Assessing Water Balance Closure Using Multiple Data Assimilation– and Remote Sensing–Based Datasets for Canada

JEFFERSON S. WONG,^{a,b,c} XUEBIN ZHANG,^c SHERVAN GHARARI,^{a,b} RAJESH R. SHRESTHA,^d
HOWARD S. WHEATER,^{a,b} AND JAMES S. FAMIGLIETTI^{a,b,e}

^a Global Institute for Water Security, University of Saskatchewan, Saskatoon, Saskatchewan, Canada

^b School of Environment and Sustainability, University of Saskatchewan, Saskatoon, Saskatchewan, Canada

^c Climate Research Division, Environment and Climate Change Canada, Toronto, Ontario, Canada

^d Watershed Hydrology and Ecology Research Division, Environment and Climate Change Canada, University of Victoria, Victoria, British Columbia, Canada

^e Department of Geography and Planning, University of Saskatchewan, Saskatoon, Saskatchewan, Canada

(Manuscript received 3 June 2020, in final form 2 April 2021)

ABSTRACT: Obtaining reliable water balance estimates remains a major challenge in Canada for large regions with scarce in situ measurements. Various remote sensing products can be used to complement observation-based datasets and provide an estimate of the water balance at river basin or regional scales. This study provides an assessment of the water balance using combinations of various remote sensing– and data assimilation–based products and quantifies the nonclosure errors for river basins across Canada, ranging from 90 900 to 1 679 100 km², for the period from 2002 to 2015. A water balance equation combines the following to estimate the monthly water balance closure: multiple sources of data for each water budget component, including two precipitation products—the global product WATCH Forcing Data ERA-Interim (WFDEI), and the Canadian Precipitation Analysis (CaPA); two evapotranspiration products—MODIS, and Global Land surface Evaporation: The Amsterdam Methodology (GLEAM); one source of water storage data—GRACE from three different centers; and observed discharge data from hydrometric stations (HYDAT). The nonclosure error is attributed to the different data products using a constrained Kalman filter. Results show that the combination of CaPA, GLEAM, and the JPL mascon GRACE product tended to outperform other combinations across Canadian river basins. Overall, the error attributions of precipitation, evapotranspiration, water storage change, and runoff were 36.7%, 33.2%, 17.8%, and 12.2%, which corresponded to 8.1, 7.9, 4.2, and 1.4 mm month^{−1}, respectively. In particular, the nonclosure error from precipitation dominated in Western Canada, whereas that from evapotranspiration contributed most in the Mackenzie River basin.

KEYWORDS: North America; Watersheds; Hydrometeorology; Water budget/balance; Error analysis; Kalman filters; Uncertainty

1. Introduction

The dynamics of the terrestrial water cycle have been altered by both climate change and human disturbances (Haddeland et al. 2014; Huntington 2006; Sterling et al. 2013; Vörösmarty and Sahagian 2000). Acceleration and intensification of the water cycle have resulted in significant impacts on the climate system, hydrological regimes, and water availability. Better understanding and characterization of the water cycle is essential for managing water resources, predicting the occurrence of hydrological extremes, and resolving the complex interactions of the land surface, oceans and atmosphere. This could be achieved by accurate and reliable estimation of the terrestrial water budget components. With the increase of data availability for each water budget component from various sources (e.g., satellite remote sensing, climate model reanalysis), the water balance and its uncertainty can be better assessed and quantified at regional or continental scales. This is especially true in regions

where in situ observations are sparse and are subject to numerous measurement errors. However, the existence of multiple products for each water budget component imposes a challenge for potential users, having to choose from a myriad of datasets for application over the areas of interest. This study aims to address this issue by evaluating the water balance performance of combinations of different products from various data sources and quantifying the error attribution from each water budget component across river basins in Canada.

With ongoing development and improvement in science and technology, various types of global and regional datasets related to components of the water budget have been developed. They include in situ station-based interpolated data (e.g., New et al. 1999, 2000), reanalysis-based multiple-source products (e.g., Mesinger et al. 2006; Weedon et al. 2011), and remote sensing–based retrieval outputs (e.g., Miralles et al. 2011; Mu et al. 2007; Swenson and Wahr 2002). As a result, there are multiple datasets for each water budget component. However, each dataset has its own strengths and weaknesses in terms of spatial coverage and temporal resolution, and different types of uncertainties are inherent in the data, including those in reanalysis data assimilation procedures and in the parameterizations of satellite retrieval algorithms (Zhang et al. 2018). Moreover, because of the differences in assumptions, methods, and use of supplementary information, the uncertainties in

Denotes content that is immediately available upon publication as open access.

Corresponding author: Jefferson S. Wong, jefferson.wong@usask.ca

DOI: 10.1175/JHM-D-20-0131.1

© 2021 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy (www.ametsoc.org/PUBSReuseLicenses).

these datasets could introduce potential inconsistencies in water balance estimation and lead to a nonclosure problem (Aires 2014). These incoherencies in datasets also complicate their use in hydrological studies that aim to calibrate/constrain model performance by incorporating other types of observations in addition to streamflow. These data errors and input uncertainties are often transferred and combined with model errors arising from model structure and parameterization; yet, assessment of the consistency of multiple datasets for water balance and exploration of their possible hydrological implications are often neglected (Kauffeldt et al. 2013). Therefore, before utilizing complementary data along with streamflow, evaluating the errors arising from different combinations of water budget component products could provide insight and information for potential users about the consistency and reliability of those products in their model calibration and hydrological applications.

Numerous studies have examined water balance closure over basins from different parts of the world at regional and/or global scales; examples include global major river basins (Munier and Aires 2018; Pan et al. 2012; Sahoo et al. 2011; Syed et al. 2009), the pan-Arctic (e.g., Landerer et al. 2010; Syed et al. 2007; Troy et al. 2011), the Amazon (e.g., Azarderakhsh et al. 2011; Marengo 2005; Oliveira et al. 2014), the Mississippi (e.g., Munier et al. 2014; Sheffield et al. 2009), the Mediterranean (e.g., Pellet et al. 2018; Pellet et al. 2019), the United States (e.g., Gao et al. 2010; Wang and Alimohammadi 2012), Canada (e.g., Szeto 2007; Szeto et al. 2008; Wang et al. 2014a; Wang et al. 2015), China (e.g., Liu et al. 2018; D. Zhang et al. 2016), and many others. These studies have three main focuses when assessing water balance closure. The first focus characterizes the water budget components and quantifies their seasonal to interannual variations. Most of the studies reported substantial water budget nonclosure/imbalances, owing to insufficient accuracies and inconsistencies among the water budget component products. The second focus examines and identifies the existence of trends in the water budget components (e.g., Liu et al. 2018; Oliveira et al. 2014; Wang et al. 2015). The third focus develops methods to account for the error uncertainties associated with the nonclosure issue and eventually combines multiple available datasets of water budget components in a consistent and coherent way (Aires 2014; Pan et al. 2012; Rodell et al. 2015; Y. Zhang et al. 2016; Zhang et al. 2018).

Large-scale water budget studies for Canadian basins have been conducted since the Global Energy and Water Exchanges project (GEWEX) study, with particular attention on the Mackenzie River basin (Louie et al. 2002; Szeto et al. 2008; Woo 2008a,b). Using similar methodology to the GEWEX study and different types of datasets, Szeto (2007) also assessed the water budget for the Saskatchewan River basin. Recent studies by Wang et al. (2014b) characterized the water budget variables for 16 large drainage basins in Canada, and Wang et al. (2014a) studied the long-term national-scale water budget closures for 370 Canadian watersheds. Based on their previous work, Wang et al. (2015) extended the water balance analysis to 19 large Canadian drainage basins, assessing the imbalance in relation to total water storage. They found that the water budget residuals were on average within 10% of the precipitation

among all the basins, and that 31% of the imbalance over 11 basins was caused by significant trends in total water storage. The above studies have made significant advances in understanding the first two focuses mentioned earlier, but investigations into reconciling the nonclosure problem and tracing the error attribution from each water budget component have been limited. In addition, the above studies are primarily based on one dataset for each water budget component without fully examining the error uncertainties arising from multiple products and their interactions.

Different methodologies have been developed to evaluate the uncertainties of each budget component and to address the nonclosure problem. Pan et al. (2012) quantified the error levels of different datasets by identifying the “best” products they believed to be less biased and more reliable and assumed those products provide the true state of the variables (i.e., the “truth”). Other studies (Sahoo et al. 2011; Y. Zhang et al. 2016; Zhang et al. 2018) calculated the error variance of each product using the ensemble mean of all available products as a benchmark. Rodell et al. (2015) used the standard deviation among the existing datasets to represent the error uncertainty. Regardless of the methods used, error uncertainties are quantified by assuming some references as the truth. Enforcing water balance closure was implemented using the constrained ensemble Kalman filter (Pan and Wood 2006), which imposed a linear equality constraint on the estimates of water budget components. Subsequently, a simplified version (nonensemble form) of the constrained Kalman filter was widely applied as a stand-alone procedure after applying a regular Kalman filter, to avoid numerical instabilities (Simon and Chia 2002). The constrained Kalman filter aims to redistribute the nonclosure errors back to each water budget component according to the magnitude of their errors and correlation among them. In this regard, the error attribution of each water budget component could be quantified by the amount of error redistribution. This is ideal for Canada, where there are limited in situ observation networks as benchmarks and large uncertainties in measurements due to the harsh environmental conditions (Wang et al. 2015).

The objectives of this study are as follows: 1) to evaluate the water balance performance of combinations from various remote sensing- and data assimilation-based products at the basin scale; 2) to examine the effects of different combinations from various products on the temporal nonclosure error characteristics; and 3) to quantify the nonclosure error attributions from each water budget component and their uncertainties due to different combinations from various products across major river basins in Canada. The importance of the consistency among the combinations and its implication for hydrological modeling were also examined. The remainder of this paper is organized as follows: a brief description of the study area and data used in this study is provided in sections 2 and 3; the methodology for evaluating water balance performance and quantifying error attribution is described in section 4; results and discussion within the context of implications for hydrological modeling are provided in sections 5 and 6, followed by a summary and conclusion in section 7.

FIG. 1. Geographical locations of the 20 river basins in Canada (in various colors). The river basins were selected to illustrate the spatiotemporal nonclosure error characteristics and attributions from each water budget components. Their hydrometric information is shown in Table 1.

2. Study region

Canada, which covers a land area of 9.9 million km², extends from 42° to 83°N latitude, and spans between 141° and 52°W longitude. With substantial landmass variations, the country can be divided into many regions using aspects such as climate, topography, vegetation, soil, geology, and land use. The standard drainage area classification was developed to delineate hydrologic areas over all the land and interior freshwater lakes of the country, with three levels of classification (11 major drainage areas, 164 subdrainage areas, and 974 sub-sub-drainage areas) (Brooks et al. 2002; Pearse et al. 1985). River basins chosen for assessment in this study were based on the following criteria: 1) the hydrometric status of the gauge station is active, 2) its operating mode is continuous, 3) the length of data is sufficient for long-term study (≥ 30 years), including recent years up to 2015, 4) the percentage of missing values is small ($\leq 5\%$), and 5) the drainage area is at least 90 000 km², following the criterion used in the study of Wang et al. (2014b). As a result, 20 major river basins, based on the Pearse et al. (1985) classification, were selected (Fig. 1 and Table 1) to illustrate the effects of different combinations from various

products on the temporal nonclosure error characteristics and attributions from each water budget component. Their drainage areas range from 90 900 to 1 679 100 km².

3. Data

The products used in this study to represent the water budget components of precipitation, evapotranspiration, and total water storage are summarized in Table 2, which shows their full names and original spatial and temporal resolutions for the versions used. A brief description of the products is provided in the following.

a. Precipitation

Based on the intercomparison work of Wong et al. (2017), two gridded precipitation P products were found to perform well over Canada, with respect to different performance measures: 1) the global product by WATCH Forcing Data ERA-Interim (WFDEI), augmented by Global Precipitation Climatology Centre (GPCC) data [WFDEI (GPCC)], and 2) the Canadian Precipitation Analysis (CaPA) product. Therefore, these two products were used in this study.

TABLE 1. Selected river basins across Canada to illustrate the spatiotemporal nonclosure error characteristics and attributions from each water budget component. The locations of the river basins are shown in Fig. 1.

No.	Station ID	Station name	River basin	Drainage area (km ²)
1	02KF005	Ottawa River at Britannia	Ottawa	90 900
2	03OE001	Churchill River above Upper Muskrat Falls	Churchill River	92 500
3	04AB001	Hayes River below Gods River	Hayes	103 000
4	04HA001	Albany River near Hat Island	Albany	118 000
5	05KL001	Saskatchewan River at Grand Rapids	Saskatchewan	406 000
6	05MJ001	Assiniboine River at Headingley	Assiniboine	162 000
7	05OC008	Red River near St. Norbert	Red	122 000
8	05PF063	Winnipeg River at Slave Falls	Winnipeg	126 000
9	05UE005	Nelson River at Kelsey Generating Station	Nelson	1 050 000
10	06FD001	Churchill River above Red Head Rapids	Churchill	289 000
11	07DA001	Athabasca River below Fort McMurray	Athabasca	133 000
12	07KC001	Peace River at Peace Point (Alberta)	Peace	293 000
13	07NB001	Slave River at Fitzgerald (Alberta)	Slave	606 000
14	08MF005	Fraser River at Hope	Fraser	217 000
15	08NE058	Columbia River at International Boundary	Columbia	156 000
16	09CD001	Yukon River above White River	Yukon	149 000
17	10ED002	Liard River near The Mouth	Liard	275 000
18	10LC014	Mackenzie River at Arctic Red River	Mackenzie	1 679 100
19	10JC003	Great Bear River at Outlet Of Great Bear Lake	Great Bear	146 400
20	10RC001	Back River above Hermann River	Back	93 900

1) WFDEI (GPCC)

The European Union Water and Global Change (WATCH) Forcing Data methodology applied to the ERA-Interim (WFDEI) was developed to provide datasets of subdaily (3-hourly) and daily meteorological data, with global coverage at 0.5° spatial resolution (~50 km) from 1979 to 2012 (Weedon et al. 2014). WFDEI has been updated to provide datasets up to 2016. Using the same methodology as WATCH (Weedon et al. 2011), WFDEI was constructed based on the European Centre for Medium-Range Weather Forecasts (ECMWF) interim reanalysis product (Dee et al. 2011), combined with the Climatic Research Unit (CRU) monthly data and the GPCC monthly variables. Therefore, WFDEI had two sets of data generated by using either CRU or GPCC precipitation totals. Only GPCC was used in this study [hereafter known as WFDEI (GPCC)].

2) CAPA

The Canadian Precipitation Analysis (CaPA) was created to provide a dataset of 6-hourly precipitation accumulation for the North America domain in real time, at a spatial resolution of 15 km (from 2002 onward) (Mahfouf et al. 2007). For Canada, short-term precipitation forecasts from the Canadian Meteorological Centre (CMC) regional Global Environmental Multiscale (GEM) model (Côte et al. 1998a,b) were used as the background field, with rain gauge measurements from the National Climate Data Archive of Environment and Climate Change Canada (NCDA) used as the observations to generate data using an optimum interpolation technique (Daley 1993). With its continuous development and different configurations, CaPA has updated its statistical interpolation method (Lespinas et al. 2015), has increased its spatial resolution to 10 km, and has

assimilated precipitation estimates from the Canadian weather radar network, as well as U.S. radar near the border (Fortin et al. 2015).

b. Evapotranspiration (ET)

Two ET products with various combinations of data sources and ET algorithms were used in this study: 1) a product derived from the NASA Moderate Resolution Imaging Spectroradiometer (MODIS) satellite, and 2) Global Land surface Evaporation: The Amsterdam Methodology (GLEAM).

1) MODIS

A remote sensing ET algorithm based on Cleugh et al. (2007) and the MODIS satellite data was developed to produce a dataset of 8-day potential and actual evapotranspiration over the globe at a spatial resolution of 500 m (from 2000 onward) (Mu et al. 2007). This dataset was constructed based on the Penman–Monteith method, driven by the Global Modeling and Assimilation Office daily meteorological reanalysis data (Global Modeling and Assimilation Office 2004) and MODIS derived data for land cover, albedo, leaf area index, and enhanced vegetation index. Mu et al. (2011) have further improved the ET algorithm by considering vegetation cover fraction, soil heat flux, daytime and nighttime evapotranspiration, including evaporation from wet canopy surfaces and from saturated wet soil surfaces, and improving calculations of stomatal conductance, aerodynamic resistance and boundary layer resistance. Hereafter we call this ET product MODIS.

2) GLEAM

The Global Land Evaporation: The Amsterdam Methodology (GLEAM) is a set of algorithms that provide datasets of daily

TABLE 2. Selected products in this study to represent water budget components of precipitation, evapotranspiration, and total water storage.

Water budget component	Dataset	Full name	Spatial resolution	Temporal resolution	Duration	Coverage	Reference
Precipitation	CaPA	Canadian Precipitation Analysis	10 km (~0.0833°)	6 h	2002–15	North America	Mahfouf et al. (2007)
	WFDEI (GPCC)	Water and Global Change Forcing Data Methodology applied to ERA-Interim (Global Precipitation Climatology Centre)	0.5° (~50 km)	3 h	1979–2016	Global	Weedon et al. (2014)
Evapotranspiration	GLEAM	Global Land Evaporation: The Amsterdam Methodology	0.25° (~25 km)	24 h	1980–2016	Global	Miralles et al. (2011)
Total water storage	MODIS	Moderate Resolution Imaging Spectroradiometer	500 m (~0.00416°)	8 days	2000–18	Global	Mu et al. (2007)
	GRACE (CSR)	Gravity Recovery and Climate Experiment (University of Texas Center for Space Research)	1° (~100 km)	1 month	2002–17	Global	Bettadpur (2007)
	GRACE (GFZ)	Gravity Recovery and Climate Experiment (Deutsches GeoForschungsZentrum)	1° (~100 km)	1 month	2002–17	Global	Dahle et al. (2012)
	GRACE (JPL)	Gravity Recovery and Climate Experiment (NASA's Jet Propulsion Laboratory)	1° (~100 km)	1 month	2002–17	Global	Watkins and Yuan (2012)
	GRACE (JPL-M)	Gravity Recovery and Climate Experiment (NASA's Jet Propulsion Laboratory-Mascons)	0.5° (~50 km)	1 month	2002–17	Global	Watkins et al. (2015)

evaporation at the global scale and 0.25° spatial resolution (~25 km) from 1984 to 2011 (Miralles et al. 2011). The components in generating GLEAM include a modified Priestly–Taylor model (Priestley and Taylor 1972), an evaporative stress module, and a Gash analytical model of rainfall interception (Miralles et al. 2010). The datasets were driven by the Climate Prediction Center morphing technique (CMORPH) precipitation data (Joyce et al. 2004), supplemented by the Global Precipitation Climatology Project (GPCP) product (Huffman et al. 2001) and various remote sensing products for net radiation, surface air temperature, surface soil moisture, vegetation optical depth, and snow depth (Miralles et al. 2011). Driven by the Tropical Rainfall Measurement Mission (TRMM) Multisatellite Precipitation Analysis (TMPA) 3B42v7 product (Huffman et al. 2007) and the Multi-Source Weighted-Ensemble Precipitation (MSWEP) dataset (Beck et al. 2017), GLEAM has been updated to include a new soil moisture data assimilation system (Martens et al. 2016; Miralles et al. 2014), a revised evaporative stress formula, and an optimized drainage algorithm (Martens et al. 2017). The datasets are currently available in three versions. This study used the 1980–2016 version 3.1a at 0.25° at daily time steps.

c. Total water storage

The Gravity Recovery and Climate Experiment (GRACE) twin satellites were launched in March 2002 to provide a dataset of monthly measurements of changes in Earth's gravity field, with global coverage at a spatial resolution of ~220 km (Swenson and Wahr 2002; Tapley et al. 2004). The changes in gravity anomalies were processed using the spherical harmonic approach at 1° spatial resolution (~100 km) to estimate the changes in total water storage (TWS) by three centers: 1) the University of Texas Center for Space Research (CSR) (Bettadpur 2007), 2) NASA's Jet Propulsion Laboratory (JPL) (Watkins and Yuan 2012), and 3) Deutsches GeoForschungsZentrum (GFZ) (Dahle et al. 2012). The JPL had alternatively processed the GRACE data using the mass concentration (mascon) solution (JPL-M) (Watkins et al. 2015). The availability of GRACE data for measuring TWS has continued with a GRACE Follow-On mission launched in 2018. The challenges in using GRACE arise from various uncertainties inherent in the data, such as errors related to atmospheric, oceanic and glacial isostatic effects, signal degradation due to gravity measurement noise, and systematic errors correlated to spatial and spectral domains (Landerer and Swenson 2012; Swenson and Wahr 2006b; Wiese et al. 2016). The data accompanied by the scaling grid (gain factor) were thus scaled to 1° (spherical harmonic) and 0.5° (mascon) datasets to minimize the attenuation of surface mass variations at small spatial scales caused by sampling and postprocessing. However, the effective scale of GRACE data is larger than the grid resolution of 1° and it is suggested that the GRACE effective scale of use is about 150 000–200 000 km² (Longuevergne et al. 2010; Rowlands et al. 2005; Swenson and Wahr 2006b). It can be noted that for the range of basins investigated here (90 900–1 679 100 km²), 10 basins fall below this suggested effective scale range. It is acknowledged that application of GRACE data at these smaller scales will result in additional errors due to spatial smoothing.

Nevertheless, the analysis reveals the extent to which the data can be useful, even at these smaller scales. These three products, average of the spherical harmonic products (equal weighting of CSR, GFZ, and JPL) (hereafter called ALL), and the coastline resolution improvement–filtered (CRI) version of the mascon product were used in this study to reflect the noise in the gravity field solutions within the available scatter of the solutions.

d. Streamflow

The HYDAT database from the National Water Data Archive of the Water Survey of Canada (<https://www.canada.ca/en/environment-climate-change/services/water-overview/quantity/monitoring/survey/data-products-services/national-archive-hydat.html>) provides datasets of daily and monthly means of streamflow Q and water levels over the Canadian hydrometric network. This database includes information on station number and name, latitude and longitude of the station, status (active or discontinued), operating mode (continuous or intermittent), length of data, and drainage area (gross and effective). Peaks and extremes information, as well as ice-influenced conditions, are also available for some stations. The stations are mostly located in southern Canada, with a progressively sparse station density in the North. This study used daily streamflow data, and the numbers of stations chosen were based on the criteria mentioned in section 2.

4. Methodology

a. Water budget analysis using water balance equation

The terrestrial water budget for a river basin over a specific time interval consists of the fluxes of precipitation, evapotranspiration, runoff, and water storage on the land surface and subsurface, which is given by the water balance equation as follows:

$$\sum_{i=t_1}^{t_2} P_i - \sum_{i=t_1}^{t_2} ET_i - \sum_{i=t_1}^{t_2} Q_i - \Delta S = 0, \quad (1)$$

where P is the total precipitation (rainfall and snow); ET is the evapotranspiration (soil and canopy water evaporation, plant transpiration and snow sublimation); Q is the runoff (surface and subsurface flow); ΔS is the change of total water storage (snowpack, vegetation canopy, lakes, wetlands, soil moisture, and groundwater); and t_1 and t_2 are the start and end time steps i used to specify the change of total water storage, respectively.

Before applying the water balance equation, data preprocessing and manipulation were required. It is noteworthy to mention that different products used in this study may have various resolution (and scale) of application that might be different from the size of basins that the products are upscaled or downscaled to. It is acknowledged that there is a scale issue, and hence the products when used at the scale of the smaller catchments will have a smoothing error associated with them. However, as our aim is to explore the implications of the consistency among different combinations from various products for hydrological modeling purposes, we used the common

practices of scaling the products to the basin of interest. Because of the differences in spatial and temporal resolutions of the various products shown in Table 2, all products were first regridded to a common $0.5^\circ \times 0.5^\circ$ spatial resolution. Streamflow data were converted to runoff by normalizing with drainage area. Since MODIS did not provide evaporation from lakes or surface water bodies which could constitute a significant portion of a river basin, the lack of evaporation from water surfaces was accounted for by adding the open water evaporation from GLEAM. The ΔS for each month should be the difference between the end and the beginning of that particular month, which corresponds to the total accumulation of water fluxes from P , ET , and Q of that month. However, as GRACE measurements were given at irregular intervals of about 30 days, this study assumed the ΔS is the difference between two GRACE estimates, representing the average change in TWS. Accordingly, all the monthly values of P , ET , and Q were calculated by adding the daily values between two GRACE intervals for consistency. Different methods of aggregation for other hydrological fluxes have been proposed (Rodell et al. 2004; Swenson and Wahr 2006a); however, it is shown that the method used in this study produced negligible difference in the water balance estimates at basin scales as compared to other methods (Syed et al. 2009).

Water balance performance of combinations from various remote sensing– and data assimilation–based products were assessed in terms of two aspects: 1) spatial variability of estimated runoff, and 2) temporal variation of nonclosure error. The estimated runoff R is defined as a residual of precipitation, evapotranspiration, and change of total water storage:

$$R = P - ET - \Delta S, \quad (2)$$

whereas the nonclosure error r is calculated as a residual of the estimated runoff and the observed streamflow:

$$r = R - Q. \quad (3)$$

b. Error attribution analysis using constrained Kalman filter

The water balance equation given by (1) assumes that there is closure among the water budget components, given the accurate measurements or estimations of the components. However, in reality, combining the estimates of the water budget components does not usually preserve this balance because of the uncertainties arising from individual sources. As a result, the measurement or estimation errors give rise to a residual or imbalance. To account for the residual attributed from each water budget component, a constrained Kalman filter with an equality constraint was applied in this study such that the water balance was closed, similar to the method described by Pan et al. (2012).

Let $\mathbf{X}$ be a complete set of water budget component that represents the “true value” of fluxes and storage in each basin over time i such that $\mathbf{X} = [P_i \ ET_i \ Q_i \ \Delta S_i]^T$. The balance constraint can be enforced by assuming both the nonclosure error and the observation error covariance are zero, which can be expressed as

$$\mathbf{r} = \mathbf{C}\mathbf{X} = 0, \quad (4)$$

where $\mathbf{r}$ is the nonclosure error or the residual, and $\mathbf{C}$ is the balance constraint defined as $\mathbf{C} = [1 \ -1 \ -1 \ -1]$.

Given a combination of the data products that is an estimate of $\mathbf{X}$, the unclosed set of water budget components, when omitting the time index i , can be written as $\hat{\mathbf{X}} = [\hat{P} \ \hat{ET} \ \hat{Q} \ \hat{\Delta S}]^T$. Since biases exist in most cases, they should be accounted for whenever information is available. However, bias information is usually difficult to obtain. With only one observation measured in $\hat{Q}$ and $\hat{\Delta S}$ (mainly from GRACE albeit different postprocessing), $\hat{Q}$ and $\hat{\Delta S}$ are assumed to be unbiased. The bias in $\hat{P}$ and $\hat{ET}$ can be treated by estimating the mean state for P and ET as the average of each of its observations such that $\bar{P} = \text{mean}(P_n)$ and $\bar{ET} = \text{mean}(ET_n)$, respectively, where the overbar represents the mean and n indicates the numbers of data product. All the sources of information P_n and ET_n are then bias corrected toward this average such that $\bar{P} = P_n - \bar{P}_n - \bar{P}_n$ and $\bar{ET} = ET_n - \bar{ET}_n - \bar{ET}_n$.

Together with their nonclosure errors, $\hat{\mathbf{r}} = \mathbf{C}\hat{\mathbf{X}} \neq 0$, the new set of water budget component estimates $\hat{\mathbf{X}}'$ is calculated as

$$\hat{\mathbf{X}}' = \hat{\mathbf{X}} + \mathbf{K}'(0 - \hat{\mathbf{r}}), \quad (5)$$

where $\mathbf{K}'$ is the Kalman gain, which is given by the following equation:

$$\mathbf{K}' = \mathbf{C}\mathbf{V}'\mathbf{C}^T(\mathbf{C}\mathbf{C}\mathbf{V}'\mathbf{C}^T)^{-1}, \quad (6)$$

where $\mathbf{C}\mathbf{V}'$ is the error covariance of the state estimate $\hat{\mathbf{X}}$, which is defined as

$$\mathbf{C}\mathbf{V}' = \overline{(\hat{\mathbf{X}} - \mathbf{X})(\hat{\mathbf{X}} - \mathbf{X})^T}, \quad (7)$$

where the overbar refers to the expectation or mean of the time series. The error covariance $\mathbf{C}\mathbf{V}'$ has a dimension of 4×4 , which decomposes the error covariance between each individual water budget component into

$$\mathbf{C}\mathbf{V}' = \begin{bmatrix} \varepsilon_{PP} & \varepsilon_{PET} & \varepsilon_{PQ} & \varepsilon_{P\Delta S} \\ \varepsilon_{ETP} & \varepsilon_{ETET} & \varepsilon_{ETQ} & \varepsilon_{ET\Delta S} \\ \varepsilon_{QP} & \varepsilon_{QET} & \varepsilon_{QQ} & \varepsilon_{Q\Delta S} \\ \varepsilon_{\Delta SP} & \varepsilon_{\Delta SET} & \varepsilon_{\Delta S} & \varepsilon_{\Delta S\Delta S} \end{bmatrix}, \quad (8)$$

such that only the lower (or upper) triangle part of the matrix is required to be calculated because of its symmetrical property. The error covariance $\mathbf{C}\mathbf{V}'$ is characterized entry by entry, which is based on the error variances in each data product from available information or expert knowledge. Since the four water budget components are measured or modeled by very different approaches and they have very different nature in errors, the error quantification process differs by variable.

For P and ET , errors of the respective products are not often disclosed by the producers and the estimation of errors for each product requires a reference dataset as “truth.” Ground measurements or interpolated station-based gridded products are often used in this regard; however, these measurements or products are lacking in Canada at the basin scale. As such, when no direct evaluation could reliably assess the product, the

uncertainties of P and ET (ε_{PP} and ε_{ETET}) can be approximated by the standard deviation from all product means of the same physical quantity. The deviations from the mean of all products represent the ensemble spread in the respective water budget component that reflects both systematic and random errors (Tian and Peters-Lidard 2010). Such an approach will probably underestimate the random errors because of the small ensemble size and the fact that ensemble members are not entirely independent of each other. However, the results are still expected to provide insight and information for potential users when the relative magnitude of the uncertainties across the basins is of interest.

For the streamflow data, without any information about the sources of errors (e.g., rating curves, water level measurements, and changes in channel geometry), the uncertainties of Q (ε_{QQ}) are assumed to be 10% (Rodell et al. 2011). The uncertainties of ΔS ($\varepsilon_{\Delta S\Delta S}$) are determined by the information associated with the GRACE data. For the spherical harmonic products, measurement and leakage errors are considered. Measurement errors ($\varepsilon_{\text{measure}}$) are related to instrument and loss signal retrieval errors, whereas leakage errors ($\varepsilon_{\text{leakage}}$) are associated with the spatial filtering process. Instrumental errors will be inversely related to the area, shape, and latitude of the river basins (Rodell and Famiglietti 1999). The combined error ($\varepsilon_{\text{combined}}$) for a given grid cell is computed by summing measurement and leakage errors in quadrature and is given as $\varepsilon_{\text{combined}} = \sqrt{\varepsilon_{\text{measure}}^2 + \varepsilon_{\text{leakage}}^2}$. Using the equation, the gridded uncertainties estimates of, for instance, the Mackenzie, Yukon, and Fraser River basins are 10.5, 20.1, and 37.2 mm, respectively (please refer to Landerer and Swenson (2012) Fig. 6 for further details). In addition, since errors in nearby grid cells are correlated, spatially correlated errors ($\varepsilon_{\text{spatial}}$) for a region are also estimated according to the guidelines provided at <http://gracetellus.jpl.nasa.gov/data/gracemonthlymassgridsland/> and are multiplied by $\sqrt{2}$ (Rodell et al. 2011). Therefore, the uncertainties of ΔS is the consideration of measurement, leakage, and spatially correlated errors such that $\varepsilon_{\Delta S\Delta S} = \varepsilon_{\text{combined}} + \varepsilon_{\text{spatial}}$. Similarly, for the mascon product, error estimates associated with each mascon calculated by Wiese et al. (2016) are used. The provided error estimates are known to be conservative and their uncertainties are scaled to roughly match the magnitude of the errors estimates, as described in Wahr et al. (2006).

The cross-covariance ε_{PET} is assumed to be the product of variances and correlations such that $\varepsilon_{PET} = \varepsilon_{PP}\rho_{PET}\varepsilon_{ETET}$, whereas ε_{PQ} , $\varepsilon_{P\Delta S}$, ε_{ETQ} , and $\varepsilon_{ET\Delta S}$ are assumed to be mutually uncorrelated and zero such that $\varepsilon_{PQ} = \varepsilon_{P\Delta S} = \varepsilon_{ETQ} = \varepsilon_{ET\Delta S} = 0$ owing to the lack of information to quantify them (Pan et al. 2012).

Combining Eqs. (5) and (6), the new set of water budget component estimates becomes

$$\hat{\mathbf{X}}' = \hat{\mathbf{X}} - \mathbf{C}\mathbf{V}'\mathbf{C}^T(\mathbf{C}\mathbf{C}\mathbf{V}'\mathbf{C}^T)^{-1}\hat{\mathbf{r}}. \quad (9)$$

In essence, the balance-constrained estimate is calculated by adjusting the unclosed set of water budget components with an adjustment term, which is the Kalman gain multiplied by the nonclosure errors. The Kalman gain states that the nonclosure error is linearly distributed back to each water

budget component based on their magnitude of errors and correlation among them. Effectively, the larger the errors associated with a specific water budget component, the more nonclosure error is attributed to that component. In other words, the adjustment term is regarded as an error attribution process which provides a breakdown of errors on each water budget component while ensuring a perfectly closed water budget.

The application of Kalman filtering usually involves a dynamical system in which a model is used such that the state variables at time t are updated by the state variables at earlier time $t - 1$ with the consideration of system noise and measurement noise. The introduction of a state constraint (e.g., water balance in this case) into the system might increase the problem dimension and cause numerical instabilities (Simon and Chia 2002). In this study, however, the constrained Kalman filter was in turn applied for each combination of the data products as a standalone procedure without regular Kalman filtering or data assimilation and without the use of any models. Therefore, it is ideal for water balance assessments because of its simplicity and convenient implementation. Nonetheless, the error attribution analysis might be very sensitive to the error terms defined in the first step as the error variances of all data products are critical in controlling the water imbalance allocation.

5. Results

a. Water balance performance

The mean monthly runoff was first estimated from 20 different combinations of various products across Canada from 2002 to 2015. Due to the minimal differences in runoff distribution among the three total water storage spherical harmonic products and the mascon product, the average of the spherical harmonic product (ALL) was only shown in Fig. 2 for better visualization. In general, regardless of the combinations, all provided very similar runoff patterns across Canada, with more than 80 mm month^{-1} estimated runoff on east and west coasts and gradual decrease in amount toward the interior. It is noticeable that there were “red patches” over various regions where estimates were physically infeasible (i.e., negative runoff). Combinations of WFDEI (GPCC) and GLEAM estimated negative runoff along the windward side of the Rocky Mountains and in the lower Yukon. This was similar to the combinations of CaPA and GLEAM. On the other hand, combinations of CaPA and MODIS showed red patches along the leeward side of the Rocky Mountains, in the Prairies, and in part of the subArctic area. These results revealed that the negative runoff could mainly be the result of the interplay between the errors in precipitation and evapotranspiration products.

At basin scales, Fig. 3 displays the best combination of data sources for the 20 chosen river basins across Canada. The combination that provided the minimum amount of mean absolute nonclosure error was regarded as the best combination. It is observed that specific combinations performed particularly well in some of the major river basins. Combination [4] (CaPA, GLEAM, JPL-M) provided the smallest nonclosure

error in the Saskatchewan, Churchill, Assiniboine, and Red River basins, whereas combination [1] (CaPA, GLEAM, and ALL) outperformed other combinations in the lower Mackenzie and Yukon River basins. Similarly, combination [7] (CaPA, MODIS, and CSR) predominantly performed the best in the Liard River basin, while combinations of WFDEI (GPCC) and MODIS with various GRACE products ([16], [17], and [19]) ranked first in the subbasins of the upper Mackenzie River basin (Athabasca, Peace, and Slave, respectively). Further examination of the combinations showed that regardless of the precipitation and evapotranspiration products, the GRACE products using the spherical harmonic approach tended to perform better than the mascon product in basins along the west and east coasts of Canada. This is probably because the overall level of error in the mascon product was increased by the implementation of the CRI filter in which the error was increased by 1 mm in the Yukon and Fraser River basins (Wiese et al. 2016). On the other hand, the areas and the magnitudes of the “red patches” could possibly determine the best combination of precipitation and evapotranspiration products over the river basins of concern, regardless of the total water storage products. Overall, combination [4] had the minimum mean absolute nonclosure error for the whole Canadian domain, probably also due to the least areas of negative runoff estimation.

b. Spatial and temporal variability of nonclosure error

Spatial characteristics of mean absolute nonclosure errors for the 20 river basins were examined at annual and seasonal scales (Fig. 4). Monthly values were averaged over the four standard seasons. At the annual scale (Fig. 4e), the largest mean absolute nonclosure errors were seen along both east and west coasts of Canada, with a range from $29.2 \text{ mm month}^{-1}$ (Ottawa) to $30.8 \text{ mm month}^{-1}$ (Columbia). Most of the basins in the Prairie province (Saskatchewan, Nelson, and Churchill) generally exhibited similar nonclosure error characteristics in which the mean absolute errors were below $19.8 \text{ mm month}^{-1}$. On the other hand, the subbasins in the Mackenzie River basin revealed an internal variability of mean absolute error, ranging from $13.4 \text{ mm month}^{-1}$ in Slave to $20.2 \text{ mm month}^{-1}$ in Liard. At the seasonal scale, most of the basins generally showed the largest mean absolute errors in summer ($24.2 \text{ mm month}^{-1}$), followed by spring ($22.9 \text{ mm month}^{-1}$), autumn ($19.4 \text{ mm month}^{-1}$), and winter ($18.3 \text{ mm month}^{-1}$) in order of decreasing value of nonclosure errors.

The temporal variabilities of nonclosure errors were further examined in the 20 major river basins in Canada (Fig. 5). By considering the spread of the nonclosure errors, consistency among the combinations could be assessed. The seasonal inconsistencies were most pronounced in the Churchill River basin where the largest amount of interquartile range (IQR) for the individual seasons was seen (from spring to winter: 36.4, 48.7, 36.1, and $40.2 \text{ mm month}^{-1}$). While the IQRs of the basins in east and central Canada (Ottawa, Churchill River, and Hayes) exhibited more seasonal variabilities, the IQRs of the Saskatchewan, Mackenzie and its subbasins were fairly even throughout the year. Specifically, when examining the extremes (differences between the 5th and 95th percentiles), the spreads of nonclosure errors were significantly larger in

FIG. 2. Mean monthly runoff estimated from different combinations (a total of 20) of various remote sensing- and data assimilation-based products (see Table 2 for products' information) from 2002 to 2015. In short, CaPA and WFDEI (GPCC) are two precipitation products, and GLEAM and MODIS are two evapotranspiration products. Due to the minimal differences in runoff distribution among the three total water storage spherical harmonic products derived from three different centers (CSR, GFZ, and JPL) using GRACE data and the mascon product processed by JPL (JPL-M), the average of the spherical harmonic products (ALL) is only shown here for simplicity.

January and from April to August among the river basins, with differences varying from 73.6 to 89.6 mm month⁻¹.

The nonclosure errors of the Mackenzie, Saskatchewan, and Yukon River basins could be further examined by comparison with previous studies (Table 3). In general, the recent study of Wang et al. (2014b) and our study show significant improvements in estimating the water balance over Canadian river basins, compared with previous studies, as they have a much smaller range of nonclosure errors. This is probably due to the use of state-of-the-art datasets for each water budget component. In particular, a more direct comparison with Wang et al. (2014b) was done by presenting the total water imbalances of the 20 selected river basins across Canada (Fig. 6). Our study similarly concluded that basins with negative imbalances were mainly found in the North and those with positive nonclosure errors were mostly located in the South, with the exception of basins on the east and west coasts, which also had substantial negative errors. Considering the multiple products for each water budget component, it is not surprising that our study showed a wider range of error uncertainties and a larger magnitude of imbalances. Overall, the total water imbalances for the selected river basins

varied from -22.6 to 10.5 mm month⁻¹, with an average of -1.8 mm month⁻¹ at an all-basin scale.

c. Nonclosure error attribution

The nonclosure error attribution from each water budget component was obtained by forcing the water balance closure through the constrained Kalman filter. Based on the ensemble mean of the 20 total combinations, Fig. 7 shows the mean error attribution from each water budget component of the 20 chosen river basins. Regarding the variability in P as an error source, the largest amount of error was found in central and western Canada, and it decreased gradually to the North. On the other hand, the largest error attribution from ET was mainly located in eastern and central Canada, and the amount of error was the least in the Saskatchewan River basin. In contrast, the error contribution from ΔS was the greatest along the west coast of Canada, followed by the Churchill River basin on the east coast, and generally below 5.8 mm month⁻¹ in central and western Canada. The amount of error attribution from Q was generally small across Canada, except in some river basins along the Rockies where the amount of error could be as high as 5.3 mm month⁻¹. The Albany River basin also

FIG. 3. Best combination of data source (in terms of minimum mean absolute nonclosure error) for the 20 chosen river basins across Canada. In essence, blue indicates the combination from data source of CaPA and GLEAM, yellow indicates the combination from CaPA and MODIS, red indicates the combination from WFDEI and GLEAM, and green indicates the combination from WFDEI and MODIS. The gradient of each color represent GRACE products derived from different centers.

had noticeable error attribution from Q ($2.9 \text{ mm month}^{-1}$), probably owing to winter ice conditions causing ice-jam flooding.

The most dominant error sources of the 20 Canadian river basins were identified by finding the largest percentage of total nonclosure errors (Fig. 8). In general, the major error source was from P , and its range varied from 37.3% to 54.0% among the river basins, which were mainly located in the Prairie province (e.g., Saskatchewan, Nelson). Errors from ET ranged from 36.4% to 56.1%, and they were mostly found in the Mackenzie River basin. The Fraser and Yukon River basins were dominated by errors from ΔS and the contributions are of 35.7% and 38.6%, respectively. It is acknowledged that the most dominant error sources could be identified by simply examining and comparing the uncertainty range of each water budget component. However, the approach used in this study provided an additional insight because the consistency among the four water budget components is considered through

imposing the water balance constraint. This adds new information that is based on the physical knowledge about the water cycle to the unconstrained set of water budget estimates and redistribution of nonclosure errors offers added value to the original estimates.

Table 4 shows the role of different combinations in the relative error contributions of each water budget component. The error source from P was consistently the largest when using WFDEI (GPCC) as the precipitation product (combinations [11]–[20]). However, similar error attributions from P and ET were seen when CaPA was used. In cases where CaPA was combined with MODIS, ET became the most dominant error source. Further examination of the percentage of error attribution from both precipitation products showed that regardless of the evapotranspiration and total water storage products, CaPA generally contributed relatively smaller errors than WFDEI (GPCC) (34.2% versus 39.2%). Similarly, the error source from ET using MODIS

FIG. 4. Mean absolute nonclosure error for the 20 river basins across Canada at different temporal scales: (a) spring, (b) summer, (c) autumn, (d) winter, and (e) annual.

(35.1%) as the evapotranspiration product was constantly larger than using GLEAM (31.4%). Despite the minimal differences in runoff distribution among the total water storage products, some variations of error contributions from ΔS were still observed, ranging from 12.1% to 25.5%. Error attribution from the mascon product (JPL-M) showed the smallest percentage on average (12.6%) among the GRACE

products. When comparing the spherical harmonic products, the error percentage of the equal-weighted GRACE of the three centers (ALL) was the smallest (ALL: 15.2%; CSR: 16.2%; GFZ: 24.3%; JPL: 20.9%) and CSR tended to outperform the other two centers. The better performance of ALL echoed the analysis conducted by Sakumura et al. (2014), which concluded that the ensemble mean of the

FIG. 5. Mean monthly water imbalance (nonclosure error) of the 20 selected river basins across Canada. The red dashed line indicates the median values from the 20 combinations of various remote sensing- and data assimilation-based products. Dark gray and light gray shaded areas represent the 25th–75th and 5th–95th percentiles of the 20 combination ensemble, respectively.

products from the three centers was able to reduce the noise in GRACE data. Regardless of the combinations, the relative error contributions from Q were fairly stable, with a smaller range of 11.4% and 12.9%.

The 20 selected river basins across Canada were used to further illustrate the uncertainties of different combinations on the error attribution of each water budget component (Fig. 9). In general, basins in Eastern and Central Canada (e.g., Hayes,

TABLE 3. Comparison of water budget analysis with previous studies for the Mackenzie, Saskatchewan, and Yukon River basins in Canada.

River basin	Water imbalance		Study period	Source
	Monthly (mm month ⁻¹)	Annual (mm yr ⁻¹)		
Mackenzie	–18 to 32		2003–06	Sahoo et al. (2011)
	–4.5 to –3.9		2002–08	Wang et al. (2014b)
		–29	1960–76	Serreze et al. (2003)
		–11 to 73	1997–2002	Szeto et al. (2008)
Saskatchewan	–5.7 to –1.1	–46.1 to –13.4	2002–15	This study
		7.3 to 142.4	1997–2002	Szeto (2007)
	3.1 to 6.1		2002–08	Wang et al. (2014b)
	0.0 to 7.2		2002–15	This study
Yukon	–57 to 55	0.2 to 86.1	2003–06	Sahoo et al. (2011)
	–9.7 to –4.3		2002–08	Wang et al. (2014b)
	–9.3 to –2.5		2002–15	This study

FIG. 6. Total water imbalance (nonclosure error) of the 20 selected river basins across Canada during 2002–15. The values were calculated from the 20 combinations of various remote sensing- and data assimilation-based products and were shown in box-and-whisker plots with the bottom, band (red line), and top of the box indicating the 25th, 50th (median), and 75th percentiles, respectively.

and Albany) tended to have smaller spread of the error attribution of each water budget component, indicating that the 20 combinations were quite consistent among each other. On the other hand, basins in the Prairies (Saskatchewan, Nelson, and Churchill) and in the upper part of the Mackenzie River basin (Athabasca, and Slave) showed wider spreads of the error contributions mainly from P and ET , whereas more inconsistencies of error attributions largely came from ET and ΔS in basins in the Yukon, Back, and the Great Bear River basin.

6. Discussion

The foregoing analyses reveal the performance of different combinations of various products from a water balance perspective. Our study identified areas and regions where negative runoff was estimated and found the best combination for each chosen river basin in Canada. Regarding which products the potential users might use in their areas of interest for their hydrological applications, finding the “best” combination of data products that provides the minimum amount of absolute nonclosure error offers an alternative and valuable information when no direct evaluation could reliably assess the product for modeling purpose. When there are very few ground measurements over a large area, a direct analysis of the product is subject to large uncertainties. For instance, only 10 precipitation-gauge stations over the Mackenzie

River basin could be used for evaluating the gridded precipitation products (Wong et al. 2017). The “best” precipitation product based on direct evaluation could still fail to represent the spatial variability of the precipitation field. This is potentially worse for evapotranspiration, for which there may be less or even no flux towers existing over the river basin. However, it is argued that the “best” combination with respect to nonclosure errors is not necessarily the most reliable combination because the best combination of P and ET could be potentially obtained by a precipitation product that suffers from systematic dry bias and an evapotranspiration product that has dry bias too. Also, considering the deficiency of the closure constraint, the “best” combination might be marginally better than other combinations and might simply a mathematical artifact. A better option of using the closure constraint could be to integrate the products of each water budget component through optimizing the nonclosure errors. Nonetheless, such practice is not common in any modeling applications and might not often be within the scope of the modeling purposes, thus, at the very least, the “best” combination could be used as the starting choice for model calibration and validation. Whether or not the “best” combination could provide better hydrological performance than other combinations or the integrated datasets remains as a hypothesis to be tested and this could be an avenue for future studies.

FIG. 7. Mean error attribution from each water budget component of the 20 chosen river basins across Canada.

Regarding the error attribution from each water budget component, the magnitude of errors from P and the variability from different precipitation products were shown to be the dominant source of water imbalance across Canada in this study. This is not surprising because P has been usually found to play the major role in the nonclosure error (e.g., [Sheffield et al. 2009](#); [Wang et al. 2014b](#)). However, error contribution from ET should not be underrated because of its dominance over a large portion of the Canadian landmass (e.g., the Mackenzie River basin) and its wider spread of error attribution in some river basins. Consistencies among the evapotranspiration products should be evaluated to reduce the error variabilities, yet, such assessment is constrained for large-scale river basin studies due to the paucity of in situ data (only 26 measurement sites from the Fluxnet Canada Research Network were located within chosen river basins).

Errors from ΔS could appear to be the most dominant source in some regions. This could be partly due to the detection of significant trends of TWS in some of the chosen basins ([Wang et al. 2015](#)). For instance, [Wang et al. \(2015\)](#) showed that a negative TWS trend of -39.5 mm yr^{-1} accounted for 65% of

the corresponding errors obtained for the Yukon River basin during 2003–08. Another reason could be because of the uncertainties arising from the application of gain factors over ice- or glacier-covered regions and the implementation of the CRI filter along some land/ocean boundaries ([Wiese et al. 2016](#)). The gain factors might not be suitable to those basins that have large snow storage and are subject to snow cover reduction, glacier melt, and permafrost thaw arising from climate change, owing to the model inadequacies in estimating ice mass balance during GRACE data postprocessing. Also, implementing the CRI filter was shown to be relatively ineffective in the Yukon and Fraser River basins where both coastal landmass and near-coastal ocean mass are spatially heterogeneous with respect to their neighboring areas ([Wiese et al. 2016](#)). Therefore, the magnitude of the error is highly sensitive to the filtering methods ([Khaki et al. 2018](#); [Vishwakarma et al. 2016](#)) and the correction schemes used to store the signal damage due to filtering ([Vishwakarma et al. 2018](#)). In addition, since the associated error in the GRACE data is inversely related to the size of the drainage area, the explanatory power of GRACE products generally decreases as the basin size decreases. With different and enhanced postprocessing techniques, river basins

FIG. 8. Most dominant error attribution from the water budget component of the 20 chosen river basins across Canada.

that are smaller than the commonly suggested effective scale range are well observed by GRACE in recent studies (e.g., [Khaki et al. 2018](#); [Lorenz et al. 2014](#); [Tourian et al. 2015](#)). The error as a function of basin scale in this study generally followed the inverse relationship identified by previous studies; however, the error associated with smaller river basins was not explicitly accounted for due to the wide spread of scatter, which made quantification of uncertainty difficult. It is thus acknowledged that the error caused by GRACE data reported in this study can be viewed as a lower bound for the actual error caused by this dataset for basins with smaller areas than the recommended scales of application.

Errors from Q and the associated uncertainties are often believed to be relatively small compared to other water budget components. The uncertainties of daily estimated Q under normal conditions are found to be 5%–10% ([Hersch 1999](#)), while estimates in rivers subject to icing are 15%–45% under the worst conditions ([Shiklomanov et al. 2006](#)). This study revealed that error contributions of Q were 12.2% on average, and the error ranges varied from 0.1% to 23.7%, with a tendency toward lower percentages in the South (e.g.,

Saskatchewan and Nelson) and higher percentages in the North (e.g., Back and Liard). This is not surprising because almost 20% of all hydrometric data are affected by ice conditions in Canada ([Hamilton and Moore 2012](#)). Even though the stations are properly maintained by field technicians, the stations (especially in some remote areas in the North) are still subject to freezing during cold and harsh winter conditions. In addition to the uncertainties caused by nonstationary river cross sections, growth of vegetation, and hysteresis in the stage–discharge relationship, ice build-up and ice-jams in winter are also well identified as one of the major sources of uncertainty which causes the true rating curve to vary on short or long time scales ([Hamilton and Moore 2012](#); [McMillan et al. 2012](#)). As a result, two or more rating curves are usually required to better capture the different hydrological conditions of the streamflow (e.g., [Lindenschmidt et al. 2019](#)). Therefore, observational uncertainties from discharge should also be considered when modeling river basins that are more likely subject to ice conditions.

The error attribution analysis provides one of the possible ways to quantify the error contribution and examine the

TABLE 4. Percentage of error attribution from each water budget component of the 20 combinations of various remote sensing– and data assimilation–based products. Please refer to Table 2 for the products' full names; the overall values are highlighted in bold font.

Combination	Data source	Precipitation (%)	Evapotranspiration (%)	Total water storage (%)	Runoff (%)
[1]	CaPA + GLEAM + ALL	36.7	34.3	16.2	12.8
[2]	CaPA + GLEAM + CSR	36.5	33.6	17.3	12.6
[3]	CaPA + GLEAM + GFZ	32.3	30.4	25.5	11.8
[4]	CaPA + GLEAM + JPL-M	36.9	37.2	13.2	12.7
[5]	CaPA + GLEAM + JPL	33.3	32.2	22.1	12.4
[6]	CaPA + MODIS + ALL	34.7	38.1	15.0	12.3
[7]	CaPA + MODIS + CSR	34.6	37.3	15.9	12.1
[8]	CaPA + MODIS + GFZ	31.0	33.7	23.9	11.4
[9]	CaPA + MODIS + JPL-M	34.4	41.4	12.1	12.1
[10]	CaPA + MODIS + JPL	31.6	35.7	20.8	11.9
[11]	WFDEI + GLEAM + ALL	41.8	29.9	15.4	12.9
[12]	WFDEI + GLEAM + CSR	41.6	29.2	16.5	12.7
[13]	WFDEI + GLEAM + GFZ	37.1	26.5	24.6	11.8
[14]	WFDEI + GLEAM + JPL-M	42.3	31.9	13.0	12.8
[15]	WFDEI + GLEAM + JPL	38.2	28.4	20.9	12.5
[16]	WFDEI + MODIS + ALL	39.6	33.7	14.3	12.4
[17]	WFDEI + MODIS + CSR	39.5	33.0	15.2	12.2
[18]	WFDEI + MODIS + GFZ	35.6	29.9	23.1	11.4
[19]	WFDEI + MODIS + JPL-M	39.7	36.0	12.2	12.1
[20]	WFDEI + MODIS + JPL	36.3	32.0	19.7	12.0
	Overall	36.7	33.2	17.8	12.2

consistency among different combinations of various products. If additional information from external sources regarding the nature of the error of any of the products is known, the information could be included to define the related error terms. For instance, if we assume the streamflow data to be correct (i.e., the usual practice of hydrological modeling in calibrating the model against streamflow without considering its observational uncertainty), the error term of the discharge is then set to zero. Due to the linear nature of the Kalman filter, other error terms will be rescaled accordingly to match to 100% of water balance error. However, if the error terms are set to any values other than zero (e.g., we assume there are 10% uncertainty in the discharge), the Kalman filter analysis could be done by assuming the informed error term, to study its effect on the error attribution of other products. On the other hand, it is entirely possible for the true uncertainties to be underestimated/overestimated when there is a lack of a priori information (especially in P and ET), and/or when there is a “false positive” of strong agreement among the products of the same geophysical variables (e.g., there are chances of strong correlation among the P products simply because all the products underestimate the precipitation amount in the North or in the mountains where very few precipitation-gauge stations are available for evaluation, resulting in low spread of P uncertainty). This could result in various combinations of error attributions from each water budget component, depending on the availability of information regarding the nature of error and bias of the products. Cautions are thus needed in interpreting the results as the error attribution is highly dependent on the specified or calculated uncertainties.

A premodeling evaluation similar to this study is essential to shed light on the coherence of the products before any calibrations of the hydrological or land surface models are done.

Knowing how cohesive the products are provides an understanding of how the other water balance components in the models are adjusted or compensated for to obtain the best agreement of the water balance closure in the models. This step is often neglected in hydrological modeling, as models are commonly driven by a given forcing product, calibrated on discharge and a spatially available storage or flux product such as GRACE or MODIS. With increasing numbers of data products for each water budget component, the combination of all error scenarios can mean it is not feasible to carry out an evaluation for all combinations of the products. However, for a limited number of basins and before any modeling practice, the error compositions could be established and evaluated under various scenarios, based on information regarding their error and bias terms. This analysis can provide an estimate of the error uncertainty in each water budget component and modelers can use this information as the weights in an objective function during optimization or as the spread for the likelihood measurement in a Bayesian framework. Either way, a better closure of water balance could be achieved by taking the error distributions of the products into consideration during model evaluation. It is realized that this study only presented the spread of error attribution based on one realization of error estimates in each combination of data products. Further research is thus required to fully account for the error distributions in the observations and the uncertainties of the knowledge of these observation errors.

7. Conclusions

Estimating the terrestrial water budget at regional or continental scales is essential for understanding and predicting water availability for human use and informing planning and

FIG. 9. Error attribution of each water budget component for the 20 selected river basins across Canada. The uncertainties were represented by the 20 combinations of various remotely sensed and data assimilation-based products and were shown in box-and-whisker plots with the bottom, band (red line), and top of the box indicating the 25th, 50th (median), and 75th percentiles, respectively.

decision-making. The availability of various remote sensing- and data assimilation-based products for different water budget components makes it possible to better evaluate the water balance and its uncertainty. This is especially relevant, given the scarcity of direct in situ measurements of various water budget components. This study evaluated the water balance performance of combinations of multiple products and quantified the nonclosure errors from each water budget component, in order to provide insight for potential users regarding choosing products for their specific interests and hydrological applications. Based on the above analysis, the following conclusions can be drawn:

- Overall, a combination of CaPA, GLEAM, and JPL-M outperformed other combinations across the chosen river basins in Canada. The ensemble mean of the GRACE products from the three centers was able to reduce the noise in the GRACE data.
- All combinations displayed the largest mean absolute nonclosure errors in summer ($24.2 \text{ mm month}^{-1}$), followed by

spring ($22.9 \text{ mm month}^{-1}$), autumn ($19.4 \text{ mm month}^{-1}$), and winter ($18.3 \text{ mm month}^{-1}$) in order of decreasing value of nonclosure errors.

- In general, the total water imbalances for the major Canadian river basins ranged from -22.6 to $10.5 \text{ mm month}^{-1}$, with an all-basin mean of $-1.8 \text{ mm month}^{-1}$. Basins with negative nonclosure errors were mostly located in the North and those with positive water imbalances were mainly found in the South. Exceptions were found, with basins on the East and West Coasts also having substantial negative errors.
- Regarding the relative error attributions of each water budget component, the major error source was from P (36.7%), followed by ET (33.2%), ΔS (17.8%), and Q (12.2%). Basins with dominant errors from P were mostly found in major river basins such as the Saskatchewan, and Nelson River basins, whereas basins with major error sources from ET were mainly located in the Mackenzie River basin.
- Errors from Q and the associated uncertainties should not be overlooked given the harsh environmental conditions, such

as winter ice, in Canada. The error contribution of Q tended to be lower in the South and became higher in the North, which could be as high as $5.3 \text{ mm month}^{-1}$.

With the use of recently developed water budget component products, water balance was better estimated as compared with previous studies. However, substantial imbalances could still be found in different basins and seasons. Further studies are needed to identify the reasons for such inconsistencies and to examine their impacts on model calibration and analysis. Findings on water budget closure assessment and uncertainty from this study address the problems of physically inconsistent and uncertain water budget component products in large-scale modeling by providing a way to evaluate those datasets independent of a specific model (Pellet et al. 2019). Such pre-modeling practice is important and is a useful tool in providing guidance to potential users regarding the reliability of different products, increasing the robustness of subsequent hydrological modeling and analyses.

Acknowledgments. Jefferson S. Wong would like to thank for the financial support the Canada Excellence Research Chair in Water Security, and the Canada 150 Research Chair in Hydrology and Remote Sensing made through the Global Institute for Water Security at the University of Saskatchewan. The financial support from the Environment and Climate Change Canada is also gratefully acknowledged. Thanks are due to Yanping Li, Zhenhua Li, and Saman Razavi for their initial thoughts and inspiring discussion.

REFERENCES

- 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>.
- 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>.
- Beck, H. E., A. I. J. M. van Dijk, V. Levizzani, J. Schellekens, D. G. Miralles, B. Martens, and A. de Roo, 2017: MSWEP: 3-hourly 0.25 degrees global gridded precipitation (1979–2015) by merging gauge, satellite, and reanalysis data. *Hydrol. Earth Syst. Sci.*, **21**, 589–615, <https://doi.org/10.5194/hess-21-589-2017>.
- Bettadpur, S., 2007: UTCSR level-2 processing standards document for Level-2 product release 0004. Center for Space Research, University of Texas at Austin, 19 pp.
- Brooks, R., K. Harvey, D. Kirk, F. Soular, P. Paul, and A. Murray, 2002: Building a Canadian digital drainage area framework. *55th Annual CWSA Conf.*, Winnipeg, Manitoba, Canada, Canadian Water Resources Association, 20 pp., https://www.rupertbrooks.ca/downloads/Drainage_Area_Framework.pdf.
- Cleugh, H. A., R. Leuning, Q. Z. Mu, and S. W. Running, 2007: Regional evaporation estimates from flux tower and MODIS satellite data. *Remote Sens. Environ.*, **106**, 285–304, <https://doi.org/10.1016/j.rse.2006.07.007>.
- Côte, J., S. Gravel, A. Methot, A. Patoine, M. Roch, and A. Staniforth, 1998a: The operational CMC-MRB Global Environmental Multiscale (GEM) model. Part I: Design considerations and formulation. *Mon. Wea. Rev.*, **126**, 1373–1395, [https://doi.org/10.1175/1520-0493\(1998\)126<1373:TOCMGE>2.0.CO;2](https://doi.org/10.1175/1520-0493(1998)126<1373:TOCMGE>2.0.CO;2).
- , J. G. Desmarais, S. Gravel, A. Methot, A. Patoine, M. Roch, and A. Staniforth, 1998b: The operational CMC-MRB global environmental Multiscale (GEM) model. Part II: Results. *Mon. Wea. Rev.*, **126**, 1397–1418, [https://doi.org/10.1175/1520-0493\(1998\)126<1397:TOCMGE>2.0.CO;2](https://doi.org/10.1175/1520-0493(1998)126<1397:TOCMGE>2.0.CO;2).
- Dahle, C., F. Flechtner, and C. Gruber, 2012: GFZ GRACE level-2 processing standards document for level-2 product release 0005. GeoForschungsZentrum Potsdam, 20 pp.
- Daley, R., 1993: *Atmospheric Data Analysis*. Cambridge University Press, 472 pp.
- 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>.
- Fortin, V., G. Roy, N. Donaldson, and A. Mahidjiba, 2015: Assimilation of radar quantitative precipitation estimations in the Canadian Precipitation Analysis (CaPA). *J. Hydrol.*, **531**, 296–307, <https://doi.org/10.1016/j.jhydrol.2015.08.003>.
- Gao, H. L., Q. H. Tang, C. R. Ferguson, E. F. Wood, and D. P. Lettenmaier, 2010: Estimating the water budget of major U.S. river basins via remote sensing. *Int. J. Remote Sens.*, **31**, 3955–3978, <https://doi.org/10.1080/01431161.2010.483488>.
- Global Modeling and Assimilation Office, 2004: File specification for GEOS-DAS gridded output version 5.3. Doc. GMAO-1001v5.3, Global Modeling and Assimilation Office, 49 pp., <https://gmao.gsfc.nasa.gov/operations/GMAO-1001v5.3.pdf>.
- Haddeland, I., and Coauthors, 2014: Global water resources affected by human interventions and climate change. *Proc. Natl. Acad. Sci. USA*, **111**, 3251–3256, <https://doi.org/10.1073/pnas.1222475110>.
- Hamilton, A. S., and R. D. Moore, 2012: Quantifying uncertainty in streamflow records. *Can. Water Resour. J.*, **37**, 3–21, <https://doi.org/10.4296/cwrj3701865>.
- Herschey, R. W., 1999: *Hydrometry: Principles and Practices*. Wiley, 511 pp.
- Huffman, G. J., R. F. Adler, M. M. Morrissey, D. T. Bolvin, S. Curtis, R. Joyce, B. McGavock, and J. Susskind, 2001: Global precipitation at one-degree daily resolution from multisatellite observations. *J. Hydrometeorol.*, **2**, 36–50, [https://doi.org/10.1175/1525-7541\(2001\)002<0036:GPAODD>2.0.CO;2](https://doi.org/10.1175/1525-7541(2001)002<0036:GPAODD>2.0.CO;2).
- , and Coauthors, 2007: The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. *J. Hydrometeorol.*, **8**, 38–55, <https://doi.org/10.1175/JHM560.1>.
- Huntington, T. G., 2006: Evidence for intensification of the global water cycle: Review and synthesis. *J. Hydrol.*, **319**, 83–95, <https://doi.org/10.1016/j.jhydrol.2005.07.003>.
- Joyce, R. J., J. E. Janowiak, P. A. Arkin, and P. P. Xie, 2004: CMORPH: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution. *J. Hydrometeorol.*, **5**, 487–503, [https://doi.org/10.1175/1525-7541\(2004\)005<0487:CAMTPG>2.0.CO;2](https://doi.org/10.1175/1525-7541(2004)005<0487:CAMTPG>2.0.CO;2).
- Kauffeldt, A., S. Halldin, A. Rodhe, C. Y. Xu, and I. K. Westerberg, 2013: Disinformative data in large-scale hydrological modelling. *Hydrol. Earth Syst. Sci.*, **17**, 2845–2857, <https://doi.org/10.5194/hess-17-2845-2013>.
- Khaki, M., E. Forootan, M. Kuhn, J. Awange, L. Longuevergne, and Y. Wada, 2018: Efficient basin scale filtering of GRACE satellite products. *Remote Sens. Environ.*, **204**, 76–93, <https://doi.org/10.1016/j.rse.2017.10.040>.
- 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>.

- , J. O. Dickey, and A. Guntner, 2010: Terrestrial water budget of the Eurasian pan-Arctic from GRACE satellite measurements during 2003–2009. *J. Geophys. Res.*, **115**, D23115, <https://doi.org/10.1029/2010JD014584>.
- Lespinas, F., V. Fortin, G. Roy, P. Rasmussen, and T. Stadnyk, 2015: Performance evaluation of the Canadian Precipitation Analysis (CaPA). *J. Hydrometeorol.*, **16**, 2045–2064, <https://doi.org/10.1175/JHM-D-14-0191.1>.
- Lindenschmidt, K. E., P. Rokaya, A. Das, Z. Q. Li, and D. Richard, 2019: A novel stochastic modelling approach for operational real-time ice jam flood forecasting. *J. Hydrol.*, **575**, 381–394, <https://doi.org/10.1016/j.jhydrol.2019.05.048>.
- Liu, W. B., 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>.
- Longuevergne, L., B. R. Scanlon, and C. R. Wilson, 2010: GRACE hydrological estimates for small basins: Evaluating processing approaches on the High Plains Aquifer, USA. *Water Resour. Res.*, **46**, W11517, <https://doi.org/10.1029/2009WR008564>.
- Lorenz, C., H. Kunstmann, B. Devaraju, M. J. Tourian, N. Sneeuw, and J. Riegger, 2014: Large-scale runoff from landmasses: A global assessment of the closure of the hydrological and atmospheric water balances. *J. Hydrometeorol.*, **15**, 2111–2139, <https://doi.org/10.1175/JHM-D-13-0157.1>.
- Louie, P. Y. T., W. D. Hogg, M. D. MacKay, X. Zhang, and R. F. Hopkinson, 2002: The water balance climatology of the Mackenzie basin with reference to the 1994/95 water year. *Atmos.–Ocean*, **40**, 159–180, <https://doi.org/10.3137/ao.400206>.
- Mahfouf, J. F., B. Brasnett, and S. Gagnon, 2007: A Canadian Precipitation Analysis (CaPA) project: Description and preliminary results. *Atmos.–Ocean*, **45**, 1–17, <https://doi.org/10.3137/ao.v450101>.
- Marengo, J. A., 2005: Characteristics and spatio-temporal variability of the Amazon river basin water budget. *Climate Dyn.*, **24**, 11–22, <https://doi.org/10.1007/s00382-004-0461-6>.
- Martens, B., D. G. Miralles, H. Lievens, D. Fernandez-Prieto, and N. E. C. Verhoest, 2016: Improving terrestrial evaporation estimates over continental Australia through assimilation of SMOS soil moisture. *Int. J. Appl. Earth Obs. Geoinfo.*, **48**, 146–162, <https://doi.org/10.1016/j.jag.2015.09.012>.
- , 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>.
- McMillan, H., T. Krueger, and J. Freer, 2012: Benchmarking observational uncertainties for hydrology: Rainfall, river discharge and water quality. *Hydrol. Processes*, **26**, 4078–4111, <https://doi.org/10.1002/hyp.9384>.
- Mesinger, F., and Coauthors, 2006: North American Regional Reanalysis. *Bull. Amer. Meteor. Soc.*, **87**, 343–360, <https://doi.org/10.1175/BAMS-87-3-343>.
- Miralles, D. G., J. H. Gash, T. R. H. Holmes, R. A. M. de Jeu, and A. J. Dolman, 2010: Global canopy interception from satellite observations. *J. Geophys. Res.*, **115**, D16122, <https://doi.org/10.1029/2009JD013530>.
- , 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>.
- , and Coauthors, 2014: El Niño–La Niña cycle and recent trends in continental evaporation. *Nat. Climate Change*, **4**, 122–126, <https://doi.org/10.1038/nclimate2068>.
- Mu, Q. Z., F. A. Heinsch, M. Zhao, and S. W. Running, 2007: Development of a global evapotranspiration algorithm based on MODIS and global meteorology data. *Remote Sens. Environ.*, **111**, 519–536, <https://doi.org/10.1016/j.rse.2007.04.015>.
- , M. S. 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>.
- 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 data sets of satellite-retrieved products for basin-scale water balance study: 2. Evaluation on the Mississippi basin and closure correction model. *J. Geophys. Res. Atmos.*, **119**, 12 100–12 116, <https://doi.org/10.1002/2014JD021953>.
- New, M., M. Hulme, and P. Jones, 1999: Representing twentieth-century space–time climate variability. Part I: Development of a 1961–90 mean monthly terrestrial climatology. *J. Climate*, **12**, 829–856, [https://doi.org/10.1175/1520-0442\(1999\)012<0829:RTCSTC>2.0.CO;2](https://doi.org/10.1175/1520-0442(1999)012<0829:RTCSTC>2.0.CO;2).
- , —, and —, 2000: Representing twentieth-century space–time climate variability. Part II: Development of 1901–96 monthly grids of terrestrial surface climate. *J. Climate*, **13**, 2217–2238, [https://doi.org/10.1175/1520-0442\(2000\)013<2217:RTCSTC>2.0.CO;2](https://doi.org/10.1175/1520-0442(2000)013<2217:RTCSTC>2.0.CO;2).
- Oliveira, P. T. S., M. A. Nearing, M. S. Moran, D. C. Goodrich, E. Wendland, and H. V. Gupta, 2014: Trends in water balance components across the Brazilian Cerrado. *Water Resour. Res.*, **50**, 7100–7114, <https://doi.org/10.1002/2013WR015202>.
- 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>.
- , A. K. Sahoo, T. J. Troy, R. K. Vinukollu, J. Sheffield, and 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>.
- Pearse, P. H., F. Bertrand, and J. W. MacLaren, 1985: Currents of change; Final Report: Inquiry on Federal water policy. Inquiry on Federal Water Policy, 222 pp.
- Pellet, V., F. Aires, A. Mariotti, and D. Fernandez-Prieto, 2018: Analyzing the Mediterranean water cycle via satellite data integration. *Pure Appl. Geophys.*, **175**, 3909–3937, <https://doi.org/10.1007/s00024-018-1912-z>.
- , —, S. Munier, D. Fernández Prieto, G. Jordá, W. A. Dorigo, J. Polcher, and L. Brocca, 2019: Integrating multiple satellite observations into a coherent dataset to monitor the full water cycle—Application to the Mediterranean region. *Hydrol. Earth Syst. Sci.*, **23**, 465–491, <https://doi.org/10.5194/hess-23-465-2019>.
- Priestley, C. H. B., and R. Taylor, 1972: On the assessment of surface heat flux and evaporation using large-scale parameters. *Mon. Wea. Rev.*, **100**, 81–92, [https://doi.org/10.1175/1520-0493\(1972\)100<0081:OTAOSH>2.3.CO;2](https://doi.org/10.1175/1520-0493(1972)100<0081:OTAOSH>2.3.CO;2).
- Rodell, M., and J. Famiglietti, 1999: Detectability of variations in continental water storage from satellite observations of the time dependent gravity field. *Water Resour. Res.*, **35**, 2705–2723, <https://doi.org/10.1029/1999WR000141>.
- , —, J. Chen, S. I. Seneviratne, P. Viterbo, S. Holl, and C. R. Wilson, 2004: Basin scale estimates of evapotranspiration

- using GRACE and other observations. *Geophys. Res. Lett.*, **31**, L20504, <https://doi.org/10.1029/2004GL020873>.
- , E. B. McWilliams, J. S. Famiglietti, H. K. Beaudoin, and J. Nigro, 2011: Estimating evapotranspiration using an observation based terrestrial water budget. *Hydrol. Processes*, **25**, 4082–4092, <https://doi.org/10.1002/hyp.8369>.
- , 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>.
- Rowlands, D. D., and Coauthors, 2005: Resolving mass flux at high spatial and temporal resolution using GRACE intersatellite measurements. *Geophys. Res. Lett.*, **32**, L04310, <https://doi.org/10.1029/2004GL021908>.
- 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>.
- Sakumura, C., S. Bettadpur, and S. Bruinsma, 2014: Ensemble prediction and intercomparison analysis of GRACE time-variable gravity field models. *Geophys. Res. Lett.*, **41**, 1389–1397, <https://doi.org/10.1002/2013GL058632>.
- Serreze, M. C., D. H. Bromwich, M. P. Clark, A. J. Etringer, T. J. Zhang, and R. Lammers, 2003: Large-scale hydro-climatology of the terrestrial Arctic drainage system. *J. Geophys. Res.*, **108**, 8160, <https://doi.org/10.1029/2001JD000919>.
- Sheffield, J., 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>.
- Shiklomanov, A. I., T. I. Yakovleva, R. B. Lammers, I. P. Karasev, C. J. Vorosmarty, and E. Linder, 2006: Cold region river discharge uncertainty—Estimates from large Russian rivers. *J. Hydrol.*, **326**, 231–256, <https://doi.org/10.1016/j.jhydrol.2005.10.037>.
- Simon, D., and T. L. Chia, 2002: Kalman filtering with state equality constraints. *IEEE Trans. Aerosp. Electron. Syst.*, **38**, 128–136, <https://doi.org/10.1109/7.993234>.
- Sterling, S. M., A. Ducharne, and J. Polcher, 2013: The impact of global land-cover change on the terrestrial water cycle. *Nat. Climate Change*, **3**, 385–390, <https://doi.org/10.1038/nclimate1690>.
- Swenson, S., and J. Wahr, 2002: Methods for inferring regional surface-mass anomalies from Gravity Recovery and Climate Experiment (GRACE) measurements of time-variable gravity. *J. Geophys. Res.*, **107**, 2193, <https://doi.org/10.1029/2001JB000576>.
- , and —, 2006a: Estimating large-scale precipitation minus evapotranspiration from GRACE satellite gravity measurements. *J. Hydrometeorol.*, **7**, 252–270, <https://doi.org/10.1175/JHM478.1>.
- , and —, 2006b: Post-processing removal of correlated errors in GRACE data. *Geophys. Res. Lett.*, **33**, L08402, <https://doi.org/10.1029/2005GL025285>.
- Syed, T. H., J. S. Famiglietti, V. Zlotnicki, and M. Rodell, 2007: Contemporary estimates of Pan-Arctic freshwater discharge from GRACE and reanalysis. *Geophys. Res. Lett.*, **34**, L19404, <https://doi.org/10.1029/2007GL031254>.
- , —, and D. P. Chambers, 2009: GRACE-based estimates of terrestrial freshwater discharge from basin to continental scales. *J. Hydrometeorol.*, **10**, 22–40, <https://doi.org/10.1175/2008JHM993.1>.
- Szeto, K. K., 2007: Assessing water and energy budgets for the Saskatchewan River basin. *J. Meteor. Soc. Japan*, **85A**, 167–186, <https://doi.org/10.2151/jmsj.85A.167>.
- , H. Tran, M. D. Mackay, and R. Crawford, 2008: The MAGS water and energy budget study. *J. Hydrometeorol.*, **9**, 96–115, <https://doi.org/10.1175/2007JHM810.1>.
- Tapley, B. D., S. Bettadpur, J. C. Ries, P. F. Thompson, and M. M. Watkins, 2004: GRACE measurements of mass variability in the Earth system. *Science*, **305**, 503–505, <https://doi.org/10.1126/science.1099192>.
- Tian, Y., and C. D. Peters-Lidard, 2010: A global map of uncertainties in satellite-based precipitation measurements. *Geophys. Res. Lett.*, **37**, L24407, <https://doi.org/10.1029/2010GL046008>.
- Tourian, M., O. Elmi, Q. Chen, B. Devaraju, S. Roohi, and N. Sneeuw, 2015: A spaceborne multisensor approach to monitor the desiccation of Lake Urmia in Iran. *Remote Sens. Environ.*, **156**, 349–360, <https://doi.org/10.1016/j.rse.2014.10.006>.
- Troy, T. J., J. Sheffield, and E. F. Wood, 2011: Estimation of the terrestrial water budget over Northern Eurasia through the use of multiple data sources. *J. Climate*, **24**, 3272–3293, <https://doi.org/10.1175/2011JCLI3936.1>.
- Vishwakarma, B. D., B. Devaraju, and N. Sneeuw, 2016: Minimizing the effects of filtering on catchment scale GRACE solutions. *Water Resour. Res.*, **52**, 5868–5890, <https://doi.org/10.1002/2016WR018960>.
- , —, and —, 2018: What is the spatial resolution of GRACE satellite products for hydrology? *Remote Sens.*, **10**, 852, <https://doi.org/10.3390/rs10060852>.
- Vörösmarty, C. J., and D. Sahagian, 2000: Anthropogenic disturbance of the terrestrial water cycle. *BioScience*, **50**, 753–765, [https://doi.org/10.1641/0006-3568\(2000\)050\[0753:ADOTTW\]2.0.CO;2](https://doi.org/10.1641/0006-3568(2000)050[0753:ADOTTW]2.0.CO;2).
- Wahr, J., S. Swenson, and I. Velicogna, 2006: Accuracy of GRACE mass estimates. *Geophys. Res. Lett.*, **33**, L06401, <https://doi.org/10.1029/2005GL025305>.
- Wang, D. B., and N. Alimohammadi, 2012: Responses of annual runoff, evaporation, and storage change to climate variability at the watershed scale. *Water Resour. Res.*, **48**, W05546, <https://doi.org/10.1029/2011WR011444>.
- Wang, S. S., D. W. McKenney, J. L. Shang, and J. H. Li, 2014a: A national-scale assessment of long-term water budget closures for Canada's watersheds. *J. Geophys. Res. Atmos.*, **119**, 8712–8725, <https://doi.org/10.1002/2014JD021951>.
- , J. L. Huang, J. H. Li, A. Rivera, D. W. McKenney, and J. Sheffield, 2014b: Assessment of water budget for sixteen large drainage basins in Canada. *J. Hydrol.*, **512**, 1–15, <https://doi.org/10.1016/j.jhydrol.2014.02.058>.
- , —, D. Q. Yang, G. Pavlic, and J. H. Li, 2015: Long-term water budget imbalances and error sources for cold region drainage basins. *Hydrol. Processes*, **29**, 2125–2136, <https://doi.org/10.1002/hyp.10343>.
- Watkins, M. M., and D.-N. Yuan, 2012: JPL Level-2 processing standards document for Level-2 product release 05. GRACE Doc. 327-744, Jet Propulsion Laboratory, California Institute of Technology, 17 pp.
- , 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>.
- Weedon, G. P., and Coauthors, 2011: Creation of the WATCH forcing data and its use to assess global and regional reference crop evaporation over land during the twentieth century. *J. Hydrometeorol.*, **12**, 823–848, <https://doi.org/10.1175/2011JHM1369.1>.
- , G. Balsamo, N. Bellouin, S. Gomes, M. J. Best, and P. Viterbo, 2014: The WFDEI meteorological forcing data set:

- WATCH forcing data methodology applied to ERA-Interim reanalysis data. *Water Resour. Res.*, **50**, 7505–7514, <https://doi.org/10.1002/2014WR015638>.
- Wiese, D. N., F. W. Landerer, and M. M. Watkins, 2016: Quantifying and reducing leakage errors in the JPL RL05M GRACE mascon solution. *Water Resour. Res.*, **52**, 7490–7502, <https://doi.org/10.1002/2016WR019344>.
- Wong, J. S., S. Razavi, B. R. Bonsal, H. S. Wheeler, and Z. E. Asong, 2017: Inter-comparison of daily precipitation products for large-scale hydro-climatic applications over Canada. *Hydrol. Earth Syst. Sci.*, **21**, 2163–2185, <https://doi.org/10.5194/hess-21-2163-2017>.
- Woo, M.-K., 2008a: *Atmospheric Dynamics*. Vol. 1, *Cold Region Atmospheric and Hydrologic Studies: The Mackenzie GEWEX Experience*, Springer Science & Business Media, 470 pp.
- , 2008b: *Hydrologic Processes*. Vol. 2, *Cold Region Atmospheric and Hydrologic Studies: The Mackenzie GEWEX Experience*, Springer Science & Business Media, 508 pp.
- Zhang, D., Q. Zhang, A. D. Werner, and R. Y. Gu, 2016: Assessment of the reliability of popular satellite products in characterizing the water balance of the Yangtze River Basin, China. *Hydrol. Res.*, **47**, 8–23, <https://doi.org/10.2166/nh.2016.138>.
- Zhang, Y., M. Pan, and E. F. Wood, 2016: On creating global gridded terrestrial water budget estimates from satellite remote sensing. *Surv. Geophys.*, **37**, 249–268, <https://doi.org/10.1007/s10712-015-9354-y>.
- , and Coauthors, 2018: A Climate Data Record (CDR) for the global terrestrial water budget: 1984–2010. *Hydrol. Earth Syst. Sci.*, **22**, 241–263, <https://doi.org/10.5194/hess-22-241-2018>.