fig. 8), has an unwelcome effect on the summer runoff. Similarly, Lorenz et al. (2015) finds that the induced change to runoff estimates by various assimilation techniques dampens the runoff maxima during summer in Yenisei, Lena, and Olenek and provides worse estimates in Ob. To understand why CLASS- Q (as opposed to pre-DAT Q) underestimates the summer peak particularly in June, we inspect the seasonal cycles of all the hydrologic variables in both pre-DAT and CLASS at the Siberian basins (Fig. 9). We exclude the Amur basin from our analysis since it is not impacted by the drop of runoff in June.

We also examine three possible explanations for the induced change in June: First, the adjustments made to the energy budget terms to achieve the energy balance can influence the changes brought to Q through ET. In particular, a rise in ET is very likely to be counterbalanced by a drop in Q . However, Fig. 9 shows that the magnitude of ET does not rise after applying the DAT in the Siberian basins, which means that no change in ET contributes to the drop in Q . A second explanation for this could be underestimation of June rainfall. As noted earlier, the employed P dataset does not account for wetting loss and gauge undercatch, which is a considerable source of error over the high latitudes. This might have caused P to be underestimated in June, during

FIG. 9. Monthly cycle of water budget variable aggregates from pre-DAT and CLASS compared with the observed streamflow over five Siberian basins.

which a marked shift in rainfall levels occurs. It is notable that the DAT achieves an increase in precipitation over Yana and Olenek basins; however, we think that despite the induced increase in precipitation, CLASS- P still underestimates P in June (and possibly in July), accounting for the negative bias in CLASS- Q . A third possible explanation is an underestimation of the deficit in water storage during June. Since June is the beginning of the hot season, snowmelt largely contributes to runoff and results in a deficit in the water storage. Accordingly, if ΔS does not represent the storage deficit well in June, likely due to inaccurate separation of land and ocean portions of mass within land/ocean mascon, there will also be an underestimation in Q . The reason why this is more appearing in June is the large magnitude of ΔS in this month. Yet, the lack of direct ΔS observations does

not allow us to assess this assumption. Finally, we think that the underestimation of June P and water storage deficit is likely to explain the large negative bias of CLASS- Q in June.

It follows from this analysis that, besides providing balanced estimates of the components of the surface terrestrial water and energy budgets, this work offers a framework for the evaluation of any estimates of a budget variable as well as a tool for comparing different estimates of the same variable. For example, by repeating the current analysis each time with a different precipitation dataset, the results of the χ^2 tests can depict areas where a particular precipitation dataset is worse than the others (this is the case where a particular dataset leads to high χ^2 values that are not produced when other datasets were employed). Further, the signs

TABLE 4. A comparison of mean annual hydrologic fluxes (mm yr^{-1}) of 20 major basins for CLASS and CDR (Zhang et al. 2018) calculated over the period 2003–09 for CLASS and 1984–2010 for CDR. The mean annual uncertainty values associated with CLASS components are shown. Values in bold indicate that CDR estimates are outside the range of CLASS estimates.

Basin	CLASS- <i>P</i>	CDR- <i>P</i>	CLASS-ET	CDR-ET	CLASS- <i>Q</i>	CDR- <i>Q</i>	CLASS- ΔS
Amazon	2265 $\pm$ 343	2182	1153 $\pm$ 245	1153	1096 $\pm$ 296	1029	16
Amur	470 $\pm$ 66	424	322 $\pm$ 99	295	171 $\pm$ 81	129	–23
Columbia	542 $\pm$ 29	624	326 $\pm$ 111	331	254 $\pm$ 95	293	–38
Congo	1425 $\pm$ 250	1449	1046 $\pm$ 187	1045	362 $\pm$ 124	404	17
Danube	751 $\pm$ 48	768	508 $\pm$ 93	503	245 $\pm$ 86	265	–2
Indigirka	281 $\pm$ 95	258	128 $\pm$ 114	138	163 $\pm$ 107	120	–10
Indus	403 $\pm$ 90	425	326 $\pm$ 129	277	67 $\pm$ 67	148	10
Kolyma	299 $\pm$ 79	283	165 $\pm$ 104	167	158 $\pm$ 84	116	–24
Lena	411 $\pm$ 88	379	212 $\pm$ 110	245	208 $\pm$ 103	134	–9
Mackenzie	392 $\pm$ 63	413	236 $\pm$ 99	241	165 $\pm$ 68	173	–9
Mississippi	780 $\pm$ 42	792	552 $\pm$ 106	577	208 $\pm$ 72	215	18
Murray–Darling	453 $\pm$ 29	452	419 $\pm$ 86	411	9 $\pm$ 5	41	25
Niger	594 $\pm$ 103	595	452 $\pm$ 130	401	100 $\pm$ 46	194	42
Northern Dvina	607 $\pm$ 101	618	315 $\pm$ 111	324	291 $\pm$ 120	294	2
Ob	445 $\pm$ 76	470	306 $\pm$ 95	323	153 $\pm$ 52	147	–13
Olenek	297 $\pm$ 84	280	160 $\pm$ 110	174	139 $\pm$ 114	106	–2
Paraná	1135 $\pm$ 143	1171	903 $\pm$ 156	892	191 $\pm$ 89	279	41
Pechora	626 $\pm$ 131	552	237 $\pm$ 120	244	386 $\pm$ 188	308	4
Yenisei	437 $\pm$ 95	460	244 $\pm$ 114	265	204 $\pm$ 118	195	–11
Yukon	358 $\pm$ 117	314	201 $\pm$ 101	175	204 $\pm$ 71	139	–46

and magnitudes of the relative adjustments brought by the DAT to each of the assessed datasets offer insights into regions where *P* is over- or underestimated. This is of great importance, especially in ungauged regions where it is difficult to evaluate the quality of global datasets.

It is worth mentioning that the approach applied in this study is flexible and can be improved in at least two ways. 1) In the generation of the hybrid budget terms (by incorporating more observational data or additional global estimates of the individual budget terms), and also 2) by adding new budget equations and linking them to the surface terrestrial water and energy budgets through a relevant budget term (e.g., by adding a carbon budget).

The shortcomings of this work arise from the short analysis period, that is, covering 2003–09 only and the spatial gaps in the derived product, which was inherent from the datasets employed in this study (Table 1). Because of this, we think that the seasonal global mean values shown in Figs. 1–4 are not very suitable for comparisons with results from other studies unless the same spatial mask is applied. Additionally, as discussed earlier in Hobeichi et al. (2018), the flux tower sites that are used to constrain the weighting of the energy budget terms are not evenly distributed across the various landscapes and do not represent all ecosystems. This also holds true for the rain gauges and stream gauges and has resulted in large biases in the water budget terms during June over the high latitudes.

Furthermore, the differences in the spatial resolution of the employed products (see Table 1) might have also

affected the values of the budget terms in the pre-DAT ensemble, and subsequently the balanced CLASS estimates, this holds especially true for REGEN dataset, which has a coarser footprint than the other budget terms.

Moreover, the DAT is based on the relative uncertainty estimates associated with the budget terms. Hence, the more uncertain the terms are, the stronger adjustments they are likely to receive. In the high latitudes, particularly in the five Siberian basins, the relative uncertainty of runoff (not shown here) is generally the highest among all the budget terms and accordingly is changed the most, while ΔS , which has the smallest relative uncertainty, is only slightly adjusted. This likely results in biases in the remaining water budget terms. Lorenz et al. (2015) suggests that when the budget terms of the water cycle are uncertain, enforcing the budget closure can lead to artifacts in the produced estimates. We agree with this, but further suggest that even if the budget terms are uncertain, the closure of the budgets can be perfectly achieved if the uncertainties in all the budget terms are equally well represented. However, uncertain budget terms with underestimated uncertainties might not receive the appropriate adjustments at the cost of the budget terms, which have well-presented but larger uncertainty estimates.

We also note that the application of the DAT has occasionally led to nonphysical values such as negative runoff or precipitation pixels, we therefore apply additional filters so that no such values appear in CLASS datasets.

FIG. 10. Annual continental averages of R_n , H , LH , and G developed in CLASS (red) and L'Ecuyer et al. (2015) (blue). The midpoint of the line segment shows the flux value while the endpoints represent the values of flux $\pm$ uncertainty.

d. Comparison with previous studies

We compare the derived CLASS hydrologic fluxes with balanced hybrid estimates of water cycle climate data record (CDR) developed in a recent study (Zhang et al. 2018) over a range of river basins. The main reasons for choosing this particular study is the common spatial (i.e., 0.5°) and temporal (i.e., monthly) resolutions despite the difference in temporal coverage that is 1984–2010 in their study while 2003–09 in the present study as well as the spatial gaps in the present study. Table 4 displays the mean annual estimates in millimeters per year (mm yr^{-1}) for CLASS components and those derived in Zhang et al. (2018) that are coded CDR- P , CDR-ET, CDR- Q , and CDR- ΔS for precipitation, evapotranspiration, runoff, and change in water storage, respectively. Zhang et al. (2018) applied additional constraints on the balanced hydrologic fluxes in a postprocessing step to ensure that the mean CDR- ΔS across each basin is equal to zero. This explains the differences with CLASS- ΔS over most of the basins. We have not implemented this additional constraint as we are looking at a shorter period, and interannual

variability can be large in some locations, especially in basins that experienced drought events (e.g., millennium drought in the Murray–Darling basin during 2002–08, long-term drought in the Congo basin since 2000, and severe drought over the Niger basin between 2000 and 2008). Due to the high anthropogenic water management in these regions that were applied during the drought events, these regions have witnessed a massive decline in the runoff that contribute to streamflow accompanied with an increase in water storage. Table 4 shows that CDR- P , CDR-ET, and CDR- Q lie within the uncertainty bounds of CLASS (i.e., between CLASS – uncertainty and CLASS + uncertainty) over all the basins except Columbia (P), Murray–Darling (P and Q), Indus (Q), and Niger (Q).

We compare CLASS energy fluxes with continental annual means of R_n , H , and LH developed in the study of L'Ecuyer et al. (2015) over the period 2000–09. As noted earlier, the same DAT implemented here is also applied in L'Ecuyer et al. (2015). In their study, the adjustment of the energy fluxes is constrained by the physical constraints of surface water and energy budgets as well as atmospheric water and energy budgets. Also,

FIG. 11. Annual continental averages of P , ET , Q , and ΔS developed in CLASS (red) and Rodell et al. (2015) (blue). The midpoint of the line segment shows the flux value while the endpoints represent the values of flux $\pm$ uncertainty.

the annual ground heat flux is considered very small over the continents and therefore neglected. We present in Fig. 10 the range of annual continental averages of R_n , H , LH , and G developed in CLASS (in red) and in L'Ecuyer et al. (2015) (in blue). The plots show that CLASS has smaller fluxes with larger associated uncertainties compared to those developed in L'Ecuyer et al. (2015). Overall, the range of their fluxes overlap with the upper range of CLASS energy fluxes over all continents.

Similarly, we compare CLASS water fluxes with continental annual means of P , ET , Q , and ΔS developed in the study of Rodell et al. (2015). In their study, the DAT is applied to derive coherent estimates of annual surface and atmospheric water budgets over the continents and oceans, and the annual mean ΔS over the period 2000–09 is neglected. We present in Fig. 11 the range of annual continental averages of P , ET , Q , and ΔS developed in CLASS (in red) and in Rodell et al. (2015) (in blue). Similar to our conclusions from the comparison with L'Ecuyer et al. (2015), the plots show that CLASS has slightly smaller fluxes with larger associated uncertainties compared to those developed in Rodell et al. (2015) and that in most cases, the range of their optimized fluxes

overlaps with the upper range of CLASS water components across most of the components. Large discrepancies between the two datasets are seen in Australia, where CLASS shows lower precipitation and runoff averages. However, our precipitation and runoff results agree better with findings from the Australian Water Availability Project (AWAP; Raupach et al. 2009) project. Also, over Eurasia, CLASS precipitation is lower than that estimated in Rodell et al. (2015). This is likely to be caused by gauge undercatch that is not accounted for in REGEN, which tends to be very large particularly during the snowy season.

6. Conclusions

This study develops conserving monthly estimates of the terrestrial water and energy budget components at a 0.5° grid scale over 2003–09.

Enforcing the closure of the water and energy budget has not radically changed the optimal estimates in the derived hybrid products, while ensuring that they are balanced with each other. Also, over most of the land, the magnitude of the adjustments required to enforce the budget is typically relatively small. On the other

hand, the implemented approach highlights particular regions where the optimal estimates have been initially least compatible, likely due to a lack of observational data there.

Furthermore, this study provides a framework for evaluating the quality of individual water or energy budget datasets and a tool for comparing various estimates of the same budget variable. This is of great importance given that direct observations, which serve to evaluate global estimates, are lacking in many regions on land.

Overall, the methods presented here starting with the merging approach that we use to develop hybrid estimates for the individual budget terms and to compute their uncertainties, and through applying the DAT to enforce the simultaneous closure of the water and energy budgets are likely to provide robust estimates of the budgets' components at the grid scale with defensible uncertainty bounds that are observationally constrained and broadly consistent across different variables in the budgets.

Expanding the time periods of CLASS products and filling the gaps in its spatial coverage will be carried out in the future.

Acknowledgments. The authors acknowledge the support of the Australian Research Council Centre of Excellence for Climate Extremes (CE170100023). NCEP Reanalysis data provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, from their website <https://www.esrl.noaa.gov/psd/> in any documents or publications using these data. NCEP–DOE 2 Reanalysis data provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, from their website <https://www.esrl.noaa.gov/psd/> in any documents or publications using these data. GRACE land data are available at <http://grace.jpl.nasa.gov>, supported by the NASA MEaSUREs Program. This research was undertaken with the assistance of resources and services from the National Computational Infrastructure (NCI), which is supported by the Australian Government. This work used eddy covariance data acquired and shared by the FLUXNET community, including these networks: AmeriFlux, AfriFlux, AsiaFlux, CarboAfrica, CarboEuropeIP, CarboItaly, CarboMont, ChinaFlux, Fluxnet-Canada, GreenGrass, ICOS, KoFlux, LBA, NECC, OzFlux-TERN, TCOS-Siberia, and USCCC. The ERA-Interim reanalysis data are provided by ECMWF and processed by LSCE. The FLUXNET eddy covariance data processing and harmonization was carried out by the European Fluxes Database Cluster, AmeriFlux Management Project, and Fluxdata project of FLUXNET, with the support of CDIAC and ICOS Ecosystem Thematic Center, and the OzFlux, ChinaFlux, and AsiaFlux offices. This

work also used eddy covariance data acquired by the FLUXNET community and in particular by the following networks: AmeriFlux [U.S. Department of Energy, Biological and Environmental Research, Terrestrial Carbon Program (DE-FG02–04ER63917 and DE-FG02–04ER63911)], AfriFlux, AsiaFlux, CarboAfrica, CarboEuropeIP, CarboItaly, CarboMont, ChinaFlux, Fluxnet-Canada (supported by CFCAS, NSERC, BIOCAP, Environment Canada, and NRCan), GreenGrass, KoFlux, LBA, NECC, OzFlux, TCOS-Siberia, and USCCC. We acknowledge the financial support to the eddy covariance data harmonization provided by CarboEuropeIP, FAO-GTOS-TCO, iLEAPS, Max Planck Institute for Biogeochemistry, National Science Foundation, University of Tuscia, Université Laval and Environment Canada, and U.S. Department of Energy and the database development and technical support from Berkeley Water Center, Lawrence Berkeley National Laboratory, Microsoft Research eScience, Oak Ridge National Laboratory, University of California, Berkeley, and the University of Virginia. The authors declare that they have no conflicts of interest.

REFERENCES

- Abatzoglou, J. T., S. Z. Dobrowski, S. A. Parks, and K. C. Hegewisch, 2018: TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015. *Sci. Data*, **5**, 170191, <https://doi.org/10.1038/sdata.2017.191>.
- Abramowitz, G., and C. H. Bishop, 2015: Climate model dependence and the ensemble dependence transformation of CMIP projections. *J. Climate*, **28**, 2332–2348, <https://doi.org/10.1175/JCLI-D-14-00364.1>.
- Adler, R. F., C. Kidd, G. Petty, M. Morissey, and H. M. Goodman, 2001: Intercomparison of global precipitation products: The Third Precipitation Intercomparison Project (PIP-3). *Bull. Amer. Meteor. Soc.*, **82**, 1377–1396, [https://doi.org/10.1175/1520-0477\(2001\)082<1377:IOGPPT>2.3.CO;2](https://doi.org/10.1175/1520-0477(2001)082<1377:IOGPPT>2.3.CO;2).
- , and Coauthors, 2018: The Global Precipitation Climatology Project (GPCP) monthly analysis (new version 2.3) and a review of 2017 global precipitation. *Atmosphere*, **9**, 138, <https://doi.org/10.3390/atmos9040138>.
- Aires, F., 2014: Combining datasets of satellite-retrieved products. Part I: Methodology and water budget closure. *J. Hydrometeorol.*, **15**, 1677–1691, <https://doi.org/10.1175/JHM-D-13-0148.1>.
- Alemohammad, S. H., D. B. McLaughlin, and D. Entekhabi, 2015: Quantifying precipitation uncertainty for land data assimilation applications. *Mon. Wea. Rev.*, **143**, 3276–3299, <https://doi.org/10.1175/MWR-D-14-00337.1>.
- Azarderakhsh, M., W. B. Rossow, F. Papa, H. Norouzi, and R. Khanbilvardi, 2011: Diagnosing water variations within the Amazon basin using satellite data. *J. Geophys. Res.*, **116**, D24107, <https://doi.org/10.1029/2011JD015997>.
- Baldocchi, D., 2008: “Breathing” of the terrestrial biosphere: Lessons learned from a global network of carbon dioxide flux measurement systems. *Aust. J. Bot.*, **56** (1), 1–26, <https://doi.org/10.1071/BT07151>.

- , and Coauthors, 2001: FLUXNET: A new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities. *Bull. Amer. Meteor. Soc.*, **82**, 2415–2434, [https://doi.org/10.1175/1520-0477\(2001\)082<2415:FANTTS>2.3.CO;2](https://doi.org/10.1175/1520-0477(2001)082<2415:FANTTS>2.3.CO;2).
- Balsamo, G., and Coauthors, 2015: ERA-Interim/Land: A global land surface reanalysis data set. *Hydrol. Earth Syst. Sci.*, **19**, 389–407, <https://doi.org/10.5194/hess-19-389-2015>.
- Beaudoing, H. K., and M. Rodell, 2016: GLDAS Noah Land Surface Model L4 monthly 0.25×0.25 degree V2.1. NASA Goddard Earth Science Data and Information Services Center, accessed 31 May 2017, <http://doi.org/10.5067/SXAVCZFAQLNO>.
- Beck, H. E., A. I. J. M. van Dijk, A. de Roo, E. Dutra, G. Fink, R. Orth, and J. Schellekens, 2017: Global evaluation of runoff from 10 state-of-the-art hydrological models. *Hydrol. Earth Syst. Sci.*, **21**, 2881–2903, <https://doi.org/10.5194/hess-21-2881-2017>.
- Best, M. J., and Coauthors, 2015: The plumbing of land surface models: Benchmarking model performance. *J. Hydrometeorol.*, **16**, 1425–1442, <https://doi.org/10.1175/JHM-D-14-0158.1>.
- Betts, A. K., J. H. Ball, and P. Viterbo, 2003: Evaluation of the ERA-40 surface water budget and surface temperature for the Mackenzie River basin. *J. Hydrometeorol.*, **4**, 1194–1211, [https://doi.org/10.1175/1525-7541\(2003\)004<1194:EOTESW>2.0.CO;2](https://doi.org/10.1175/1525-7541(2003)004<1194:EOTESW>2.0.CO;2).
- Bishop, C. H., and G. Abramowitz, 2013: Climate model dependence and the replicate Earth paradigm. *Climate Dyn.*, **41**, 885–900, <https://doi.org/10.1007/s00382-012-1610-y>.
- Bowen, I. S., 1926: The ratio of heat losses by conduction and by evaporation from any water surface. *Phys. Rev.*, **27**, 779–787, <https://doi.org/10.1103/PhysRev.27.779>.
- Contractor, S., 2018: Rainfall estimates on a gridded network based on long-term station data v1.0. NCI Australia, accessed 30 January 2019, <https://doi.org/10.25914/5b9fa55a8298c>.
- , and Coauthors, 2019: Rainfall Estimates on a Gridded Network (REGEN)—A global land-based gridded dataset of daily precipitation from 1950–2013. *Hydrol. Earth Syst. Sci. Discuss.*, <https://doi.org/10.5194/hess-2018-595>.
- Dai, A., 2016: Historical and future changes in streamflow and continental runoff: A review. *Terrestrial Water Cycle and Climate Change: Natural and Human-Induced Impacts*, *Geophys. Monogr.*, Vol. 221, Amer. Geophys. Union, 17–37, <https://doi.org/10.1002/9781118971772.ch2>.
- Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137**, 553–597, <https://doi.org/10.1002/qj.828>.
- Driemel, A., and Coauthors, 2018: Baseline Surface Radiation Network (BSRN): Structure and data description (1992–2017). *Earth Syst. Sci. Data*, **10**, 1491–1501, <https://doi.org/10.5194/essd-10-1491-2018>.
- Ferreira, V. G., and Z. Asiah, 2015: An investigation on the closure of the water budget methods over Volta basin using multi-satellite data. *IGFS 2014*, International Association of Geodesy Symposia Series, Vol. 144, Springer, 171–178.
- GMAO, 2015a: MERRA-2 tavgM_2d_lnd_Nx: 2d, Monthly mean, Time-Averaged, Single-Level, Assimilation, Land Surface Diagnostics V5.12.4. NASA Goddard Earth Science Data and Information Services Center, accessed 31 May 2017, <http://doi.org/10.5067/8S35XF81C28F>.
- , 2015b: MERRA-2 tavgM_2d_flux_Nx: 2d, Monthly mean, Time-Averaged, Single-Level, Assimilation, Surface Flux Diagnostics V5.12.4. NASA Goddard Earth Science Data and Information Services Center, accessed 31 May 2017, <http://doi.org/10.5067/0JRLVL8YV2Y4>.
- Haddeland, I., and Coauthors, 2011: Multimodel estimate of the global terrestrial water balance: Setup and first results. *J. Hydrometeorol.*, **12**, 869–884, <https://doi.org/10.1175/2011JHM1324.1>.
- Heger, N., G. Abramowitz, R. Knutti, O. Angéilil, K. Lehmann, and B. M. Sanderson, 2018: Selecting a climate model subset to optimise key ensemble properties. *Earth Syst. Dyn.*, **9**, 135–151, <https://doi.org/10.5194/esd-9-135-2018>.
- Hobeichi, S., 2017: Derived optimal linear combination evapotranspiration v1.0. NCI National Research Data Collection, accessed 30 January 2019, <https://doi.org/10.4225/41/58980b55b0495>.
- , 2018: Linear Optimal Runoff Aggregate (LORA) v1.0. NCI National Research Data Collection, accessed 30 January 2019, <https://doi.org/10.25914/5b612e993d8ea>.
- , G. Abramowitz, J. Evans, and A. M. Ukkola, 2018: Derived Optimal Linear Combination Evapotranspiration (DOLCE): A global gridded synthesis ET estimate. *Hydrol. Earth Syst. Sci.*, **22**, 1317–1336, <https://doi.org/10.5194/hess-22-1317-2018>.
- , —, —, and H. E. Beck, 2019: Linear Optimal Runoff Aggregate (LORA): A global gridded synthesis runoff product. *Hydrol. Earth Syst. Sci.*, **23**, 851–870, <https://doi.org/10.5194/hess-23-851-2019>.
- Jiménez, C., and Coauthors, 2011: Global intercomparison of 12 land surface heat flux estimates. *J. Geophys. Res.*, **116**, D02102, <https://doi.org/10.1029/2010JD014545>.
- , B. Martens, D. M. Miralles, J. B. Fisher, H. E. Beck, and D. Fernández-Prieto, 2018: Exploring the merging of the global land evaporation WACMOS-ET products based on local tower measurements. *Hydrol. Earth Syst. Sci.*, **22**, 4513–4533, <https://doi.org/10.5194/hess-22-4513-2018>.
- Jung, M., M. Reichstein, and A. Bondeau, 2009: Towards global empirical upscaling of FLUXNET eddy covariance observations: Validation of a model tree ensemble approach using a biosphere model. *Biogeosciences*, **6**, 2001–2013, <https://doi.org/10.5194/bg-6-2001-2009>.
- Kalman, R. E., 1960: A new approach to linear filtering and prediction problems. *J. Basic Eng.*, **82**, 35–45, <https://doi.org/10.1115/1.3662552>.
- Kalnay, E., 2003: *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge University Press, 368 pp., <https://doi.org/10.1017/CBO9780511802270>.
- , and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.*, **77**, 437–471, [https://doi.org/10.1175/1520-0477\(1996\)077<0437:TNYRP>2.0.CO;2](https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2).
- Kanamitsu, M., W. Ebisuzaki, J. Woollen, S.-K. Yang, J. J. Hnilo, M. Fiorino, and G. L. Potter, 2002: NCEP-DOE AMIP-II Reanalysis (R-2). *Bull. Amer. Meteor. Soc.*, **83**, 1631–1644, <https://doi.org/10.1175/BAMS-83-11-1631>.
- Kato, H., M. Rodell, M. Beyrich, H. Cleugh, E. van Gorsel, H. Liu, and T. P. Meyers, 2007: Sensitivity of land surface simulations to model physics, land characteristics, and forcings, at four CEOP sites. *J. Meteor. Soc. Japan*, **85A**, 187–204, <https://doi.org/10.2151/jmsj.85A.187>.
- Landerer, F. W., and S. C. Swenson, 2012: Accuracy of scaled GRACE terrestrial water storage estimates. *Water Resour. Res.*, **48**, W04531, <https://doi.org/10.1029/2011WR011453>.
- L'Ecuyer, T. S., and G. L. Stephens, 2002: An estimation-based precipitation retrieval algorithm for attenuating radars. *J. Appl. Meteor.*, **41**, 272–285, [https://doi.org/10.1175/1520-0450\(2002\)041<0272:AEBPRA>2.0.CO;2](https://doi.org/10.1175/1520-0450(2002)041<0272:AEBPRA>2.0.CO;2).
- , and Coauthors, 2015: The observed state of the energy budget in the early twenty-first century. *J. Climate*, **28**, 8319–8346, <https://doi.org/10.1175/JCLI-D-14-00556.1>.

- Liu, W., and Coauthors, 2018: Investigating water budget dynamics in 18 river basins across the Tibetan Plateau through multiple datasets. *Hydrol. Earth Syst. Sci.*, **22**, 351–371, <https://doi.org/10.5194/hess-22-351-2018>.
- Loeb, N., 2017: CERES Level 3B EBAF-Surface Terra+AquapnetCDF file—Edition 4.0. NASA Langley Atmospheric Science Data Center DAAC, accessed 31 August 2017, https://doi.org/10.5067/terra+aquap/ceres/ebaf-surface_l3b004.0.
- Loew, A., J. Peng, and M. Borsche, 2016: High-resolution land surface fluxes from satellite and reanalysis data (HOLAPS v1.0): Evaluation and uncertainty assessment. *Geosci. Model Dev.*, **9**, 2499–2532, <https://doi.org/10.5194/gmd-9-2499-2016>.
- Lorenz, C., M. J. Tourian, B. Devaraju, N. Sneeuw, and H. Kunstmann, 2015: Basin-scale runoff prediction: An ensemble Kalman filter framework based on global hydrometeorological data sets. *Water Resour. Res.*, **51**, 8450–8475, <https://doi.org/10.1002/2014WR016794>.
- Luo, L., E. F. Wood, and M. Pan, 2007: Bayesian merging of multiple climate model forecasts for seasonal hydrological predictions. *J. Geophys. Res.*, **112**, D10102, <https://doi.org/10.1029/2006JD007655>.
- Martens, B., and Coauthors, 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>.
- Menne, M. J., I. Durre, R. S. Vose, B. E. Gleason, and T. G. Houston, 2012: An overview of the global historical climatology network-daily database. *J. Atmos. Oceanic Technol.*, **29**, 897–910, <https://doi.org/10.1175/JTECH-D-11-00103.1>.
- Miralles, D. G., T. R. H. Holmes, R. A. M. De Jeu, J. H. Gash, A. G. C. A. Meesters, and A. J. Dolman, 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>.
- Mitchell, K. E., 2004: The multi-institution North American Land Data Assimilation System (NLDAS): Utilizing multiple GCIP products and partners in a continental distributed hydrological modeling system. *J. Geophys. Res.*, **109**, D07S90, <https://doi.org/10.1029/2003JD003823>.
- Mu, Q., M. Zhao, and S. W. Running, 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>.
- Mueller, B., and Coauthors, 2011: Evaluation of global observations-based evapotranspiration datasets and IPCC AR4 simulations. *Geophys. Res. Lett.*, **38**, L06402, <https://doi.org/10.1029/2010GL046230>.
- Munier, S., and F. Aires, 2018: A new global method of satellite dataset merging and quality characterization constrained by the terrestrial water budget. *Remote Sens. Environ.*, **205**, 119–130, <https://doi.org/10.1016/j.rse.2017.11.008>.
- , —, S. Schlaffer, C. Prigent, F. Papa, P. Maisongrande, and M. Pan, 2014: Combining datasets of satellite retrieved products for basin-scale water balance study. Part II: Evaluation on the Mississippi Basin and closure correction model. *J. Geophys. Res.*, **119**, 12 100–12 116, <https://doi.org/10.1002/2014JD021953>.
- ORNL DAAC, 2015: FLUXNET. Oak Ridge National Laboratory, accessed 1 December 2016, <https://fluxnet.ornl.gov>.
- Pan, M., and E. F. Wood, 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>.
- , —, R. Wójcik, and M. McCabe, 2008: Estimation of regional terrestrial water cycle using multi-sensor remote sensing observations and data assimilation. *Remote Sens. Environ.*, **112**, 1282–1294, <https://doi.org/10.1016/j.rse.2007.02.039>.
- , A. K. Sahoo, T. J. Troy, R. K. Vinukollu, J. Sheffield, and A. E. F. Wood, 2012: Multisource estimation of long-term terrestrial water budget for major global river basins. *J. Climate*, **25**, 3191–3206, <https://doi.org/10.1175/JCLI-D-11-00300.1>.
- Papale, D., D. A. Agarwal, D. Baldocchi, R. B. Cook, J. B. Fisher, and C. van Ingen, 2012: Database maintenance, data sharing policy, collaboration. *Eddy Covariance*, Springer, 399–424.
- Raupach, M. R., P. R. Briggs, V. Haverd, E. A. King, M. Paget, and C. M. Trudinger, 2009: Australian Water Availability Project (AWAP): CSIRO Marine and Atmospheric Research Component: Final report for phase 3. CAWCR Tech. Rep. 013, 67 pp., http://www.csiro.au/awap/doc/CTR_013_online_FINAL.pdf.
- Reichle, R. H., R. D. Koster, G. J. M. De Lannoy, B. A. Forman, Q. Liu, S. P. P. Mahanama, and A. Touré, 2011: Assessment and enhancement of MERRA land surface hydrology estimates. *J. Climate*, **24**, 6322–6338, <https://doi.org/10.1175/JCLI-D-10-05033.1>.
- Rodell, M., and Coauthors, 2004: The Global Land Data Assimilation System. *Bull. Amer. Meteor. Soc.*, **85**, 381–394, <https://doi.org/10.1175/BAMS-85-3-381>.
- , and Coauthors, 2015: The observed state of the water cycle in the early twenty-first century. *J. Climate*, **28**, 8289–8318, <https://doi.org/10.1175/JCLI-D-14-00555.1>.
- Rodgers, C. D., 2000: *Inverse Methods for Atmospheric Sounding: Theory and Practice*. Series on Atmospheric, Oceanic and Planetary Physics, Vol. 2, World Scientific, 256 pp., <https://doi.org/10.1142/3171>.
- Rudolf, B., T. Fuchs, U. Schneider, and A. Meyer-Christoffer, 2003: Introduction of the Global Precipitation Climatology Centre (GPCC). Deutscher Wetterdienst, <https://www.dwd.de/EN/ourservices/gpcc/gpcc.html>.
- Sahoo, A. K., M. Pan, T. J. Troy, R. K. Vinukollu, J. Sheffield, and E. F. Wood, 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>.
- Schamm, K., M. Ziese, K. Raykova, A. Becker, P. Finger, A. Meyer-Christoffer, and U. Schneider, 2016: GPCC Full Data Daily Version 1.0: Daily Land-Surface Precipitation from Rain Gauges built on GTS based and Historic Data. National Center for Atmospheric Research Computational and Information Systems Laboratory Research Data Archive, accessed 31 January 2019, <https://doi.org/10.5065/D6V69GRT>.
- Schellekens, J., and Coauthors, 2017: A global water resources ensemble of hydrological models: The earth2Observe Tier-1 dataset. *Earth Syst. Data*, **9**, 389–413, <https://doi.org/10.5194/essd-9-389-2017>.
- Sheffield, J., and E. F. Wood, 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>.
- , G. Goteti, and E. F. Wood, 2006: Development of a 50-year high-resolution global dataset of meteorological forcings for land surface modeling. *J. Climate*, **19**, 3088–3111, <https://doi.org/10.1175/JCLI3790.1>.
- , C. R. Ferguson, T. J. Troy, E. F. Wood, and M. F. McCabe, 2009: Closing the terrestrial water budget from satellite remote sensing. *Geophys. Res. Lett.*, **36**, L07403, <https://doi.org/10.1029/2009GL037338>.
- Siemann, A. L., N. Chaney, and E. F. Wood, 2018: Development and validation of a long term, global, terrestrial sensible heat

- flux dataset. *J. Climate*, **31**, 6073–6095, <https://doi.org/10.1175/JCLI-D-17-0732.1>.
- Skok, G., N. Žagar, L. Honzak, R. Žabkar, J. Rakovec, and A. Ceglar, 2016: Precipitation intercomparison of a set of satellite- and raingauge-derived datasets, ERA Interim reanalysis, and a single WRF regional climate simulation over Europe and the North Atlantic. *Theor. Appl. Climatol.*, **123**, 217–232, <https://doi.org/10.1007/s00704-014-1350-5>.
- Stackhouse, P. W., Jr., S. K. Gupta, S. J. Cox, T. Zhang, J. C. Mikovitz, and L. M. Hinkelman, 2011: 24.5-year surface radiation budget data set released. *GEWEX News*, International GEWEX Project Office, Silver Spring, MD, No. 21, http://www.gewex.org/gewex-content/files_mf/1432209318Feb2011.pdf.
- Stephens, G. L., and Coauthors, 2012: An update on Earth's energy balance in light of the latest global observations. *Nat. Geosci.*, **5**, 691–696, <https://doi.org/10.1038/ngeo1580>.
- Su, Z., 2002: The Surface Energy Balance System (SEBS) for estimation of turbulent heat fluxes. *Hydrol. Earth Syst. Sci.*, **6**, 85–100, <https://doi.org/10.5194/hess-6-85-2002>.
- Swenson, S., and J. Wahr, 2006: Post-processing removal of correlated errors in GRACE data. *Geophys. Res. Lett.*, **33**, L08402, <https://doi.org/10.1029/2005GL025285>.
- , P. J. F. Yeh, J. Wahr, and J. Famiglietti, 2006: A comparison of terrestrial water storage variations from GRACE with in situ measurements from Illinois. *Geophys. Res. Lett.*, **33**, L16401, <https://doi.org/10.1029/2006GL026962>.
- Trenberth, K. E., and J. T. Fasullo, 2013: North American water and energy cycles. *Geophys. Res. Lett.*, **40**, 365–369, <https://doi.org/10.1002/grl.50107>.
- , L. Smith, T. Qian, A. Dai, and J. Fasullo, 2007: Estimates of the global water budget and its annual cycle using observational and model data. *J. Hydrometeorol.*, **8**, 758–769, <https://doi.org/10.1175/JHM600.1>.
- , J. T. Fasullo, and J. Kiehl, 2009: Earth's global energy budget. *Bull. Amer. Meteor. Soc.*, **90**, 311–324, <https://doi.org/10.1175/2008BAMS2634.1>.
- Van Dijk, A. I. J. M., L. J. Renzullo, Y. Wada, and P. Tregoning, 2014: A global water cycle reanalysis (2003–2012) merging satellite gravimetry and altimetry observations with a hydrological multi-model ensemble. *Hydrol. Earth Syst. Sci.*, **18**, 2955–2973, <https://doi.org/10.5194/hess-18-2955-2014>.
- Vinukollu, R. K., J. Sheffield, E. F. Wood, M. G. Bosilovich, and D. Mocko, 2012: Multimodel analysis of energy and water fluxes: Intercomparisons between operational analyses, a land surface model, and remote sensing. *J. Hydrometeorol.*, **13**, 3–26, <https://doi.org/10.1175/2011JHM1372.1>.
- Vörösmarty, C. J., B. Moore III, A. L. Grace, M. P. Gildea, J. M. Melillo, B. J. Peterson, E. B. Rastetter, and P. A. Steudler, 1989: Continental scale models of water balance and fluvial transport: An application to South America. *Global Biogeochem. Cycles*, **3**, 241–265, <https://doi.org/10.1029/GB003i003p00241>.
- Wang, K., and R. E. Dickinson, 2012: A review of global terrestrial evapotranspiration: Observation, modeling, climatology, and climatic variability. *Rev. Geophys.*, **50**, RG2005, <https://doi.org/10.1029/2011RG000373>.
- Watkins, M. M., D. N. Wiese, D.-N. Yuan, C. Boening, and F. W. Landerer, 2015: Improved methods for observing Earth's time variable mass distribution with GRACE using spherical cap mascons. *J. Geophys. Res. Solid Earth*, **120**, 2648–2671, <https://doi.org/10.1002/2014JB011547>.
- Wiese, D. N., D.-N. Yuan, C. Boening, F. W. Landerer, and 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, accessed 5 April 2019, <https://doi.org/10.5067/TEMSC-3MJC6>.
- Wild, M., D. Folini, C. Schär, N. Loeb, E. G. Dutton, and G. König-Langlo, 2013: The global energy balance from a surface perspective. *Climate Dyn.*, **40**, 3107–3134, <https://doi.org/10.1007/s00382-012-1569-8>.
- , and Coauthors, 2015: The energy balance over land and oceans: An assessment based on direct observations and CMIP5 climate models. *Climate Dyn.*, **44**, 3393–3429, <https://doi.org/10.1007/s00382-014-2430-z>.
- Yao, Y., and Coauthors, 2017: Improving global terrestrial evapotranspiration estimation using support vector machine by integrating three process-based algorithms. *Agric. For. Meteorol.*, **242**, 55–74, <https://doi.org/10.1016/j.agrformet.2017.04.011>.
- Zhang, K., J. S. Kimball, and S. W. Running, 2016: A review of remote sensing based actual evapotranspiration estimation. *Wiley Interdiscip. Rev.: Water*, **3**, 834–853, <https://doi.org/10.1002/wat2.1168>.
- Zhang, Y., and Coauthors, 2015: Multi-decadal trends in global terrestrial evapotranspiration and its components. *Sci. Rep.*, **6**, 19124, <https://doi.org/10.1038/srep19124>.
- , and Coauthors, 2018: A climate data record (CDR) for the global terrestrial water. *Earth Syst. Sci.*, **22**, 241–263, <https://doi.org/10.5194/hess-22-241-2018>.